AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Personality and Social Psychology:
Interpersonal Relations and Group Processes

© 2022 American Psychological Association
ISSN: 0022-3514

https://doi.org/10.1037/pspi0000264

Who Sees Which Political Falsehoods as More Acceptable and Why: A New Look at In-Group Loyalty and Trustworthiness

Jeff Galak[1] and Clayton R. Critcher[2]
[1] Tepper School of Business, Carnegie Mellon University
[2] Haas School of Business, University of California, Berkeley

Many politicians—even those who occupy some of the most powerful offices in the world—lie. Five studies examined how conservative and liberal Americans responded to media reports of politicians' falsehoods—that is, flagged falsehoods (FFs). Even accounting for partisan biases in how much participants dismissed such reports as fake news and assumed that such lies were unintentional, we consistently observed partisan evaluations in how much FFs were seen as justifiable: Republicans and Democrats alike saw their own party's FFs as more acceptable (Studies 1–4). This charitability did not reflect unconditional in-group favoritism. Instead, it was strongest for policy FFs—those meant to advance a party's explicit agenda—as opposed to personal FFs about a politician's past (Study 2) or electoral FFs that strayed from parties' explicit goals by aiming to disenfranchise legally eligible voters (Study 4). Although FFs can undermine general trustworthiness in the eyes of both in-group and out-group members, policy FFs in particular signal partisan trustworthiness (Studies 3–5)—the belief that a politician can be trusted by their own political side and not by the other. For likeminded partisans, such partisan trustworthiness predicted not only the perceived acceptability of FFs, but also perceptions of the politician as a more prototypically moral actor, even outside of the political sphere. These findings validate the importance of our dual conception of trustworthiness in intergroup contexts.

Keywords: falsehoods, politics, trustworthiness, moral evaluation, loyalty

Supplemental materials: https://doi.org/10.1037/pspi0000264.supp

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

In 2016, the Oxford Dictionary selected the word "posttruth," used frequently with "politics," as its word of the year (Oxford Dictionary, 2016). Although politicians may be lying to an increasing degree (Oborne, 2014), the joke "How do you know a politician is lying? His lips are moving," dates to at least the 1950s. It would seem the phenomenon is not completely new. What is notable is that neither a watchful media nor the informational empowerment afforded by the digital age has kept liars out of the highest offices in the world (Holan, 2015; Kessler & Kelly, 2018). Why is it that calling out their prevarications does not elicit career-killing outrage?

Most simply, we might expect the hyperpolarization in American politics to find its way into evaluations of politicians who are called out for telling untruths. Group memberships guide people's sense of self (Brewer, 1979; Tajfel & Turner, 1979, 1986). Political partisanship has been conceptualized as a core social identity (Green et al., 2004; Mason, 2015). Such social identities influence judgments of the seriousness of their own group members' misdeeds (Dunbar et al., 2016). And, indeed, partisanship has been shown to color perceptions of fellow political in-group members: Democrats and Republicans are more likely to believe false reports that are beneficial to their own political group (Allcott & Gentzkow, 2017), especially when they can convince themselves that those statements *could* have been true even when they are clearly not (Effron, 2018). People discount the unethical actions of other members of their group (Valdesolo & DeSteno, 2007), such that people are less likely to engage in altruistic (i.e., personally costly) punishment of in-group members (Bernhard et al., 2006). Although we do not question these findings, we see this account—merely rooted in in-group favoritism and out-group derogation—to be incomplete.

In this article, we more deeply examine how, when, and why partisanship may color the perceived acceptability or justifiability of politicians' false statements. To approach this question, consider how lies are detected in the real world. The electorate rarely has perfect access to what is true, much less what a politician actually believes is true. For that reason, we identify our focus of study as *flagged falsehoods* (FFs), statements identified by nonpolitical authoritative entities (e.g., the media) as being untrue. Although researchers could bypass these limitations by playing the role of

The research reported in this article was supported in part by U.S. National Science Foundation Award 1749608, awarded to Clayton R. Critcher. Interested readers can find data and materials from all studies reported below here: https://osf.io/j5g6n/?view_only=6967fbc462c146c4abf2608800274962.

Jeff Galak played lead role in data curation, visualization, and creation of study materials; supporting role in validation and writing of original draft; and equal role in conceptualization, formal analysis, funding acquisition, investigation, project administration, and resources. Clayton R. Critcher played lead role in writing of original draft; supporting role in data curation, visualization, and creation of study materials; and equal role in conceptualization, formal analysis, funding acquisition, investigation, project administration, and resources.

Correspondence concerning this article should be addressed to Clayton R. Critcher, Haas School of Business, University of California, Berkeley, 545 Student Services Building #1900, Berkeley, CA 94720, United States. Email: ClaytonCritcher@haas.berkeley.edu

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

omniscient narrator (e.g., "The senator knew what she was saying was untrue . . . "), such a methodological approach may not then offer ecologically valid insights. For example, such an empirical approach may miss that partisans respond to falsehoods differently, at least in part, because they disagree about the extent to which the FF is an intentional inaccuracy. In this vein, we first identify two ways in which partisans may arrive at different conclusions about a political statement flagged as a falsehood, which would then justify different evaluations regarding the acceptability of the utterance. After articulating such routes, we move to what is the central focus on the present work: development of an explanation for why partisans may still display disagreement about FFs' acceptability, above and beyond these excuse-offering pathways that we describe next.

One such excuse-making route may be found in Collins Dictionary's 2017 word of the year: "fake news" (Collins Dictionary, 2017). In the leadup to the 2016 U.S. presidential election, most Americans both saw and believed intentionally inaccurate news stories (Allcott & Gentzkow, 2017). And as fake news spreads and is thus seen repeatedly, it is more likely to be believed (Pennycook et al., 2018), in particular by those for whom such stories may help create a sense of personal control over an often chaotic and unpredictable political world (Whitson et al., 2019). The known proliferation of fake news has colored some people's perceptions of legitimate information sources. For example, former President Donald Trump frequently weaponized this phrase in an effort to cast doubt on an often-skeptical media. Liars may thus remain in power if the electorate distrusts the sources trying to hold these fibbers to account. That is, the media may not be trusted as a neutral arbiter of truth. When the media calls out a politician for a false statement, sympathetic partisans may think it is the media, not the politician, who have shared an inaccuracy. Consistent with this possibility, Swire et al. (2017) showed that although explanations of why statements made by former President Trump were true or false did lead his supporters to better discriminate his true and false statements, they still showed worse discrimination than did those who were not fans of the (now former) president. Especially given that even nonpartisan fact-checkers can disagree about how exactly to rate the truth value of statements (Lim, 2018), there may be plausible latitude for likeminded partisans to see more truth in their own side's flagged falsehoods.

A second excuse-making route may be found in the feature that is core to definitions of lying. Notably, most definitions of lying are agnostic as to whether the statement is actually true. Instead, what matters is whether the speaker believes their statement to be true, and thus whether they intended to communicate a false statement (Leth, 2021; Mahon, 2008; Walker et al., 2021). Intentionality has long been identified as a key precondition for moral culpability (Cushman, 2008; Malle et al., 2007; Shaver, 1985; Weiner, 1995; Young & Saxe, 2011), especially when it comes to deception (Croson et al., 2003). When intentionality is absent, actions are not seen as indicators of stable character, and faith in an actor can be restored or maintained (Kim et al., 2009). Thus, regardless of whether a member of the electorate thinks a politician's statement is actually true, they may think that the politician believed his or her own statement to be true.

FFs May Invite Partisan Evaluations, Independent of Partisan Excuse-Making

Although we do believe (and indeed show) that both of these routes—each of which involves a (re)interpretation of an intentional inaccuracy in more innocuous terms—often explain partisan evaluations of the perceived acceptability or justifiability of flagged falsehoods, we suggest that these routes offer an incomplete picture. Our theoretical analysis relies on the convergence of two ideas. In so doing, we move beyond a simple appeal to in-group favoritism by fleshing out an account of how political falsehoods can offer signals of moral character. This will offer more nuance to when and why it would be the case that evaluative charitability is extended to some who issue flagged falsehoods.

First, we lean on a recently articulated person-centered approach to morality (Uhlmann et al., 2015). People care about morally relevant actions because of what they signal about moral character (Landy & Uhlmann, 2018)—in particular, whether someone will make a good social partner (Heiphetz et al., 2017; Helzer & Critcher, 2018). Central to such considerations is who can be trusted (Everett et al., 2016). Indeed, trust—although easily broken (Iwai et al., 2018)—has been identified as the most important ingredient for well-functioning relationships (Simpson, 2007). Furthermore, trust—as a willingness to make oneself vulnerable to exploitation by another (Mayer et al., 1995; Rousseau et al., 1998)—almost perfectly describes what the electorate is doing in selecting representatives to act in government on their behalf. Others' perceived trustworthiness makes one willing to take risks by deferring or delegating to them (Colquitt et al., 2007; Swider et al., 2022). The power that political leaders have over those they represent (Dunn, 1988) reinforces the key role that perceived trustworthiness likely serves as a filter for making sense of and evaluating politicians' behavior and character.

Second, and building on this point, we consider that falsehoods vary in what truth they misrepresent, and thus what they signal about the critical quality of trustworthiness. This analysis begins with a consideration of the social function of trust. More generally, moral systems are adaptive because they encourage social harmony (Tepe & Aydinli-Karakulak, 2019) and cooperation (Curry et al., 2019; Delton & Krasnow, 2015). And it is trust within social groups, in particular, that permits cooperation (Tomasello et al., 2012), stability (Van Vugt & Hart, 2004), and effective coordination (Mayer et al., 1995). Because groups reap the benefits of their size through collective action, the ability to trust fellow group members as loyal to the mission of their group (and not merely their own personal agenda) is a key condition for group survival (Misch et al., 2014). This logic foreshadows that those who take a functional approach to morality—seeing it as a system of rules that facilitate the healthy and smooth functioning of group life (Haidt, 2008)—identify in-group loyalty as one of the major foundations of morality (Haidt & Graham, 2007). Indeed, loyalty itself is among a broad set of values that people embrace (Shweder et al., 1997), leading it to also be a key feature of Fiske and colleagues' relational approach to morality (Fiske, 1991; Fiske & Haslam, 2005; Rai & Fiske, 2011), Schwarz's (1992) core value of benevolence, and Curry and colleagues' morality-as-cooperation framework (Curry, 2016; Curry et al., 2019).

Loyalty has an inherent partiality to it (Hirschman, 1970; Oliver, 1999). People signal their loyalty to a group when they take actions that show that they will help to advance the cause and purpose of a group even at a cost to an individual, especially the self (Crone & Laham, 2015; Hildreth et al., 2016). The perceived moral worth of such team players is appreciated quite early in the developmental process (Misch et al., 2014). And although loyalty or commitment to one's group can seem selflessly commendable and can sometimes

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

motivate moral behavior (Hildreth et al., 2016), loyalty can also lead people to compromise other moral concerns (e.g., fairness) to engage in cronyism, nepotism, and other morally questionable acts (Heilman et al., 1992; Hildreth et al., 2016; Padgett & Morris, 2000; Suh et al., 2020; Thau et al., 2015; Umphress et al., 2010). In other words, groups can trust their loyal members, even if loyalty does not always mark those who are generally trustworthy, moral individuals.

In some situations, group commitment is demonstrated quite directly: by explicitly sticking with or favoring one's own group at a cost or risk to the self. But in other cases, an individual's degree of loyalty (or disloyalty) to a group is signaled more indirectly. Somewhat ironically, whistleblowers are often disliked by other members of an organization (Cortina & Magley, 2003), for the negative signal of their disloyalty outweighs the fact that they are trying, quite literally, to enhance the morality of the group (McManus, 2021; Waytz et al., 2013). Groups often prefer that members betray their moral standards than that they betray the group. Instead of considering the repercussions of calling out another for spreading falsehoods, we consider the consequences for the fibbers themselves. On this question, we suggest that the nature of the falsehoods—namely, what they reflect about an individual's commitment or loyalty to their own (political) group and its goals—may predict partisan disagreement regarding the moral meaning of such actions and thus the character of the prevaricator.

We consider how the *content* of people's flagged falsehoods does or does not serve to undermine perceptions of the liars as trustworthy, loyal members of the in-group. Numerous taxonomies have been introduced to differentiate lies by their content, motivation, and acceptability (Kashy & DePaulo, 1996; Lindskold & Walters, 1983; Seiter et al., 2002). As one example of how such distinctions bear on the current questions, prosocial deception can be seen as ethical (Levine & Schweitzer, 2014) and even breed trust (Levine & Schweitzer, 2015). In politics, what types of lies may not fully undermine trust—at least among likeminded partisans—and may thus give rise to partisan differences in perceived acceptability?

To begin, we introduce a distinction between policy and personal falsehoods. Policy falsehoods express a false premise upon which a policy position could be based. Personal falsehoods make false claims about an individual's own life or actions. When 2016 U.S. presidential candidate Ben Carson said, "Every time we raise the minimum wage, the number of jobless people increases," he told a policy falsehood (Jacobson, 2015). When Carson also recounted his adolescent involvement in knife fights—part of an evangelical Christian redemption narrative that does not match any contemporaries' recollection of him—or an episode in which he was almost shot while eating at "a Popeye's organization"—a similarly unverified story that was part of an attempt to paint the wealthy neurosurgeon as personally familiar with the challenges of urban life (Gosa, 2017)—he (most likely) told a personal falsehood (Resnick, 2015). This distinction is important to our analysis because these falsehoods may vary in what they seem to reveal about the speaker's trustworthiness (at least to their own political team or party), which may explain why partisanship may color perceptions of some falsehoods more than others.

Political parties organize in an effort to elect candidates who, they hope, will enact the party's policy platform. So central is a policy agenda to political identity that people will endorse contradictory policies because they believe their preferred party backs them (Cohen, 2003). Although such work is often characterized as reflecting the flimsiness of people's ideologies, this work also reinforces the notion that party loyalty demands policy loyalty. After all, loyalty has been defined as the steady, sustained, and practical devotion to a cause (Royce, 1908). Because policy agendas reflect political groups' cause, policy FFs may reveal an individual to be a loyal group member who can still be trusted (by fellow group members), meaning that such loyal in-group members may be seen to merit less reproach by fellow partisans.

Note how this analysis has considered the question of what politicians are trustworthy through a group-centered or partisan lens. Writing about trust, Gambetta (1988) said that "we implicitly mean that the probability that [the trustee] will perform an action that is beneficial or at least not detrimental to us is high enough for us to consider engaging in some form of cooperation with him" (p. 217). But the pronoun "us" is ambiguous in terms of the breadth of the antecedent. Does it refer to people in general, or members of one's in-group (vs. one's out-group)? Considered in the context of political partisanship, we differentiate *general trustworthiness* (a perception that someone can be trusted by people more generally) from *partisan trustworthiness* (a perception that someone can be trusted by their political in-group more than their political out-group).

We suggest this distinction is crucial. After all, our line of argumentation has emphasized the moral relevance of partisan trustworthiness: Flagged falsehoods can signal a politician's commitment to their group and its goals at the expense of a political out-group and its opposing agenda. Thus, by taking a narrow conception of "us," this suggests that flagged falsehoods should have different moral meaning to fellow political in-group members than to out-group members. But previous work has also linked telling falsehoods with the general erosion of trust (Schweitzer et al., 2006). Our dual conception of trustworthiness accommodates such findings by considering that flagged falsehoods may also damage general trustworthiness. From this broader conception of "us," FFs should carry more universal negative moral meaning. Ultimately, we will document empirically how this dual conception of trustworthiness is crucial to making sense of (sometimes partisan) responses to FFs.

Overview of Theoretical Goals and Empirical Approach

We present five studies that aim to contribute to literatures on moral psychology, political psychology, and the role and moral meaning of in-group loyalty and perceived trustworthiness in intergroup contexts. In our studies, participants of varied political orientations learn about a Democratic or Republican politician whose public statements have been called out as falsehoods by a fact-checking source. We then examine whether, when, and why people display partisan evaluations: judging some flagged falsehoods as more acceptable when they come from politicians of their own stripes. We aim to make progress in four main ways, the theoretical value of each we highlight below:

Partisan Evaluations of FF Acceptability

First, we ask whether partisan disagreements about the acceptability of political FFs—if observed—merely reflect disagreement about what occurred (i.e., that the politician actually did tell a falsehood and did so intentionally), or whether they also emerge

as different subjective evaluations of the same agreed-upon facts. This moves beyond previous demonstrations of in-group leniency that have not unconfounded the interpretation from the evaluation of what occurred. Second, we ask whether these effects apply similarly to all political FFs (thereby illustrating simple favoritism toward in-group members; e.g., Locksley et al., 1980) or especially to falsehoods that signal commitment to the political party's explicit agenda. Notably, we make use of the fact that certain falsehoods may stray from what groups explicitly embrace even as those FFs are group-serving.

The Importance of a Dual Conception of Trustworthiness

Third, we explore the role of perceived trustworthiness in these results. We distinguish whether there is a partisan disagreement about whether politicians who tell FFs are generally trustworthy (to Democrats and Republicans alike) or whether partisans actually lean on politicians' apparent partisan trustworthiness in forming divergent moral judgments and evaluations. These efforts have the potential to validate the importance of treating perceived trustworthiness not merely as a unitary construct, but of distinguishing general from partisan trustworthiness. Fourth, we move beyond our examinations of explaining variability in how different people respond to different types of political falsehoods to directly compare the effects of speaking falsehoods versus telling the truth. In so doing, we examine how moral perceptions of those who traffic in untruths may both be tainted (due to the degradation in general trustworthiness they suffer) but also salvaged in the eyes of some (due to the perceptions of partisan trustworthiness they encourage). Such efforts may establish the dual (and sometimes competing) signals sent by flagged falsehoods.

Studies 1a–1d

The first set of four studies tests whether people offer partisan leniency when evaluating their own partisans' policy FFs relating to immigration (Study 1a), the minimum wage (Studies 1b–1c), or school vouchers (Study 1d). Critically, in each study, we also measure participants' beliefs that the statement (flagged by the media as a falsehood) is and that the politician believed it to be true. We appreciate that readers may be interested in these forms of excuse-making in their own right: how they are (often) subject to their own partisan biases and how strongly they predict perceptions of FF acceptability. But our primary purpose in including these measures is to allow them to serve as covariates for our critical analyses. In this way, our central hypothesis is that people will display partisan evaluations of these political FFs in such a way that cannot simply be explained by partisan excuse-making about the nature of the flagged falsehoods themselves (i.e., that they are actually true and/or believed to be so by the speaker).

Method

Participants and Design

American participants were recruited from Amazon's Mechanical Turk participant panel ($N_{1a} = 401$; $N_{1b} = 401$; $N_{1c} = 1{,}042$; $N_{1d} = 199$).[1] Of note, Amazon's Mechanical Turk participant panel has been shown to offer a viable population for conducting research on political ideology and beliefs. Specifically, both liberal and conservative participants on Amazon Mechanical Turk mirror those from more traditional national panels in terms of their demographics, psychological differences, and ideologies (Clifford et al., 2015).[2] We used two exclusion criteria. To begin, we excluded all completions from duplicate IP addresses ($n_{1a} = 10$; $n_{1b} = 10$; $n_{1c} = 8$; $n_{1d} = 4$). Moreover, in each study, two or three memory-based attention checks—multiple-choice questions designed to determine whether participants had carefully read and remembered key details from the experimental materials—were included. Participants who incorrectly responded to more than one of these checks were excluded from all further analyses ($n_{1a} = 27$; $n_{1b} = 34$; $n_{1c} = 89$; $n_{1d} = 22$). This left final sample sizes of 364 (Study 1a), 357 (Study 1b), 945 (Study 1c), and 173 (Study 1d). Although including all participants does not change our key results in any meaningful way, we report the results of all key analyses for each study, without exclusions, in the Supplemental Materials.

In each study, participants were randomly assigned to learn about a *Democrat* or a *Republican* who had issued a tweet or a public statement. A media outlet called out the person (almost always a politician) for making a false statement, referenced evidence in support of its claim, and indicated that the person did not reply to requests to comment. Two studies included an additional exploratory condition that had no influence on our effects of interest. In Study 1a, the liar was either a politician or an ordinary citizen. In Study 1c, the lie was said to be the basis or not the basis of the politician's own policy belief (see Appendix A).

Procedure

Although each study focused on a different FF related to one of three distinct policy domains, the experiments' basic structure was almost identical. Participants first saw what was designed to look like a webpage from *The Albuquerque Journal,* the most widely circulated newspaper in New Mexico. To promote experimental realism, the image displayed not merely the newspaper story but a series of headers, links, and banner ads that online readers are accustomed to seeing (see Appendix B). The article was recently dated and had the title "Stay Informed of The Truth" (Study 1a, 1c) or "Steve Wooley ([*party affiliation*]) Tweets False Information about [*policy issue*]" (Studies 1b, 1d).

Although the evidence calling into question the policy FF's veracity was real, the news story (including the politician) was not. We debriefed participants on these facts at the study's conclusion. The newspaper article began by reporting that Steve Wooley, the leading [*party affiliation*] on the Regulatory and Public Affairs committee in the New Mexico State Legislature,[3] had tweeted or

[1] For all studies that were funded by a single lab (Studies 1a–1d, 2a–2b, 3, and 5), sample sizes were based on the available resources of the funding lab. One lab applied this rule more imprecisely by setting sample sizes at either 100 or 200 per condition. The other lab chose more exact sample sizes based on its monthly budget.

[2] Although we ran (and replicated) some of the studies presented in this article using in-lab university participant samples, the liberal skew of such samples makes extrapolation to conservatives suspect. To avoid engaging in such a problematic research practice, we focus on online samples whose political orientation is more representative of the American electorate.

[3] Half of participants in Study 1a were instead told Steve Wooley was an ordinary citizen.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

publicly stated supposed facts relating to his (depending on party affiliation) support for or opposition to the issue. The article displayed a screenshot of the tweet (Studies 1a, 1b, and 1d) or quote from a speech (Study 1c), and claimed that nonpartisan groups had pointed out the claim was false (see Appendix A for the specific language used in each study). To bolster the charge that Wooley's claim was indeed false, the article always closed by describing that the "calls to Wooley ... asking for comments on the charge that Wooley had not been truthful ... were not returned."

Reactions to the Politician's FF. Following exposure to the newspaper article, participants again read Wooley's tweet or statement and answered three questions on 7-point scales. Each was bounded at −3 and +3. First, we determined whether participants came to the politician's defense with a charge that the media rebuke must have been *fake news*: "Given the information you have, how likely do you think it is that [Wooley's] statement is true?" Second, we assessed whether participants thought the politician was *unaware* that he had told a falsehood: "Regardless of whether you think [Wooley's] statement is true, how likely do you think it is that the representative believes his statement is true?" Third, participants evaluated whether the FF was *acceptable*: "Regardless of your answers above, do you think it was justifiable or unjustifiable for [Wooley] to post this tweet?" Each response scale included labels at −3 (*very unlikely to be true/very unlikely to have believed it to be true/very much not justifiable*), 0 (*equally likely to be true and untrue/equally likely to have believed it to be true and untrue/equally justifiable and unjustifiable),* and +3 (*very likely to be true/very likely to have believed it to be true/very much justifiable*).

Political Ideology. To assess participants' political ideology, we used a combination of global and (issue-)specific measures. In all studies, participants answered two general questions by identifying themselves along Likert-type scales. One read, "Although some people do not strongly align themselves along this single continuum, please do your best to determine where you fit." The 7-point scale was anchored at −3 (*staunchly Democrat*) and +3 (*staunchly Republican*), with the midpoint 0 labeled "*equal Democrat and Republican*." A second asked, "How would you describe your political views?" Participants selected one of seven categories, presented in this order: *extremely conservative, very conservative, somewhat conservative, neither liberal nor conservative, somewhat liberal, very liberal, extremely liberal.*

Next, participants indicated their stance on seven policy issues on a scale from −3 (*totally opposed*) to +3 (*totally supportive*), with 0 labeled "*neutral.*" One issue was relevant to the FF in each particular study: "policies that aim to limit immigration of any kind into the country" (Study 1a), "policies that favor implementing a $15/hr minimum wage" (Studies 1b), "policies that would raise the minimum wage" (Study 1c), or "policies that support school vouchers" (Study 1d). We scored each scale so that it was centered at 0 and higher numbers reflected a more conservative or Republican-leaning position. Given all items used 7-point response scales, we created a composite by averaging the two general and one relevant specific item (all αs > .72). To norm these composites, we divided this average by the standard deviation. Note that this standardized participant *political orientation* composite preserves the individual items' meaningful 0, which reflects neutrality on the standard left–right dimension.

Results

We conduct analyses that combine across Studies 1a–1d, though see the Supplemental Materials for disaggregated results. All analyses include a random effect of study, to account for the nonindependence of errors for responses that come from the same study. Each model includes three fixed-effects predictors: the politician's *party* (+1: Republican, −1: Democratic), the participant's *political orientation* (centered at political neutrality), and the Party × Political orientation interaction. In presenting our results, we refer to our model's predictions for "conservatives" and "liberals": those who are +1 or −1, respectively, on our political orientation composite. This allows us to roughly equate the groups in terms of their distance from our meaningful, neutral 0. Predicting outcomes at deviations from the sample mean (as would be the case if, for example, we *z*-scored the individual items) would leave our labels and results at the whims of the sample's average political orientation. Although we proceed to describe key results below, key predicted means appear in Figure 1, and the full regression output appears in Table 1.

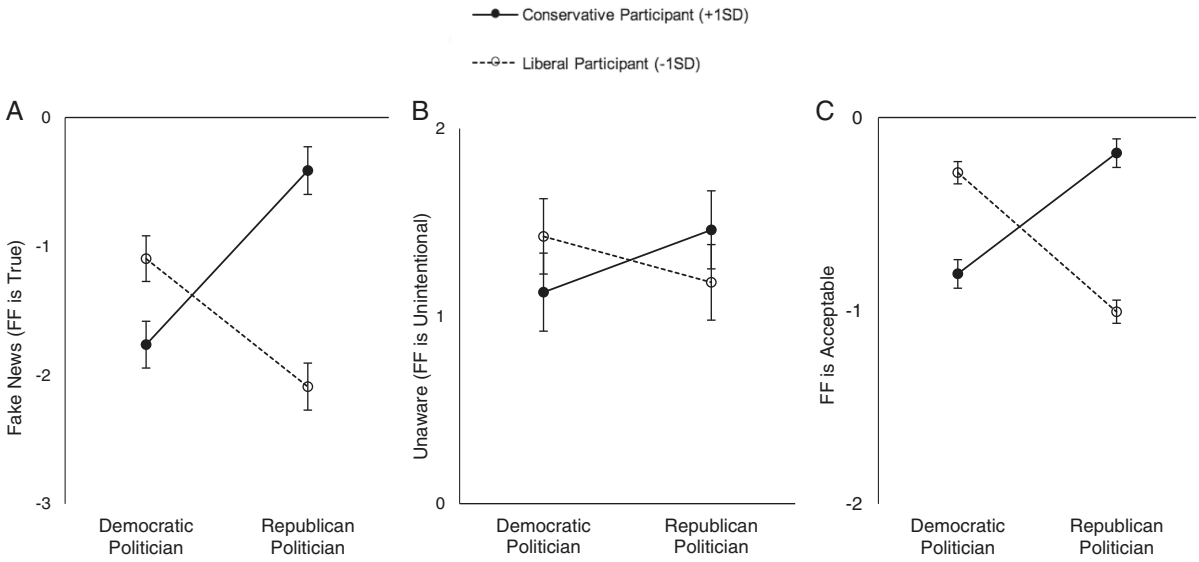
After examining effects on the most basic excuse for a FF (fake news), we control for this perception in examining the second excuse (unaware). Analyses for the final measure (FF is acceptable) control for both fake news and unaware. In this way, we examine variability in perceptions of FF acceptability that cannot be attributed to variation in the first two forms of excuse-making (and the partisan biases that may characterize them). In all cases, we standardize the covariate(s). The covariates are meaningful in how our results are interpreted in two ways.

First, the mean value of each covariate is useful in considering at what level an effect is estimated (Table SM19). In every study in which these two measures were included (Studies 1–4), the average participant reported that the politician had in fact told a falsehood but directionally (even if not always statistically significantly) thought the politician was unaware of stating a falsehood. In the General Discussion, we return to all studies to test whether evidence of partisan evaluations of policy FFs emerges similarly regardless of the participants' standing on the two covariates. To foreshadow, we will find that this is the case.

Second, the inclusion of the covariate(s) means that any effects of partisan leniency (Party × Political orientation) on one excuse are partialed out before examining the next excuse. In essence, this allows us to avoid double (or triple) counting the effects of partisan biases (i.e., the extent to which responses are a function of a match between the participant and the target). This helps us to avoid the problem of attributing effects to partisanship that are actually explained by correlated inputs to such evaluations (Druckman & McGrath, 2019; Pennycook & Rand, 2021). That is, if likeminded partisans display a partisan bias in determining which politicians are telling a falsehood at all, then inclusion of the fake news covariate allows us to control for such variability (and the partisan bias that explains it) in estimating whether people are likely to excuse their own side's politicians as being unaware of making a false statement. Note that as a result, the models estimate the judgments of liberals and conservatives as being more similar than the covariate-free, raw means reflect. For example, averaged across our studies, participants who were likeminded partisans tended to see FFs as slightly more justifiable than not (an important finding in its own right in capturing their friendly responses to these FFs), though the model output depicted below will identify likeminded partisans as seeing FFs as

Figure 1

Excuses for FFs by Participant Political Orientation and the Politician Political Party (Studies 1a–1d)



Note. FFs = flagged falsehoods; *SD* = standard deviation. Depicted values are predicted means (and standard errors) as a function of the party of the politician and participants' political orientation (± 1 *SD* from political neutrality) in Studies 1a–1d for (A) fake news (FF is actually true), (B) unaware (FF is unintentional), and (C) FF acceptability. Means are predicted at the overall average response for the dependent variable or variables predicted in the earlier panel or panels.

mildly unacceptable.[4] Furthermore, we report in the Supplemental Materials the zero-order correlations among the two forms of excuse-making and FF acceptability (Table SM21). VIF statistics suggest our results are not threatened by multicollinearity (Table SM22).

Fake News and Unaware

Full analyses of these excuses (fake news and unaware) in the Supplemental Materials. Of note, however, for fake news, we observed a significant Party × Political orientation interaction, $B = 0.59$, 95% CI [0.52, 0.65], $t(1832.26) = 16.91$, $p < .001$. Likewise, for unaware, when controlling for fake news, we also observe a significant Party × Political orientation interaction, $B = 0.14$, 95% CI [0.07, 0.22], $t(1831.071) = 3.86$, $p < .001$. As can be seen in Figure 1 (Panels A and B), there was a partisan bent to this excuse-making. Though these results may be of interest in their own right as demonstrations of partisan excuse-making, these measures' primary functions are to serve as covariates in the next set of analyses.[5]

FF Acceptability

Do people merely show partisan patterns in explaining away their politicians' FFs as inaccurately implicated or at least as unintended sins, or—above and beyond partisan differences in this excuse-making—are people actually more likely to feel their own side's FFs are more justified? We included both previously identified excuses—fake news and unaware—as standardized covariates in this model. Unsurprisingly, each excuse predicted FF acceptability (fake news: $B = 0.83$, 95% CI [0.76, 0.91], $t(1830.86) = 22.25$, $p < .001$; unaware: $B = 0.36$, 95% CI [0.29, 0.43], $t(1830.28) = 10.47$, $p < .001$). Of more central interest was whether there were partisan differences in the perceived acceptability of issuing FFs that could not be traced back to these two (partisan) excuses. Although we did not observe a main effect of party, $B = -0.02$, 95% CI [−0.09, 0.04], $t < 1$, we did see a main effect of political orientation, $B = 0.07$, 95% CI [0.01, 0.14], $t(1832.58) = 2.29$, $p = .022$. Of more central interest, we also identified a significant Party × Political orientation

[4] Consider further why predicting means at the sample average of each covariate understates how much likeminded partisans find their own politicians' FFs to be acceptable. Because likeminded participants are more likely to excuse political FFs through claims of fake news or skepticism that the politician knowingly lied, predicting acceptability judgments for someone at the sample average of the two covariates (thereby estimating a conservative and liberal participant's responses if they interpret the statement's truth and intentionality as the average person would) in effect shrinks the partisan gap and thus understates just how acceptable likeminded partisans judge the politicians' lies to be. To illustrate this and to give a more direct sense of partisans' comfort with their own side's lies, we provide a table of predicted means from models that do not include the two other excuses for lying covariates when predicting FF acceptability in the Supplemental Materials for Studies 1–5 (Table SM18). Still, the reader should be mindful that when we say likeminded partisans excuse FFs as more acceptable, a more nuanced characterization is that opposing partisans see the lies as unacceptable whereas likeminded partisans see the lies as neither acceptable nor unacceptable.

[5] Studies 1–5 permit tests of partisan biases in excusing policy FFs as actually true (fake news) and as likely believed to be true by the target (unaware). Although complete reporting of these results can be found in the Supplemental Materials, similar partisan effects emerged on fake news in all five studies (and in Studies 1a, 1b, 1c, 1d, 2a, and 2b, individually) and in three of five studies on unaware (controlling for fake news). In the two cases in which the partisan effect did not emerge, it did without controlling for fake news. This suggests that partisan effects on unaware are sometimes explained by partisan differences in believing the statement was a falsehood.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 1
Estimated Betas (Standard Errors) From Models for Studies 1a–1d

Predictors	Fake news (FF is true)	Unaware (FF is unintentional)	FF is acceptable
Party (+1: Republican, −1: Democratic)	0.09 (0.04)*	0.02 (0.04)	−0.02 (0.03)
Political orientation	0.25 (0.04)***	−0.00 (0.04)	0.07 (0.03)*
Party × Political orientation	0.59 (0.03)***	0.14 (0.04)***	0.34 (0.03)***
Fake news	—	0.41 (0.04)***	0.83 (0.04)***
Unaware	—	—	0.36 (0.03)***

Note. FF = flagged falsehood. All models also include a random effect of study (1a, 1b, 1c, or 1d). All row variables are simultaneous predictors of the variable that is the column header.
* $p < .05$. *** $p < .001$.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

interaction, $B = 0.34$, 95% CI [0.27, 0.40], $t(1830.25) = 9.82$, $p < .001$.

When considering a Republican, conservative participants found his FFs more acceptable ($M = -0.13$) than did liberals ($M = -0.94$), $t(1832.85) = 8.52$, $p < .001$. But in judging a Democrat, liberal participants found his FFs more acceptable ($M = -0.22$) than did conservatives ($M = -0.75$), $t(1832.42) = -5.66$, $p < .001$. The main effect of political orientation suggested that liberals were more likely to see FFs as unjustifiable. Although we report the (stronger) covariate-free results in the Supplemental Materials for completeness, such analyses have less value given they, in part, reflect the partisan bias in excuse-making (on fake news and unaware) that was reported above, excuses that themselves would make the FF more acceptable. Regardless, readers may want to consult these unadjusted results to see how all three mechanisms (i.e., partisan biases in the two forms of excuse-making that inform but are further twisted in a partisan way to arrive the acceptability judgments) combine to produce especially divergent perspectives between liberal and conservative participants. We follow this same reporting procedure for all future studies and report all unadjusted results in the Supplemental Materials.

These findings converge in suggesting strong partisan differences in how people explain and evaluate politicians' policy-focused FFs. We found weak evidence that conservative participants express more skepticism toward the media's identifications of FFs (fake news; Supplemental Materials) and see FFs as more justifiable (acceptable). We found strong evidence that conservatives and liberals alike come to their own politicians' defense, seeing their policy FFs as: less likely to be false, more likely to be delivered from a belief of truthfulness, and more acceptable (even controlling for these prior assessments, which were subject to their own partisan biases).

Studies 2a and 2b

We have argued not that there is similar partisan leniency in judging the acceptability of all flagged falsehoods, but that FFs that signal allegiance to the defining goals of one's political group are especially likely to get this gentler treatment. Studies 2a and 2b thus test whether this partisan charitability is extended less willingly to autobiographical FFs about politicians' own pasts—that is, personal FFs—than to FFs that advocate for policy positions. We extend our investigation by testing for reactions to policy and personal FFs that relate to new issues: gun control (Study 2a) and affirmative action (Study 2b). Of course, if our reasoning is wrong, and Study 1's findings merely reflect unconditional charitability toward one's ingroup, then the nature of the FF should not moderate the degree to which we observe partisan evaluations of FF acceptability.

Method

Participants and Design

American participants were recruited from Amazon's Mechanical Turk participant panel ($N_{2a} = 400$; $N_{2b} = 401$). Participants were randomly assigned to one of four conditions in a 2 (party: Democrat or Republican) × 2 (FF: policy or personal) full-factorial design. As before, we excluded participants with duplicate IP addresses ($n_{2a} = 12$; $n_{2b} = 15$). Next, and as in all studies, we removed participants who answered the two memory-based attention checks incorrectly ($n_{2a} = 17$; $n_{2b} = 11$). This left sample sizes of 371 (Study 2a) and 375 (Study 2b).

Procedure

The basic procedure followed the same structure as that of Studies 1a–1d, but included the following key change: The state representative tweeted about a reason to support or oppose a policy position (policy FF) or a detail from his own past (personal FF). Policy FFs provided a false fact that helped make the case for a particular policy position, whereas the personal FFs provided a false autobiographical detail that suggested the politician had a personal connection to the issue. To bolster the article's claim that the tweet was indeed false, the article always closed by describing that the "calls to Wooley ... asking for comments on the charge that Wooley had not been truthful ... were not returned."

The measures took a similar form to those used in Studies 1a–1d. Participants began by responding to the two excuses for the FF (fake news, unaware) before offering their opinion on the FF's acceptability. Then, participants indicated their own political ideology on the two general scales and seven issue-specific ones. The relevant items were "all citizens' right to bear arms" (Study 2a) and "affirmative action policies" (Study 2b). As before, we created a three-item composite for each participant so that 0 reflected absolute neutrality and higher values reflected a more conservative or Republican-leaning position (both αs $> .78$).

Results and Discussion

Much as in Studies 1a–1d, we conducted a single analysis that combined across Studies 2a and 2b. (See Supplemental Materials for

disaggregated results.) Each model includes the following fixed-effects predictors: the politician's *party* (+1: Republican, −1: Democratic), the participant's *political orientation* (centered at political neutrality), and the type of *FF* (+1: policy, −1: personal). The three possible two-way and one three-way interaction terms were included as well. We treated study as a random factor. The full model output is in Table 2, and the adjusted means (whose interpretation are subject to the caveat described in Footnote 4) for FF acceptability are displayed in Figure 2. To streamline presentation, we report only in the Supplemental Materials the analyses on the two excuses for lying (fake news and unaware). These two measures are included as (standardized) covariates in the focal analyses that follow:

We aimed to understand when FFs are seen to be more or less acceptable. We observed effects of the two excuses for lying: fake news, $B = 0.90$, 95% CI [0.79, 1.00], $t(736) = 16.37$, $p < .001$; unaware, $B = 0.42$, 95% CI [0.30, 0.53], $t(736) = 7.32$, $p < .001$. Their inclusion allows us to test for partisan effects on perceptions of FF acceptability independent of what is explained by variation in the two forms of excuse-making. And indeed, we observe a Party × Political orientation × FF interaction, $B = 0.16$, 95% CI [0.07, 0.26], $t(736) = 3.50$, $p < .001$. This provides initial evidence that partisan biases in FF acceptability do not simply reflect an in-group bias. We decompose this interaction by examining policy and personal FFs separately.

Policy FFs

In predicting the perceived acceptability of policy FFs, we observe a marginal main effect of party, $B = 0.14$, 95% CI [−0.00, 0.29], $t(736) = 1.95$, $p = .052$, and a significant effect of political orientation, $B = 0.21$, 95% CI [0.07, 0.34], $t(736) = 3.07$, $p = .002$. Of more key import, we find a Party × Political orientation interaction, $B = .42$, 95% CI [0.29, 0.56], $t(736) = 6.06$, $p < .001$. When considering a Republican politician's FFs, conservative participants ($M = -0.22$) deemed it more acceptable than did liberal participants ($M = -1.45$), $t(736) = 5.97$, $p < .001$. But when evaluating a Democratic politician's FFs, liberals saw it as more acceptable ($M = -0.92$) than did conservatives ($M = -1.38$), $t(736) = -2.50$, $p = .013$.

Personal FFs

This partisan bias did not extend to personal FFs. We did not observe main effects of party, $B = -0.03$, 95% CI [−0.16, 0.10], $t < 1$, or political orientation, $B = -0.06$, 95% CI [−0.19, 0.06], $t(736) = -1.01$, $p = .314$. Furthermore, the critical Party × Political orientation interaction failed to reach significance, $B = 0.09$, 95% CI [−0.03, 0.22], $t(736) = 1.47$, $p = .143$. This reflects that partisan leniency was applied to group-serving FFs designed to advance a party's policy agenda, but not those written merely to exaggerate a politician's own connection to an issue. Stated differently, partisan effects in seeing FFs as more acceptable excused displays of loyalty to a partisan agenda, not to all of a party's members.

Study SM1: Conceptual Replication, Distinguishing Mechanistic Accounts

We completed a study that: (a) used a new domain (minimum wage hikes and unemployment) to conceptually replicate the finding that partisan evaluations emerge clearly for policy but not personal FF and (b) provided initial evidence that perceived trustworthiness explains these contrasting patterns. The study failed to find support for three alternative mechanisms that could explain why policy FFs in particular invite partisan evaluations of acceptability, that policy FFs (more than personal FFs): (a) encourage a perception that the other side's politicians actually tell more lies (a *prevalence hypothesis*), (b) will be more effective in accomplishing a legislative goal (an *effectiveness hypothesis*), or (c) will aid with candidates' reelection efforts (a *reelection* hypothesis). The results are reported in full in the Supplemental Materials. The study did not empirically distinguish general from partisan trustworthiness, a differentiation that is theoretically core to our contribution and will be a focus of our remaining three studies.

Study 3

Study 2 demonstrated that partisan evaluations of FF acceptability do not apply to all falsehoods, but to policy FFs in particular. Study SM1 offered initial evidence that perceived trustworthiness explains these contrasting patterns. In Study 3, participants considered one of the FFs used in Study SM1. Those tweets mischaracterized the strength of the connection between the minimum wage and unemployment rates. In addition to completing our standard trio of measures that permit participants to excuse politicians' FFs (fake news and unaware) before commenting on the FF's acceptability,

Table 2
Estimated Betas (Standard Errors) From Models for Studies 2a–2b

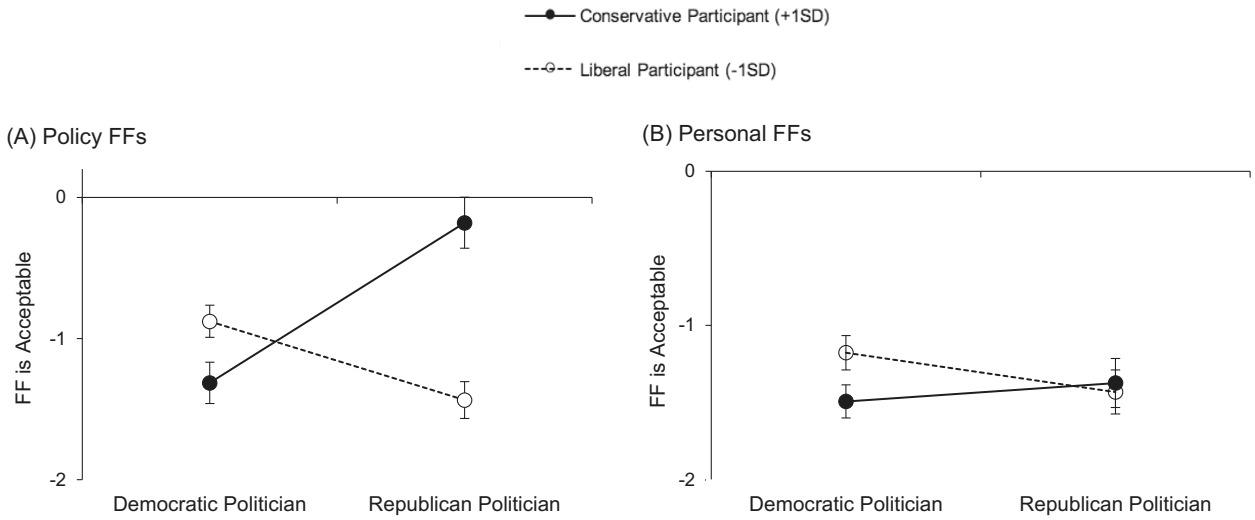
Predictors	Fake news (FF is true)	Unaware (FF is unintentional)	FF is acceptable
FF (+1: policy, −1: personal)	0.11 (0.06)*	0.85 (0.06)***	0.21 (0.06)***
Party (+1: Republican, −1: Democratic)	0.19 (0.06)***	0.06 (0.06)	0.06 (0.05)
Political orientation	0.16 (0.05)***	0.06 (0.06)	0.07 (0.05)
FF × Party	0.08 (0.06)	−0.02 (0.06)	0.09 (0.05)
FF × Political orientation	0.10 (0.05)	−0.04 (0.06)	0.13 (0.05)***
Party × Political orientation	0.37 (0.05)***	0.03 (0.06)	0.26 (0.05)***
FF × Party × Political orientation	0.30 (0.05)***	−0.02 (0.06)	0.16 (0.05)***
Fake news	—	0.69 (0.06)***	0.90 (0.05)***
Unaware	—	—	0.42 (0.06)***

Note. FF = flagged falsehood. All models also include a random effect of study (2a or 2b). All row variables are simultaneous predictors of the variable that is the column header.
* $p < .05$. *** $p < .001$.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 2

Perceived Acceptability of (A) Policy FFs and (B) Personal FFs by Participant Political Orientation and Politician Political Party (Studies 2a–2b)



Note. FFs = flagged falsehoods; *SD* = standard deviation. Depicted values are predicted means (and standard errors) as a function of the party of the politician and participants' political orientation (± 1 *SD* from political neutrality). Means are predicted at the overall average response for the fake news and unaware measures.

participants completed two new measures. This allowed Study 3 to build on our earlier results in two primary ways.

First, we empirically differentiate general trustworthiness and partisan trustworthiness. We asked whether Democrats and, separately, Republicans could trust the politician. A generally trustworthy politician can be trusted by people—Democrats and Republicans—more generally. Instead, a partisan trustworthy politician can be trusted by his own party more than the other party. These measures thus permit distinguishing two pathways by which perceived trustworthiness may relate to FF acceptability. By one hypothesis, participants display a partisan bias in evaluating which politicians are generally trustworthy. But by a separate hypothesis that follows from our reasoning in the Introduction, perceived partisan trustworthiness may have different moral meaning to the politician's likeminded (vs. opposing) partisans, thereby explaining the partisan bent to judgments of FF acceptability.

Second, we developed a measure of whether the politician was seen as a generally moral actor, one who is likely to take moral actions and avoid immoral ones in his day-to-day life. We used two rounds of pretesting to identify prototypical everyday moral and immoral behaviors. Participants forecast how likely the politician was to engage in each when provided the opportunity. This allows us to probe the breadth of the partisan bias we have identified. That is, one possibility is that the partisan bias is narrow, applying specifically to how political partisans judge political FFs. By this hypothesis, partisans may not be too bothered by the FFs themselves (explaining why they are judged to be less unacceptable), but may then take a harsher view of the politician's moral character outside of the political arena. But a competing hypothesis is that the same psychology that explains who excuses certain political FFs may encourage more fundamental shifts in how that politician is viewed—that is, as a better or worse moral actor more generally. After examining the role of perceived (general and partisan) trustworthiness in evaluations of FF acceptability, we then proceed to test whether the same partisan biases extend to forecasts of the politician's (nonpolitical) moral and immoral behaviors.

Method

Participants and Design

We recruited 756 Americans from Amazon's Mechanical Turk. These participants were randomly assigned to one of two *party* conditions; participants learned about either a Democratic or a Republican politician whom the media had identified as issuing a policy-related FF. We excluded from further consideration the 43 participants who missed both memory-based attention checks. This left 713 participants in the analyses reported below.

Procedure

Participants considered one of two policy FFs. In each case, a politician tweeted that minimum wage increases always produced a consistent effect on unemployment. Each exaggerated when tweeting that minimum wage increases *always* reduced (Democrat) or increased (Republican) unemployment. Next, participants completed our standard trio of measures that permitted: denying the media report that the tweet was false (fake news), claiming the politician believed he was telling the truth (unaware), and labeling the FF as justified (acceptable). After completing two new measures (described below), participants indicated their own political orientation using the same measures administered in Study SM1 ($\alpha = .81$).

General and Partisan Trustworthiness. Participants indicated whether "Wooley is the type of person Democrats can trust" and someone "Republicans can trust." Responses were offered 7-point scales anchored at 1 (*not at all*) and 7 (*completely*). By

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

summing the two measures, we created an index of *general trustworthiness* ($M = 6.60$, $SD = 3.05$). That is, a candidate who was seen as someone whom both Democrats and Republicans could trust would score the highest on this composite. We also created a difference score by taking perceptions of how much Wooley's own party could trust him and subtracting how much the other party could trust him. This *partisan trustworthiness* composite is maximized when Wooley is seen as someone whom his own side can trust but whom the other side cannot ($M = 0.94$, $SD = 2.24$). Perceptions of general and partisan trustworthiness were only marginally significantly correlated, $r(711) = .07$, $p = .084$, and thus were largely independent.

Moral Behaviors. We developed a new measure to help us understand how participants viewed the target's general moral character, outside of the political realm. To avoid experimenter stimulus-selection biases, we conducted two pretests to identify a set of everyday moral (5) and immoral (5) behaviors. The first pretest asked participants ($N = 93$) to list five everyday moral and five everyday immoral acts. The second pretest showed participants ($N = 94$) the 15 mostly frequently listed behaviors from each category and had them rate how representative they were of moral and immoral behaviors. Through this process (described in greater detail in the Supplemental Materials), we selected the five most representative moral behaviors (return a lost item, help a stranger retrieve dropped possessions, help someone cross the street, return excess change to a cashier, and give up a seat so family members can sit together) and the five most representative immoral behaviors (knowingly lie on tax returns, pretend not to hear someone calling for help, make fun of someone in front of other people, make a racist joke, share a secret that one was asked to keep). Crucially, these moral and immoral behaviors are not described as directed toward the political in-group or out-group. In this way, the measure assesses to what extent the politician is seen as a moral exemplar more generally, not simply someone who tends to display partiality toward or concern for his own party.

In the main study, we asked participants, "In your best estimate, what percentage chance is there that Rep. Wooley will do each of the behaviors in the context or circumstance described?" Participants provided a percentage response for each behavioral forecast. In general, perceptions that the target would perform one moral behavior predicted perceptions that the target would perform other moral behaviors ($\alpha = .92$). The same was true of immoral behaviors ($\alpha = .89$). Furthermore, the more the target was believed to perform moral behaviors, the less he was seen as likely to perform immoral ones, $r(711) = -.36$, $p < .001$. As such, we made a single *moral behavior* composite by taking the average forecast for the moral behaviors and subtracting the average forecast for the immoral behaviors ($M = 11.19\%$, $SD = 42.17\%$).

Results and Discussion

We proceed in three parts. First, we attempt to replicate our central finding, that (controlling for the two excuses that are susceptible to partisan biases themselves) there is a partisan bent to the perceived acceptability of FFs. Second, we disentangle whether it is perceptions of the candidate's general or partisan trustworthiness that help to explain why participants see like-minded politicians' FFs as more acceptable. Third, we probe this trustworthiness distinction more deeply by examining how general and partisan trustworthiness predict perceptions of the politician as a moral person more generally, even outside of political contexts.

FF Excuses and Acceptability

Given that all participants considered a policy FF, we returned to our simplified model used in Studies 1a–1d. Because we report the results of a single study, we omitted the random effect of study and again tested whether people thought the FF was relatively justifiable while including both of the first two excuses—fake news and unaware—as standardized covariates in this model. (Full results showing partisan effects on both forms of excuse-making are presented in the Supplemental Materials.) Doing so, we found a partisan bias in characterizations of the FF as acceptable, $B = 0.33$, 95% CI [0.22, 0.44], $t(707) = 6.05$, $p < .001$ (see Figure 3). In short, we again find that participants more sympathetically evaluated FFs from politicians of their own party.

Trustworthiness

We begin by asking (a) whether there is a partisan bias in evaluations of general trustworthiness (how much Democrats + Republicans can trust him) and/or (b) whether the perception that the lying politician can be trusted more by his own party than by the opposing party is connected in a partisan way (i.e., differently by members of his own vs. the other party) to a sense that the candidate's behavior was acceptable. Either route (or both routes) could produce the partisan evaluations of FF acceptability that we have consistently observed.

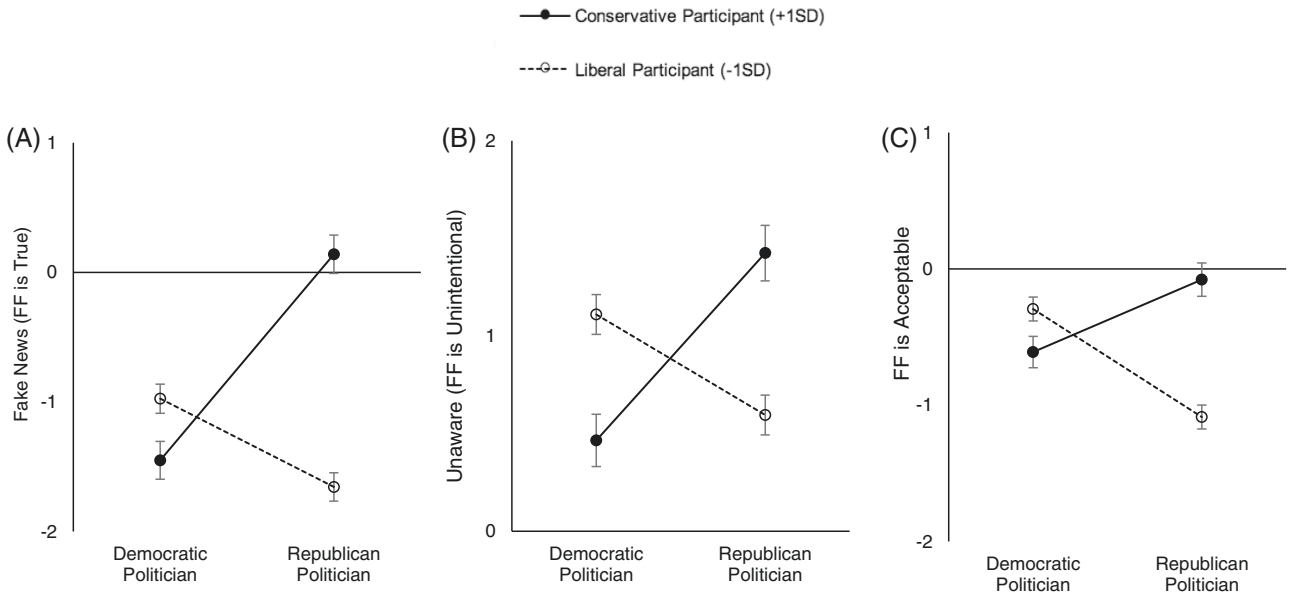
General Trustworthiness. We used the same model as above that predicted FF acceptability to predict general trustworthiness, the sum of perceptions that Democrats and Republicans can trust the politician. As a reminder, this model controls for beliefs that the tweet was actually true and that the FF was perhaps unintentional. We observed a main effect of party, $B = -0.33$, 95% CI [−0.54, −0.13], $t(707) = 3.21$, $p = .001$: The Republican politician ($M = 6.23$) was seen as less generally trustworthy than the Democratic one ($M = 6.90$). We did not, however, observe a significant Party × Political orientation interaction, $B = 0.19$, 95% CI [−0.02, 0.40], $t(707) = 1.80$, $p = .071$. In other words, participants' own political orientation did not significantly relate to their perceptions of the politician as generally trustworthy. Of course, this effect was marginally significant, but this can be contrasted with our findings on FF acceptability, for which a strong partisan bias was observed. This suggests that partisan biases in FF acceptability may instead be more rooted in perceptions of partisan trustworthiness and how those perceptions are connected with FF acceptability. We now turn to those tests.

Partisan Trustworthiness. We started by testing what seemed straightforward and intuitive, whether Rep. Wooley was seen as being more trustworthy to his own party than to the opposing party. We used the same model as that predicting general trustworthiness, but predicted partisan trustworthiness instead. The partisan trustworthiness composite has a meaningful zero: It reflects being equally trustworthy to both members of his political in-group and out-group. For this reason, we were interested in this first model's intercept.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 3

Excuses for FFs by Participant Political Orientation and Politician Political Party (Study 3)



Note. FFs = flagged falsehood; *SD* = standard deviation. Depicted values are predicted means (and standard errors) as a function of the party of the politician and participants' political orientation (± 1 *SD* from political neutrality) in Study 3 for (A) fake news (FF is actually true), (B) unaware (FF is unintentional), and (C) FF acceptability. Means are predicted at the overall average response for the dependent variable or variables predicted in the earlier panel or panels.

The intercept was significantly greater than 0, $B = 0.87$, 95% CI [0.70, 1.04], $t(707) = 10.07$, $p < .001$. Wooley was seen as one who could be trusted more by members of his own party than by members of the opposing party. Furthermore—much like with perceptions of general trustworthiness—this perception of partisan trustworthiness did not vary as a function of participants' own political orientation: The Party × Political orientation interaction did not approach significance, $B = -0.10$, 95% CI [−0.28, 0.08], $t(707) = 1.12$, $p = .262$. Combined, these analyses demonstrate that there was broad (bipartisan) agreement that Wooley can be trusted by members of his own party more than by members of the other party—that is, that he is partisan trustworthy. But next we ask whether partisan trustworthiness operates differently as a cue to FF acceptability when evaluating an in-group politician as opposed to an out-group one. If so, this could offer insight into why there are partisan biases in perceived FF acceptability.

Connecting Trustworthiness to FF Acceptability. We return to our model predicting FF acceptability but add several additional predictors. We include both general trustworthiness and partisan trustworthiness (each standardized). But because we thought that the connection between trustworthiness and FF acceptability may differ for those considering a politician of their own political stripes, we also permit each trustworthiness measure to interact with party, political orientation, and the interaction of these two variables (what we have called partisan bias). Although we will guide readers through the most relevant findings, the results of the full model are detailed in Table 3.

Let us begin with the connection between general trustworthiness and FF acceptability. We observe a main effect of general trustworthiness, $B = 0.54$, 95% CI [0.43, 0.66], $t(699) = 9.28$, $p < .001$. This reflects that the more generally trustworthy the politician was seen to be, the more acceptable participants found his FF. Just as there was not a significant partisan bias in judgments of general trustworthiness, there was also no partisan bias in the connection between general trustworthiness and FF acceptability. That is, the General trustworthiness × Party × Political orientation interaction did not approach significance, $B = -0.04$, 95% CI [−0.13, 0.06], $t(699) = .76$, $p = .445$. This shows that targets' general trustworthiness does relate to the perceived acceptability of their FFs, but also that general trustworthiness will not help to make sense of the partisan bias in FF acceptability. As we turn our attention to partisan trustworthiness, keep in mind that effects observed below are those that exist independently of (or above and beyond) the nonpartisan effects of general trustworthiness just discussed.

Turning to the effects of partisan trustworthiness, we did not in this case observe a main effect, $B = -0.04$, 95% CI [−0.15, 0.06], $t(699) = .81$, $p = .417$, but we did observe a Partisan trustworthiness × Party × Political orientation interaction, $B = 0.17$, 95% CI [0.07, 0.26], $t(699) = 3.54$, $p < .001$ (see Figure 4). This reflects a partisan bias in the connection between trustworthiness and FF acceptability. Those who saw the politician as especially trustworthy to his own political party at the expense of the other (+1 *SD* partisan trustworthiness) showed the partisan bias (Party × Political orientation interaction) in evaluations of FF acceptability, $B = 0.43$, 95% CI [−0.31, 0.56], $t(699) = 6.74$, $p < .001$. But for those who did not assume greater trustworthiness to the target's own party (−1 *SD* partisan trustworthiness), this partisan bias evaporated, $B = 0.10$, 95% CI [−.05, .25], $t(699) = 1.34$, $p = .179$. In other words, everyone sees politicians' political FFs as more acceptable to the extent that the politician seems generally trustworthy, but partisans

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 3
Estimated Betas (Standard Errors) From Models for Study 3

Predictors	FF is acceptable	Moral behavior composite
Party (+1: Republican, −1: Democratic)	0.01 (0.05)	−4.70 (1.41)***
Political orientation (conservative, higher)	0.27 (0.05)***	2.69 (1.42)
Party × Political orientation	0.27 (0.05)***	13.07 (1.45)***
Belief FF is true (standardized)	0.68 (0.06)***	−0.42 (1.63)
Belief speaker thought FF was true (standardized)	0.37 (0.05)***	8.79 (1.40)***
General trustworthiness (standardized)	**0.54 (0.06)*** **	**6.54 (1.58)*** **
General trustworthiness × Party	−0.16 (0.05)**	1.81 (1.40)
General trustworthiness × Political orientation	−0.01 (0.05)	−1.09 (1.34)
General trustworthiness × Party × Political orientation	−0.04 (0.05)	0.14 (1.33)
Partisan trustworthiness (standardized)	−0.04 (0.05)	2.90 (1.44)
Partisan trustworthiness × Party	−0.02 (0.05)	−3.66 (1.42)*
Partisan trustworthiness × Political orientation	−0.10 (0.05)*	0.96 (1.26)
Partisan trustworthiness × Party × **Political orientation**	**0.17 (0.05)*** **	**4.35 (1.27)*** **

Note. FF = flagged falsehood. The key terms discussed in the main text are in bold. Values represent regression betas, and values in parentheses represent standard errors.
* $p < .05$. ** $p < .01$. *** $p < .001$.

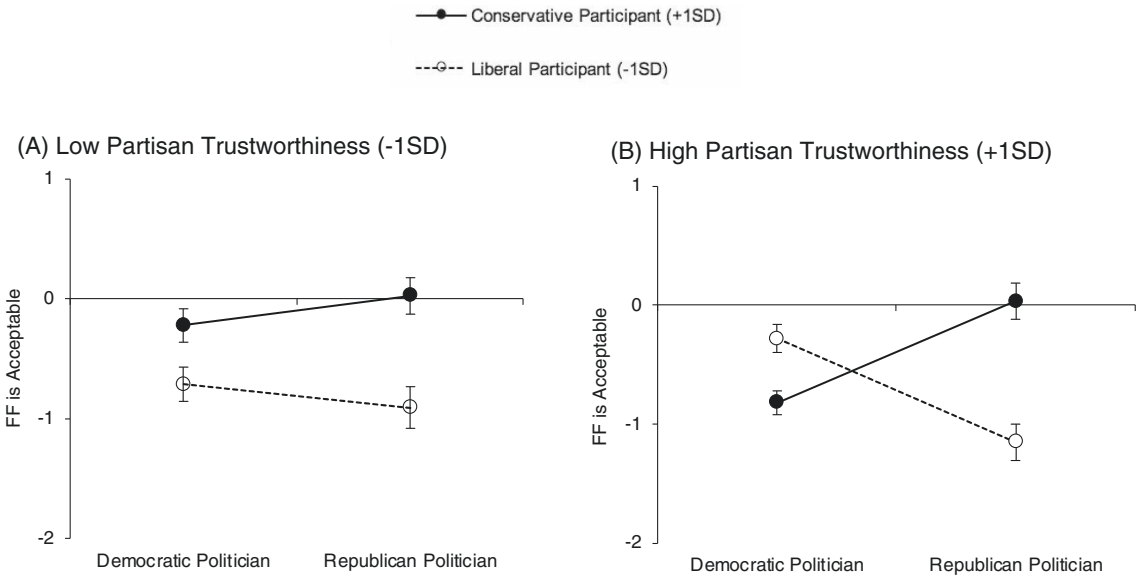
also see FFs as more acceptable to the extent that politicians can be trusted by perceivers' own party and not the other political side.

Moral Behaviors

Our next analyses distinguish two possibilities. One is that we have been examining the narrow effect of excusing one's own side's political misdeeds (misdeeds that are at least consistent with one's own political ideology). The second is we have been studying a more general effect, that partisan trustworthiness is a divergent cue to the morality of others' actions and character more generally. If the latter, we should see evidence that the partisan bias in evaluating political FFs is one instantiation of a broader phenomenon by which cues to partisan trustworthiness take on different moral meaning depending on the political allegiances of the perceiver. To explore this question, we ran the same model but predicted the moral behavior composite, the forecast that the politician would engage in more prototypically good and less prototypically bad (nonpolitical) behaviors in his daily life.

We begin with general trustworthiness, which our above analyses found was neither assessed nor leaned upon in a partisan way. We observed a main effect of general trustworthiness, $B = 6.54$, 95%

Figure 4
FF Acceptability by Politician Party, Participant Political Orientation, and Perception of the Politician as (A) Relatively High (+1 SD) or (B) Low (−1 SD) in Partisan Trustworthiness (Study 3)



Note. FF = flagged falsehood; *SD* = standard deviation. Depicted values are predicted means (and standard errors) as a function of the party of the politician and participants' political orientation (± 1 *SD* from political neutrality). Means are predicted at the overall average response for the fake news and unaware measures.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

CI [3.44, 9.94], $t(699) = 4.14$, $p < .001$. In a sense, this is something of a validation of our new moral behavior measure: The more that the politician was viewed to be generally trustworthy, the more he was forecast to display prototypically moral (and not immoral) behaviors. Furthermore, this relationship was similar when judging a politician from a participant's own side or the other: The General trustworthiness × Party × Political orientation interaction was not significant, $B = 0.13$, 95% CI [−2.48, 2.75], $t < 1$. To everyone, general trustworthiness means being a generally moral actor. We also observed a main effect of partisan trustworthiness, $B = 2.90$, 95% CI [0.08, 5.72], $t(699) = 2.02$, $p = .044$. But of more central interest, and as depicted in Figure 5, that effect was qualified by partisan bias: The Partisan trustworthiness × Party × Political orientation interaction was significant, $B = 4.35$, 95% CI [1.86, 6.84], $t(699) = 3.43$, $p < .001$. In other words, the more that the politician was seen to be trustworthy to a participant's own political group (as opposed to the opposing one), the more he was seen to be a good person more generally.

These final results—when considered in the broader context of our studies—begin to paint a more complete picture of why the electorate can take such sharply divergent views of politicians who are called out for making false statements. Specifically, *partisan* trustworthiness explains these divergent perceptions of the acceptability of policy FFs—not directly, but because partisan trustworthiness was imbued with different moral meaning by different people. This contrasts with general trustworthiness, which had the same positive, moral significance to everyone. A politician's *partisan* trustworthiness helped to blunt the perceived unacceptability of his FF to the extent that the perceiver and the politician were on the same political team. And as these findings show, partisan trustworthiness is similarly used as a cue of more general moral character. Those who can be trusted by one's own political in-group but not by the political out-group are also those who are assumed to go through life doing more good and less bad.

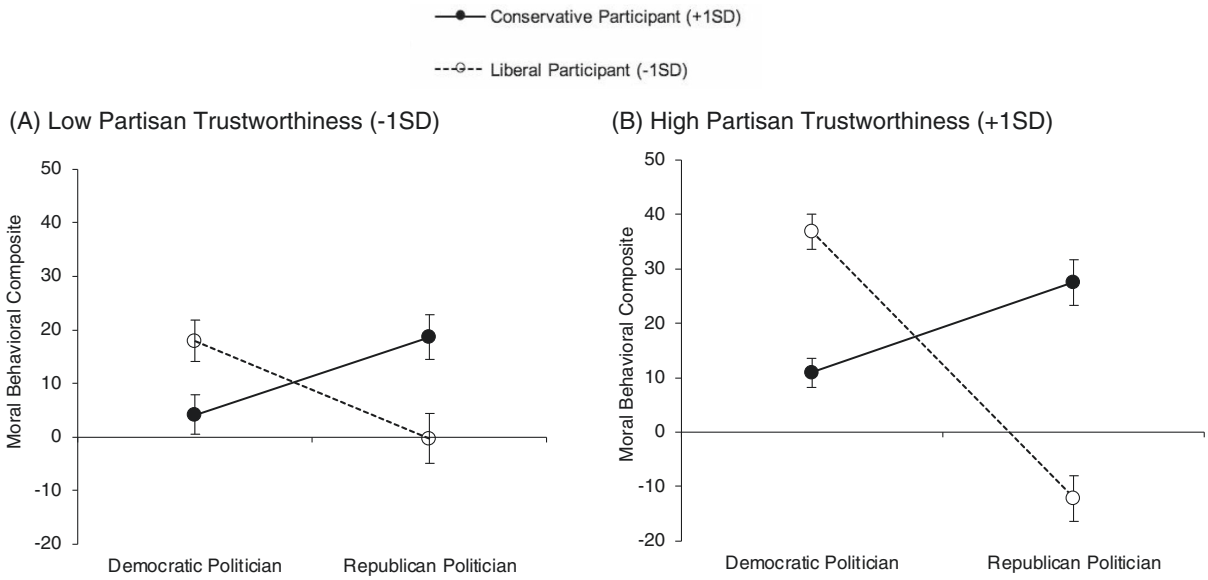
These findings with regard to general and partisan trustworthiness both reinforce and add nuance to previous claims that perceived trustworthiness is crucial to creating expectations about a target's future behavior (Mayer et al., 1995). When the electorate is confronted by a politician who is willing to throw around political policy FFs, they quite literally see a person of two different moral resumés. In this way, two opposing partisans can be in complete agreement regarding the extent to which a policy FF has been intentionally issued, but then diverge sharply in the perceived moral acceptability and implications of such behavior. Loyalty is a positive cue to morality, but only when those loyalties belong to one's own (and not another's) group.

Further Consideration of the Role of Partisan Trustworthiness

We have intentionally avoided making strong claims that our dependent measures are necessarily connected in a single linear causal sequence. After all, we are most interested in these results because of what they reveal about how partisan trustworthiness has different moral connotations to different people. But consider a feature that differentiates perceptions of partisan trustworthiness from perceptions of both FF acceptability and forecasts of moral behaviors: Liberal and conservative participants did not differ in perceptions of a target's partisan trustworthiness. Instead, they seemed to vary only in their views about whether a partisan

Figure 5

Forecasted Moral Behavior by Politician Party, Participant Political Orientation, and Perception of the Politician as Relatively High (+1 SD) or Low (−1 SD) in Partisan Trustworthiness (Study 3)



Note. *SD* = standard deviation. Depicted values are predicted means (and standard errors) as a function of the party of the politician and participants' political orientation (± 1 *SD* from political neutrality). Means are predicted at the overall average response for the fake news and unaware measures.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

trustworthy politician is a good person who should escape special condemnation for his falsehoods. It would be surprising if, for example, liberals and conservatives first drew diametrically opposing conclusions about a FF's acceptability and the politician's moral character, only to then ultimately perfectly reconverge at precisely the same perceptions of the politician's partisan trustworthiness. Instead, a more parsimonious interpretation is that identical perceptions of partisan trustworthiness have different meaning for (and thus have different causal effects on) perceptions of FF acceptability and moral character.

Similarly, consider our findings regarding forecasts of moral behavior. Although we think it is most parsimonious that perceptions of partisan trustworthiness shape our participants' forecasts of another's moral behavior (after all, we doubt participants were actively thinking about forecasts of those 10 behaviors spontaneously), it is certainly possible that when people learn of another's nonpolitical moral or immoral deeds, perceptions of partisan trustworthiness shift in turn. If so, this would reinforce our basic point that moral evaluations relate not only to general trustworthiness but also to a special trustworthiness to one's own group in particular. We leave this question of whether such reverse-causality exists for future research.

A separate question is whether the political FF itself exaggerated perceptions of partisan trustworthiness, or whether it simply failed to undermine such preexisting perceptions. On the one hand, policy FFs may simply not call into question assumptions of partisan trustworthiness that people already hold. On the other hand, policy FFs—as acts that serve to advance one side's partisan agenda—may exaggerate such perceptions. Although the resolution of this question is not necessary to substantiate our central point that partisan trustworthiness holds different moral connotations for different people, Study 5 will begin to address it by testing whether those who embrace a policy FF are seen to be more partisan trustworthy than those who instead aim to truthfully correct the FF.

Finally, we considered whether these results shed better light on our previous findings that policy FFs are treated with more partisan leniency than are personal FFs. The present study suggests that may be because politicians issuing policy FFs are seen to be more partisan trustworthy than politicians issuing personal FFs. We conducted two posttests to examine this possibility in convergent ways. In our first posttest, ($N = 220$ Americans on AMT, 185 after excluding those who failed a memory-based attention check), we described to participants the policy/personal distinction (in a counterbalanced order) by referencing the FFs of each type from Studies 2a, 2b, and SM1. We then asked (within-subjects), "If you knew a politician told a [policy, personal] lie, how much do you think he or she could be trusted by ... : members of his/her own party, members of the other party?" A 2 (falsehood: policy or personal) × 2 (target: own party, other party) interaction suggested that partisan trustworthiness differed in response to a policy and a personal falsehood, $F(1, 184) = 9.21$, $p = .003$, $\eta_p^2 = 0.05$. That is, partisan trustworthiness was inferred to be greater about a politician issuing a policy falsehood ($M_{dif} = 0.81$, $SD = 1.73$) than a personal FF ($M_{dif} = 0.42$, $SD = 1.49$).

In a second posttest (preregistration: https://aspredicted.org/MM5_N26), we did not use the word "lie" but instead asked participants to consider "statements that are flagged (by nonpolitical authoritative sources) as being falsehoods." After reading the same distinction between policy/personal falsehoods, participants ($N =$ 230 Americans on AMT, 229 after excluding one who failed a memory-based attention check) indicated on two 5-point scales what type of politician would tell a policy falsehood and a personal falsehood: −2 = definitely *nonpartisan* politicians, +2 = definitely *partisan* politicians. We defined a "partisan" politician as, "someone who has a reputation as someone who can be trusted by members of his or her own party more than by members of the other side." As predicted, participants believed that policy falsehoods are more likely to be told by partisan politicians ($M = 0.67$, $SD = 1.17$) than are personal falsehoods ($M = 0.24$, $SD = 1.10$), paired $t(229) = 4.63$, $p < .001$, $d = 0.31$.

Study 4

To this point, we have argued that policy FFs—in part because they signal a politician's partisan trustworthiness—are deemed more acceptable by likeminded partisans. Consider further why policy FFs send this signal. They bend the truth in the service of supporting policy goals that are part of what defines a party's purpose and thus a partisan identity. People affiliate with parties because of what those parties represent. A party's agenda is reflected in a party's platform. Supporting that agenda—even, or (as the last study will explore) perhaps especially, with a set of "alternative facts"—displays one's loyalty to one side of a partisan divide.

But one can also think of dishonest statements that could serve to help one's own party, but not by displaying one's unwavering commitment to advancing what one's party explicitly aims to accomplish. For example, a politician could be flagged as telling an *electoral falsehood*, one that tries to meddle in the conduct of a free and democratic election. Although recent events in American history have revealed alarming antidemocratic sentiment in American society, neither the Democratic nor Republican platforms explicitly include a commitment to denying members of the other party knowledge of how to exercise their right to vote. Although the major parties differ in how they balance concerns about election security and voting access, neither party's platform endorses fraud. Even the recent pro-Trump "Stop the Steal" movement was couched in the language—even if unsupported by facts—of concern for election integrity.

Motivated by this reasoning, we had all participants in Study 4 consider a politician who tweeted a FF that had the potential to help his own party. This communication either offered unsubstantiated facts in support of a policy position (policy FF) or provided false information about certain voters' ability to cast their ballot (electoral FF). Whereas the former FF reflects a dishonest approach to supporting the party line, the latter FF reflects a questionable tactic to more directly undermine democracy. In light of Study SM1's finding that partisan differences in FF acceptability cannot be traced to an ends-justify-the-means sort of thinking, we suspected that electoral FFs—given they do not reflect a direct commitment to a party's aims—would not send the same signal of partisan trustworthiness as would policy FFs. As a result, we hypothesized that judgments of the acceptability of policy FFs would be characterized by more partisan leniency than would judgments of electoral FFs, and that these results could be traced to the stronger signal of partisan trustworthiness sent by the policy than the electoral FF.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Method

Participants and Design

American participants were recruited from Amazon's Mechanical Turk participant panel (N = 4,175). Participants were randomly assigned to one of four conditions in a 2 (party: Democrat or Republican) × 2 (FF: policy or electoral) full-factorial design. Following the same standard as in our earlier studies, the 394 participants who responded incorrectly to both memory-based attention checks were excluded. This left 3,781 participants in our analyses reported below.[6]

Procedure

Participants read a newspaper article detailing a tweet from (fictitious) New Mexico State Representative Steven Wooley. In all cases, the tweet was reported to be a falsehood, but the topic of that falsehood varied by condition. For those exposed to a *policy* FF, the tweet was similar to that used in Study 2a. More specifically, the Republican politician stated that higher rates of gun ownership produce less crime, whereas the Democratic politician stated that higher rates of gun ownership produce more crime. The newspaper article cited evidence that there is no relationship between gun ownership and crime.

But for those who saw the *electoral* FF, the politician tweeted false information about voting access. More specifically, the politician—regardless of his political party—was reported to have tweeted inaccurate information on election day regarding the ability to still vote in Downtown Albuquerque precincts. Crucially, those precincts' electorate was said to be disproportionately allegiant to the opposing party. The tweet from the afternoon of the election read, "Lines in Downtown Albuquerque voting precincts are 4 hr long. If you're not already in line, you won't get to vote because polls close at 7." The article then explained that the newspaper's own journalists saw no such lines and that, by law, anyone who was in line to vote prior to 7 p.m. would be able to cast a ballot. In so doing, the article made clear that the tweet was not true, but also explained how it could ultimately help Wooley's own party—not by promoting its policy agenda (the aim of the party-serving FFs in our previous studies), but by attempting to disenfranchise voters from the opposing party. Both FF conditions' articles concluded with the statement that Wooley's office had not responded to media inquiries about the tweet.

Next, participants completed our standard slate of three dependent measures. These allowed participants to deny the conclusion of the news story that the tweet was a falsehood (fake news), to claim that the politician believed himself to be telling the truth (unaware), and to label the FF itself as justifiable (acceptable). Then, participants completed the two trustworthiness measures introduced in Study 3 from which we could calculate general trustworthiness (how much Democrats + Republicans can trust him) and partisan trustworthiness (how much his own party—the other party can trust him). As in that study, the two composites were correlated, but weakly so, $r(3781) = .15$, $p < .001$. Finally, participants completed the same two general political orientation questions ($r = .73$) used in all previous studies. Although participants did complete issue-specific items as well (including one on gun control), we did not use this measure in our political orientation composite given its differential applicability to the policy and electoral FFs.

Results and Discussion

General and Partisan Trustworthiness

To begin, we ask whether the policy FF and electoral FF differ in the extent to which they signal general trustworthiness and partisan trustworthiness. For both, they did. When the politician told a policy FF, he was seen as more generally trustworthy (M = 6.96, SD = 2.91) than when he told an electoral FF (M = 5.38, SD = 3.19), $t(3759.83) = 15.93$, $p < .001$, $d = 0.52$. Furthermore, the politician issuing the policy FF was also seen as more partisan trustworthy (M = 1.34, SD = 2.11) than the politician issuing the electoral FF (M = 0.87, SD = 2.05), $t(3745.41) = 6.78$, $p < .001$, $d = 0.22$. Even when we controlled for the first two excuses for FFs (fake news, unaware), we observed a diminished but still-significant difference on general trustworthiness, $t(3777) = 6.25$, $p < .001$, and a largely unchanged difference on partisan trustworthiness, $t(3777) = 6.18$, $p < .001$. Especially for partisan trustworthiness, the differential signals sent by policy and electoral FFs were not explained by differences in their perceived believability and intentionality.

FF Acceptability

We again tested whether people thought the FF was relatively justifiable while including both of the first two excuses—fake news and unaware—as standardized covariates in this model. (Results of models explaining variation in these two forms of excuse-making are presented in the Supplemental Materials.) If the policy and electoral FFs differ in the extent to which they reflect trustworthiness—and partisan trustworthiness, in particular—then by our logic the two FFs should predictable differ in the magnitude of the partisan evaluations they invite. Not only did we replicate the finding that there was a partisan bias in evaluating the acceptability of FFs (even with fake news and unaware controlled), $B = 0.23$, 95% CI [0.19, 0.27], $t(3771) = 11.05$, $p < .001$, but this effect was further qualified by the type of FF, $B = 0.11$, 95% CI [0.07, 0.15], $t(3771) = 5.21$, $p < .001$ (see Figure 6). For the policy FF, there was a strong partisan bias concerning FF acceptability, $B = 0.34$, 95% CI [0.28, 0.40], $t(3771) = 11.32$, $p < .001$. For the electoral FF, this partisan bias was still present, though significantly diminished, $B = 0.13$, 95% CI [0.07, 0.18], $t(3771) = 4.39$, $p < .001$.

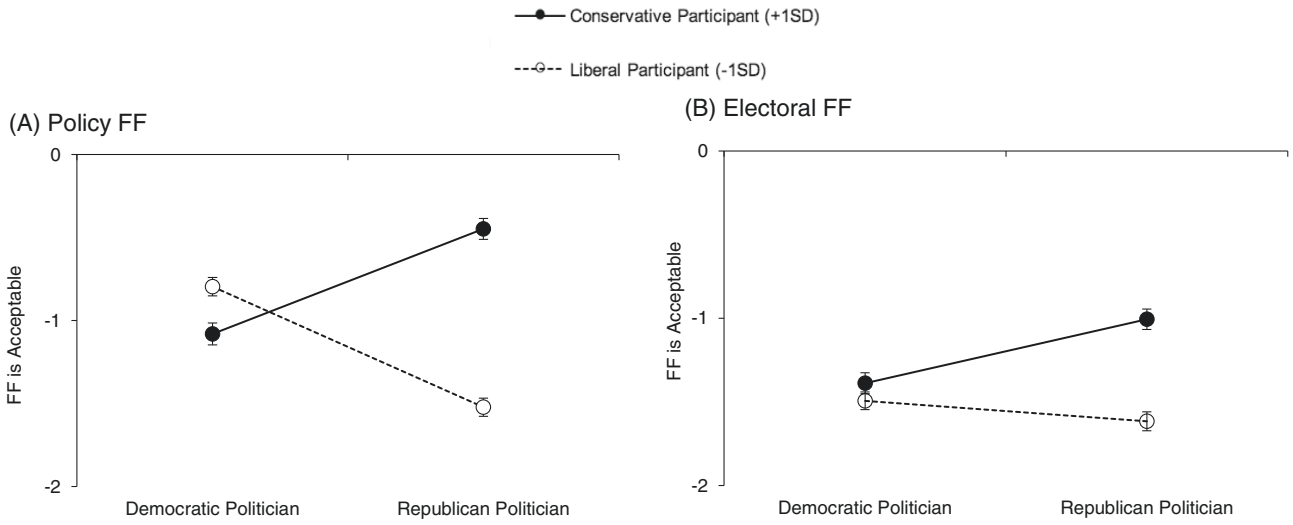
Although we have argued that this effect follows from the policy FF being a stronger signal of partisan trustworthiness, it was also the case that policy FFs were stronger signals of general trustworthiness. As such, we proceeded to conduct a more precise test of our hypotheses. We built on the first model by adding in terms for general trustworthiness (standardized), partisan trustworthiness (standardized), as well as an additional slate of interaction terms that simply replaced our FF variable with each form of trustworthiness. That is, the three interaction terms from the previous model that included the FF condition remained but were complemented by six interactions terms (three using general trustworthiness and three

[6] In this study, we knew on the basis of a previously run study that participants were likely to display some partisan leniency even toward electoral lies. This means that our central prediction was that a partially attenuated three-way interaction would emerge. Given that partially attenuated interactions make achievement of adequate statistical power difficult without very large sample sizes, we combined resources across two funding labs to reach this enormous (by experimental social psychological standards) sample size.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 6

Acceptability of the FF by Politician Party, Participant Political Orientation, and Whether the Politician Issued (A) a Policy FF or (B) an Electoral FF (Study 4)



Note. FF = flagged falsehood; *SD* = standard deviation. Depicted values are predicted means (and standard errors) as a function of the party of the politician and participants' political orientation (± 1 *SD* from political neutrality). Means are predicted at the overall average response for the fake news and unaware measures.

using partisan trustworthiness). This is because we have argued that the FF manipulation (policy or electoral) has the moderating effect it does on FF acceptability because of the different signal it sends about partisan trustworthiness (but not general trustworthiness). Consistent with this hypothesis, we found that to the extent that the politician was assumed to be more partisan trustworthy, the partisan bias in FF acceptability grew, $B = 0.17$, 95% CI [0.14, 0.21], $t(3763) = 9.38$, $p < .001$. But the more the politician was assumed to be generally trustworthy, the partisan bias in FF acceptability did not significantly vary, $B = 0.02$, 95% CI [−0.02, 0.06], $t(3763) = 1.19$, $p = .235$. In short, it was the elevated perceptions of partisan trustworthiness (as opposed to general trustworthiness) that policy FFs prompted that explained the greater partisan leniency toward policy FFs. The full results of this model can be found in the Supplemental Materials.

Additional Analyses

We reasoned that the electoral FF was less a signal of partisan trustworthiness than was the policy FF given that the policy FF directly promotes a goal a party actively embraces (thereby signaling partisan loyalty). In contrast, the electoral FF displays partisanship by transparently aiming to disenfranchise voters, something that parties do not intentionally affiliate themselves with. This logic suggests that if there are partisan biases in responding to the electoral FF, they may take a different form. That is, whereas likeminded partisans may have no problem admitting that a politician issuing a policy falsehood is intending to advance a party's goal (a worthy outcome in the eyes of the party), such partisans may simply deny that a politician issuing an electoral falsehood is truly trying to interfere with the election (the action that could reflect group loyalty).

And, indeed, we saw evidence of this in a study (in full disclosure, run for another purpose[7]) using only the electoral FF (N = 1,006 Americans from AMT; 907 after exclusions based on memory-based attention checks). We added the question, "Did Representative Wooley hope or intend that his tweet would decrease the number of voters in Downtown Albuquerque?" (1 = *not at all*, 9 = *definitely*). Although not predicted a priori, we observed a partisan bias in responses to this question, one that reflected that likeminded partisans *denied* that depressing voting was the politician's intent, $t(897) = 6.35$, $p < .001$. Of course, if likeminded partisans dismiss the politician's malintent, then in essence they are denying that the electoral FF actually had a group-serving aim. This could offer the more nuanced answer to why electoral FFs are weaker signals of partisan trustworthiness (and thus are greeted with less partisan leniency).

Guided by this intriguing finding, we returned to Study 4's data to probe perceptions of partisan trustworthiness more carefully. We used our initial model that probed for partisan biases in FF acceptability, but used the model to predict partisan trustworthiness instead. We observed a Party × Political orientation × FF interaction, $B = 0.08$, 95% CI [0.01, 0.15], $t(3771) = 2.39$, $p = .017$. The relevance of this effect becomes clear upon examining the partisan bias (the Party × Political orientation) in perceptions of partisan trustworthiness for each FF separately. Like in Study 3, policy FFs signaled elevated partisan trustworthiness fairly universally, for likeminded partisans as strongly as those from the other side, $B = -0.06$, 95% CI [−0.16, 0.04], $t(3771) = 1.25$, $p = .210$. But for

[7] We varied whether the tweet was actually visible to voters during or (due to the liar's error) only after the election to see whether its ability to actually influence the election influenced the partisan nature of FF acceptability. Although we replicated the partisan bias in FF acceptability, $t(897) = 2.34$, $p = .019$, this did not depend on the ability of the falsehood to actually influence the election, $t < 1$. Although we appreciate that null effects should be interpreted with caution, this reinforces a point from Study 3: The partisan bias in evaluating a FF's acceptability did not stem from the falsehood's ability to exert an actual effect on a group-relevant outcome.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

electoral FFs, there was a partisan bias in the extent to which they signaled partisan trustworthiness, $B = -0.23$, 95% CI [−0.32, −0.13], $t(3771) = 4.71$, $p < .001$. The negative coefficient reflects that it was likeminded partisans who were especially likely to see *diminished* evidence of partisan trustworthiness in this FF. After all, it is these politically sympathetic participants who were prone to see less malintent in this out-of-bounds act. It seems the policy FF (unlike the electoral FF) was universally accepted for what it was—an effort to promote the party's aims—thereby explaining the clearer partisan take on that FF's acceptability.

Study 5

Our previous studies identified, documented systematic variation in, and aimed to explain the occurrence of partisan evaluations of politicians' FFs. This final study extends on our previous experiments in three ways. First, we manipulated whether a politician told a FF or the truth. In both cases, the politician expressed support for a policy position that is compatible with his own party's platform; however, we varied whether, as part of that support, he cited an unsubstantiated research finding that supported the position (*falsehood* condition) or explicitly disavowed the false premise (*truth* condition). To begin, this allowed us to assess the effect of telling a FF (vs. the truth) on perceptions of a candidate's general trustworthiness and partisan trustworthiness. Recall that Studies 3 and 4 showed that general trustworthiness and partisan trustworthiness connected to our dependent variables (FF acceptability, moral behavior forecasts) in a nonpartisan and partisan way, respectively. If telling a policy FF (vs. the truth) has different implications for general and partisan trustworthiness, then we may be able to identify dual, distinct effects of issuing FFs on moral evaluation.

After all, prevaricating has been shown to erode trust (Schweitzer et al., 2006). A lack of perceived trustworthiness has been equated with a lack of perceived credibility (Hovland et al., 1953). Within our framework, it may be that telling a policy FF diminishes perceptions of general trustworthiness. On the other hand, other-oriented falsehoods (especially compared to egoistic or self-oriented ones) have been shown to be more acceptable (Lindskold & Walter, 1983; see also DePaulo et al., 1996)—especially when the perceiver (or, by extension, the perceiver's political group) stands to gain from it (Bocian & Wojciszke, 2014). Given policy FFs display a differential willingness to help one side of the political divide over the other, they should also communicate partisan trustworthiness. Whether heightened partisan trustworthiness stems not merely from support for the party's policy position (a feature of both the falsehood and the truthful statement) but also from a willingness to add a position-bolstering falsehood as part of that support will be revealed by the comparison between the falsehood and truth conditions. If so, this suggests that making false statements (vs. telling the truth) may generally depress perceptions of a politician as a moral actor (to the extent it depresses general trustworthiness) while simultaneously increasing partisan disagreement on that perspective (to the extent it increases partisan trustworthiness). This allows us to move beyond explaining variation in responses to FFs to identify these (possible dual) effects of issuing falsehoods.

Second, we probed for these dual effects using only the moral behavior composite, without first measuring the standard slate of excuses for FFs. One reason was practical: These excuse-making measures have less clear interpretations in the context of the new truth condition. Additionally, excluding these measures allowed us to understand perceivers' (nonpolitical) perceptions of the target, without such impressions being influenced by the excuse-making suggested by the measures themselves.

Third, we added a baseline moral behavior measure. Before reading the media account of the politician's tweet, participants made similar moral forecasts about all elected officials from one party—in particular, the party of which the target politician was a member. After all, group memberships may bias preexisting perceptions of in-group versus out-group members' trustworthiness and morality (e.g., Bergstrom et al., 2022). This allows our tests to isolate the effects of our manipulations on perceptions of the specific politician, independent of participants' preexisting beliefs about the moral character of elected officials from that party.

Method

Participants and Design

We recruited 1,274 Americans from Amazon's Mechanical Turk. We randomly assigned participants to one of four conditions in a 2 (party: Democrat or Republican) × 2 (veracity: truth or falsehood) full-factorial design. Applying the same standard used in our previous studies, we identified and excluded from further consideration the 184 participants who missed more than one memory-based attention check. This left 1,090 participants in the analyses reported below.

Procedure

To begin, participants completed the moral behavior composite used in Study 3, but, rather than the prompts being about the specific politician who issued the tweet, these items were about Democratic or Republican officeholders in general (depending on party condition). Participants considered the five everyday moral and five everyday immoral behaviors, and indicated what percentage of Democratic (or Republican) elected officials "in your best estimate, would do each of the behaviors in the context or circumstance described?" This *baseline moral behavior composite* served as a covariate in our key analyses.

At that point, participants read a newspaper article describing a tweet from (fictitious) New Mexico State Representative Steven Wooley. The tweet always related to the relationship between gun ownership and crime, but its veracity varied by condition. When the tweet was identified as a falsehood, the politician cited an unsubstantiated research claim. When the politician was said to be a Democrat, he indicated his support for gun control by tweeting that higher rates of gun ownership produce more crime; when a Republican, he indicated support for gun rights by stating higher rates of gun ownership produce less crime.

For other participants, the news article reported on a tweet from the politician that was said to be *true*. When the politician was a Democrat, he said, "The research is clear: Higher rates of gun ownership DON'T produce more crime (or have any effect on crime)." The tweet from the Republican was similar, but noted gun ownership does not "produce less crime (or have any effect on crime)." But just as in the falsehood versions, the politician followed up with a statement supporting "getting guns off NM streets" (Democratic) or "protecting NM gun owners' rights" (Republican). In both cases, the newspaper article did not take a position on the

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

wisdom of gun control, but simply focused on the truth-evaluable component of the tweets.

After participants completed three memory-based attention checks, they completed the two trustworthiness items (from Studies 3 and 4). In this case, general and partisan trustworthiness were weakly *negatively* correlated, $r(1,090) = -.07$, $p = .019$. At that point, participants completed the moral behavior measures used in Study 3. Unlike the baseline measure, this one asked about the likelihood that Representative Wooley in particular would engage in five everyday moral ($\alpha = .90$) and five everyday immoral behaviors ($\alpha = .89$) when in the next relevant situation. Finally, participants responded to the same two general political orientation questions used in all previous studies. They also indicated their support for or opposition to a number of more specific policy positions. Their response to "a citizen's right to bear arms" was combined with the two general items to identify participants' political orientation ($\alpha = .75$), such that higher values reflected greater conservatism.

Results and Discussion

General and Partisan Trustworthiness

To begin, we tested whether learning that the politician's statement was flagged as a falsehood as opposed to the truth encouraged different inferences about his general trustworthiness and partisan trustworthiness. The truth-telling politician was seen to be more generally trustworthy ($M = 9.08$, $SD = 2.58$) than the falsehood-uttering one ($M = 7.64$, $SD = 2.88$), $t(1083.22) = 8.70$, $p < .001$, $d = 0.53$. In contrast, when the politician issued a FF, he was seen as more partisan trustworthy ($M = 1.59$, $SD = 2.08$) than when the politician's statement was flagged as truthful ($M = 1.29$, $SD = 2.10$), $t(1088) = 2.34$, $p = .019$, $d = 0.14$. Note that these two effects have distinct implications for how the politician should be morally evaluated. The greater general trustworthiness that telling the truth implies should lead liberals and conservatives alike to see the politician as more moral. But the greater partisan trustworthiness that telling a (policy) FF prompts should, independently, add more of a partisan bent to the falsehood-delivering politician's perceived morality. Whereas Walker et al. (2021) argued that a major reputational risk for lying is undermining trustworthiness, the present findings suggest the importance of distinguishing general from partisan trustworthiness.

Moral Behaviors

We proceed by running a regression that predicts the moral behavior composite, the belief that the target politician would be more likely to engage in everyday moral behaviors and less likely to engage in everyday immoral behaviors. Crucially, this model controls for the baseline moral behavior composite. The predictors were party (+1: Republican, −1: Democratic), veracity (+1: falsehood, −1: truth), and political orientation, as well as all interaction terms that could be made from these variables. As anticipated by the veracity manipulation's effect on perceptions of general trustworthiness, we observed a main effect of veracity: The politician who issued a falsehood was forecast to behave less morally than the truthful one, $B = -5.28$, 95% CI [−6.87, −3.68], $t(1081) = 6.50$, $p < .001$. In addition, we observed a partisan bias (i.e., a Party × Political orientation interaction) on the moral behavior composite, $B = 2.95$, 95% CI [1.11, 4.79], $t(1081) = 3.16$, $p = .002$. But, as foreshadowed by the effects of the veracity manipulation on perceptions of partisan trustworthiness, this partisan bias was magnified when the politician issued a falsehood as opposed to the truth, $B = 1.81$, 95% CI [0.22, 3.39], $t(1081) = 2.23$, $p = .026$ (see Figure 7). More specifically, we observed clear evidence of this partisan bias when the politician told a falsehood, $B = 4.76$, 95% CI [2.39, 7.12], $t(1081) = 3.95$, $p < .001$. But when the politician told the truth, the partisan bias disappeared, $B = 1.14$, 95% CI [−1.33, 3.64], $t < 1$.

In a final model, we aimed to connect the two sets of effects just discussed. We not only included general trustworthiness (standardized) and partisan trustworthiness (standardized), but also added the six interaction terms that come from replacing the veracity variable with each of these two trustworthiness composites. That is, we retained the three interaction terms from the previous model that included the veracity manipulation. We thus created six new interaction terms that replaced the veracity component with general trustworthiness (3 terms) or partisan trustworthiness (3 terms).

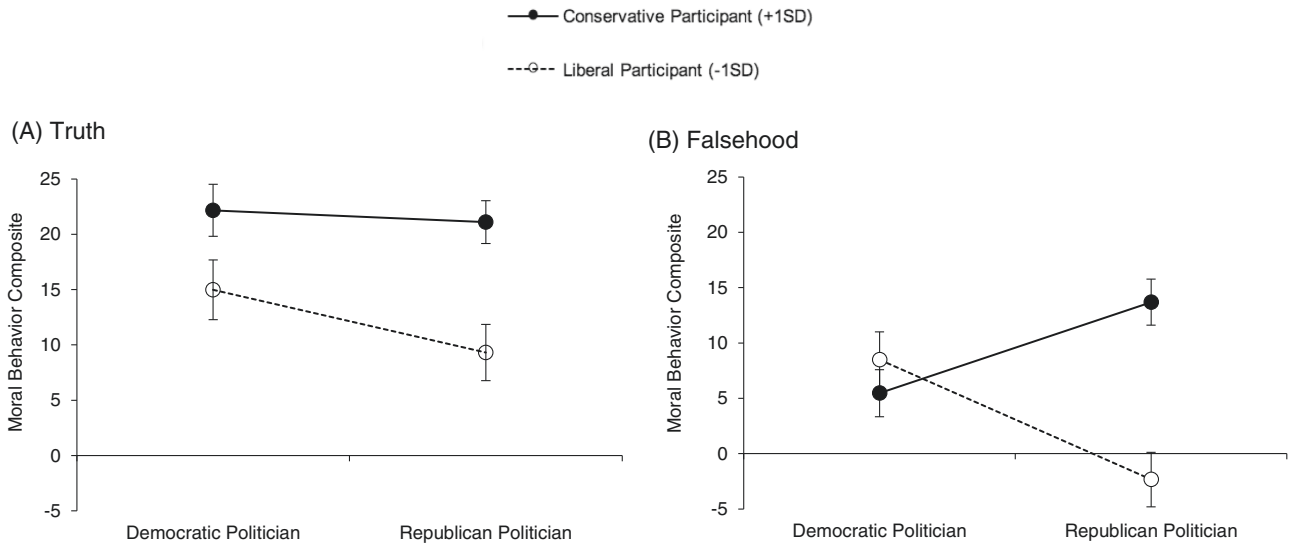
Whereas the full model output is supplied in the Supplemental Materials, we focus on the most relevant results here. We observe a main effect of general trustworthiness, $B = 9.42$, 95% CI [7.58, 11.26], $t(1073) = 10.06$, $p < .001$, one that was not qualified by a partisan bias, $B = 0.68$, 95% CI [−0.89, 2.26], $t < 1$. In other words, when the politician was seen as more generally trustworthy (as the truth-telling politician was), he was seen as a more generally moral actor, regardless of whether his political party matched the perceiver's political orientation. And even though we did not observe a main effect of partisan trustworthiness, $B = 0.34$, 95% CI [−1.25, 1.93], $t < 1$, its predictive power was qualified by a partisan bias, $B = 3.65$, 95% CI [2.08, 5.22], $t(1073) = 4.56$, $p < .001$. This reflects that to the extent a politician was seen as more partisan trustworthy (e.g., as a falsehood-issuing politician was) then like-minded partisans saw this politician as a more moral actor even in the nonpolitical sphere.

Whereas in our previous studies, we aimed to understand when and why there is variation in the extent to which political FFs are judged to be acceptable and a signal of a target's moral character, Study 5 instead documents dual signals that are sent by telling a (policy) FF as opposed to the truth. On the one hand, politicians issuing FFs are seen as less morally upstanding people, an effect that can in part be traced to their being seen as less generally trustworthy. But, on the other hand, politicians who have delivered FFs prompt more politically polarized perceptions, an effect that can in part be traced to the greater perceptions of partisan trustworthiness that the policy FFs encouraged. Note that this helps to resolve a tension between likeminded partisans' apparent commitment to honesty (reflected in their bias toward saying that their own side's politicians were likely telling the truth) and their leniency toward dishonesty (as observed in our FF acceptability measure). If general trustworthiness (as reflected in telling the truth) and partisan trustworthiness (as reflected in a willingness to issue a falsehood to advance the group's agenda) are independent (but directionally opposing) contributors to or signals of moral character, then it makes sense why likeminded partisans may both be skeptical that their own side's politicians have indeed uttered a falsehood even as they see some redeeming virtue in the added in-group loyalty that such falsehoods reflect. Loyalty to a group is demonstrated by maintaining one's group affiliation even at a personal cost (Levine & Moreland, 2002). These dual effects of dishonesty demonstrate both that personal cost and the stronger signal of partisan trustworthiness such actions prompted.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 7

Forecasted Moral Behavior by Politician Party, Participant Political Orientation, and the Veracity of the Tweet (Study 5)



Note. *SD* = standard deviation. Depicted values are predicted means (and standard errors) as a function of the party of the politician and participants' political orientation (± 1 *SD* from political neutrality). Means are predicted at the average level of the baseline moral behavior composite.

This provides another demonstration of how ethicality and trustworthiness—especially when distinguishing general and partisan trustworthiness—can be dissociated (see Levine et al., 2018).

General Discussion

Quite literally engraved into Judeo-Christian morality as part of The Ten Commandments, "Thou shalt not bear false witness against thy neighbor" is typically interpreted as a simple moral prohibition against lying. Calling someone a liar can seem tantamount to claiming they are a bad, untrustworthy person. From this perspective, it is surprising how much false political rhetoric has been tolerated in those who occupy the most powerful political positions in the U.S. and around the world. In this article, we find that who tells a flagged falsehood, what the FF is about, and who is listening all help predict how people explain and evaluate politicians who do not speak the truth. In so doing, we highlight that the moral acceptability of *bearing false witness* really depends on the extent to which such falsehoods are used in support of or against the explicit aims of *thy neighbors,* one's political in-group.

Summary of Empirical and Theoretical Contributions

In what follows, we organize our empirical findings by the contributions they make in four broad areas. We emphasize both the general theoretical take-aways and the empirical nuances that add color to them.

Partisan Evaluations of Flagged Falsehoods: Nature and Scope

First, we showed there is partisan leniency in how people evaluate the acceptability of political flagged falsehoods. Falsehoods that supported a policy position—regarding immigration reform (Study 1a), minimum wage laws (Studies 1b–1c, 3), school voucher policy (Study 1d), gun control measures (Studies 2a, 4), and affirmative action (Study 2b)—were seen to be more justifiable when they emanated from the perceiver's own political side. Partisans do disagree about the acceptability of FFs in part because they disagree about the facts of what they observed (i.e., whether such statements are false and believed by the speaker to be false). But, independent of these disagreements, partisans go further and differ in these FFs' perceived acceptability.

This characterization raises important theoretical questions regarding the scope of partisan evaluations. Empirically, instead of considering the two excuses for lying (fake news, unaware) as covariates, we can treat them as moderators. This would allow us to differentiate whether the partisan evaluations we have documented emerge quite generally (regardless of people's perceptions of the FFs as intentional falsehoods, actual truths, or something in between) or instead only in specific circumstances. For example, one possibility is that partisan evaluations may emerge only among those who deny that any misdeed was committed. If, for example, people believe a FF was in fact true and believed to be true by the speaker, then judgments of its acceptability would simply reflect an endorsement of the content instead of any leniency in evaluations of it.

We conducted two follow-up tests to determine: (a) whether the partisan leniency in FF acceptability depended on participants' making these previous excuses, and relatedly (b) whether even those who indicated the politician was indeed telling a falsehood (fake news < 0) and was likely aware of doing so (unaware < 0) would also show partisan leniency toward the acceptability of policy FFs. Given these analyses require the examination of higher-order interactions or a subsetted sample of the data, respectively, we have waited until now to maximize power for these analyses by pooling across the 4,925 participants in Studies 1–4 and Study SM1 who considered a policy FF. A random effect of study was included in each model.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

For our first cross-study analysis, we included fixed effects of the fibber's party, the participant's political orientation, the fake news excuse (centered at the scale midpoint), the unaware excuse (centered at the scale midpoint), as well as all interaction terms that can be made from these four variables. The dependent variable was FF acceptability. First, we replicate the effect of partisan leniency (Party × Political orientation) on FF acceptability, $B = 0.32$, $t(4905.88) = 7.78$, $p < .001$, now estimated for someone perfectly unsure if the statement was indeed false and perfectly unsure if the fibber was aware of issuing a falsehood (instead of at the sample means). Partisan evaluations of FF acceptability did not depend on participants' beliefs about whether the article made a false accusation and was thus fake news, $B = 0.00$, $t < 1$, whether the target was assumed to be unaware of telling a falsehood, $B = 0.03$, $t(4905.48) = 1.49$, $p = .137$, nor the interaction of those two excuses, $B = -0.00$, $t < 1$.[8] In other words, although FF acceptability was highly sensitive to the size of these two excuses (fake news: $B = 0.55$, $t(4905.47) = 32.19$, $p < .001$; unaware: $B = 0.19$, $t(4904.61) = 10.26$, $p < .001$), partisan leniency in FF acceptability was not moderated by the degree of this excuse-making. Considered in light of Study 5's findings, which showed that partisan evaluations *were* sensitive to whether the politician actively embraced or disavowed the false premise when offering support for a party-platform-consistent position, this reflects that what invites partisan evaluation is the target's loyal statement of support, regardless of the platform-friendly premise's perceived truth value. Overall, this provides evidence that the partisan evaluation effect is quite general.[9]

The Content of the FF

Our second central contribution is that partisan leniency did not extend equally to all FFs (Studies 2a–2b, 4). Partisans came down less hard on those who fibbed on behalf of their party's agenda, those called out for distorting the facts in a way that signaled commitment to the political in-group and its explicit goals. These partisan differences reduced or disappeared when considering mischaracterizations of the fibber's own life story (personal FFs) or group-serving prevarications that were inconsistent with a party's explicit values (electoral FFs). This clarifies our theoretical contribution. Partisan evaluations of FFs do not simply reflect an evaluative charitability extended to members of the perceivers in-group (e.g., Valdesolo & DeSteno, 2007), nor do they simply reflect the excusing of actions that could benefit the in-group. Given that elite political figures can have more potential to shape their own side's positions and beliefs than do nonpolitical subject matter experts (e.g., Sauer et al., 2021), then the partisan reluctance to condemn policy FFs may contribute to misinformation and division among the electorate.

Trustworthiness: Partisan and General

Consistent with a person-centered approach to morality that has identified trustworthiness as a core component to moral evaluations, the present work implicated perceived trustworthiness in explaining partisan evaluations of FFs. We introduced a distinction between general and partisan trustworthiness. Studies 3–5 established that it is perceptions of partisan (instead of general) trustworthiness—a perception that a politician can be trusted more by their own party than the opposing party—that identifies when likeminded partisans give those called out for issuing falsehoods a relative (though not necessarily an absolute) pass. Policy falsehoods were found to signal more partisan trustworthiness than personal falsehoods (follow-up to Study 3), electoral FFs (Study 4), or telling the truth (Study 5); in turn, partisan leniency toward policy FFs was especially strong.

Although we tested our effects with various FFs, the generality of these efforts' conclusions is still necessarily constrained by the specific FFs used to test them. Within each category of falsehood (e.g., policy, personal, electoral), there is no doubt variability in how much they signal the speaker's partisan trustworthiness. Future theoretical development will progress by achieving a more fine-grained understanding of what features of a falsehood (or actions more generally) communicate partisan trustworthiness. This would allow one to be able to better predict a priori precisely what misdeeds will be reacted to most divergently by partisans of different stripes. Furthermore, it would allow for the identification of exceptions to the general rule that policy FFs are particularly likely to be greeted with more partisan leniency.

How Falsehoods Signal Moral Character

Finally, we moved beyond commentaries on the FFs themselves by examining the speakers' perceived moral standing outside of the political domain, in their everyday lives (Studies 3, 5). As part of this effort, we developed a new measure, one that removed potential experimenter bias in stimulus selection in arriving at a set of the most prototypical everyday moral and immoral behaviors. We hope that this measure will be of use to moral psychologists, who are more accustomed to measuring moral evaluations through more subjective Likert-type scales that ask about agents' praiseworthy or blameworthy character. Applying this measure to the present research questions, we demonstrated how and why issuing (flagged) falsehoods influences moral evaluations compared to telling the truth. On the one hand, those who issue policy FFs (compared to the truth) are seen as less generally trustworthy, which explains why likeminded and opposing partisans see them to be of lower moral standing. But on the other hand, those who issue policy FFs (again, compared to the truth) are seen as more partisan trustworthy, which serves as a sign of better or worse moral character in the eyes of political in-group and out-group members, respectively.

First, this reflects how partisan leniency toward political FFs does not simply reflect shifting standards for excusing political misbehavior in particular. Instead, it reflects a broader moral standard—one that prioritizes in-group loyalty—that is used to infer general moral character. Second, the contrasting signals sent by fibbers and

[8] This foreshadows the results of our second cross-study analysis: Those participants who perceived the policy FF as an intentional falsehood ($N = 692$) showed a partisan bias in evaluations of these lies, $B = .27$, $t(686) = 6.17$, $p < .001$ (see Supplemental Materials for details).

[9] Vosgerau et al. (2019) compellingly argued that internal meta-analyses have the potential to inflate Type-I error because of researcher degrees of freedom regarding what measures or what studies to include in an internal meta-analysis. This concern arises as a researcher may hope to argue for aggregate evidence of an effect even when such evidence could not be consistently observed in individual studies. Note that the just-presented analyses instead show a null effect in the aggregate: That is, the cross-study analysis finds no evidence that partisan evaluations of FF acceptability are moderated by the two forms of excuse-making (fake news, unaware). The aggregate analyses simply offered more power to detect this effect—if it is indeed present—than did each study alone.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

truth-tellers further validates the importance of our novel distinction between general and partisan trustworthiness. Perceived trustworthiness is already known to signal benevolence and ethical integrity (Mayer et al., 1995). But the present findings speak to the importance of appreciating the dual nature of perceived trustworthiness in intergroup contexts: One type (general trustworthiness) has more universal meaning, whereas the other type (partisan trustworthiness) holds different significance depending on the eye (and really, the group membership) of the beholder.

Questions Posed by and Theoretical Implications for Related Literatures

We close by discussing other theoretical implications of our findings, addressing some superficial inconsistencies between the present and past research, and identifying some open questions for future research.

Moral Mandates

We argued that partisan leniency extends to policy FFs in particular because such falsehoods signal commitment to a group's agenda. But not all policy lies may serve this function. Political attitudes that tie into an individual's moral mandates (Skitka, 2002) help to define the individual's identity and thus delineate group membership (Rokeach, 1973; Taylor, 1989), thereby warping the individual's assessments of the justice of actions taken in defense of them. This suggests that the partisan leniency we explored may have been particularly robust if the issues we chose were related to participants' own moral mandates. Future research should explore whether policy lies in support of less morally relevant issues (e.g., economic issues like tariff policies; Mooney, 2001; Permoser, 2019; Wendell & Tatalovich, 2021) or those that are framed in less morally relevant terms (Mucciaroni et al., 2019) would receive the same magnified partisan evaluations. This may depend on whether such falsehoods still signal trustworthy allegiance to a group that advances a morally laden agenda in other ways.

In-Group Loyalty

Moral foundations theory adopts a functional perspective, arguing that morality is that which promotes harmony and cooperation within social groups (Haidt, 2008). In this tradition, loyalty to one's in-group is a binding foundation, one that encourages group solidarity; it encourages coalitional thinking, motivating individuals to get on board with and not question the aims of the group (Graham et al., 2013). Judging by responses to the Moral Foundations Questionnaire (MFQ, in-group loyalty is more core to conservatives' than to liberals' morality (Clifford et al., 2015; Graham et al., 2009, 2011; Weber & Federico, 2013). It also correlates with darker markers of conservatism, such as right-wing authoritarianism (Federico et al., 2013) and social dominance orientation (Kugler et al., 2014; Milojev et al., 2014). This might have led one to expect that conservatives would have been more likely than liberals to excuse their own politicians' FFs, an asymmetry we rarely observed.

More recently, it was argued that liberals and conservatives may show differences in how they evaluate moral principles or *issues*, but that need not imply they show the same differences in their evaluations of *people* (Frimer et al., 2013). Indeed, liberals and conservatives were fairly similar in how they leaned on moral foundations to judge historical figures' moral standing. In that work, for neither liberals nor conservatives did targets' behavioral reputation of in-group loyalty predict characterizations of targets as moral exemplars. Given we found that both liberals and conservatives displayed moral leniency toward those seen as partisan trustworthy, a clear marker of in-group loyalty, how do we resolve these apparent inconsistencies?

First, it is certainly possible that conservatives—responding to the sort of abstract questions that constitute the MFQ—might be more likely (than liberals) to endorse the principle that fealty to one's own party is paramount, but conservatives and liberals could still display the same partisan biases when evaluating actual political FFs. After all, strong partisans are marked by greater loyalty, whether they be on the left or the right (Clifford, 2017). Second, although Frimer et al. (2013) found that in-group loyalty predicted moral evaluations for neither liberals nor conservatives, it is important to note that their targets who were judged high in in-group loyalty (e.g., Ayatollah Khomeini, Pope John Paul II, Che Guevara) did not necessarily display loyalty to the perceiver's own group. Whereas Royce (1908) characterized loyalty—independent of the cause to which one is loyal—as humans' chief moral good, our own results showed it is not targets' loyalty to the targets' own group, but loyalty to the judge's own political group, that is taken as a reflection of better moral behavior and character. Thus, whereas Mayer et al. (1995) argued that those who are of high integrity (defined, in part, as being consistent and principled) are seen to be especially trustworthy, our own findings emphasize that this perspective requires nuance. It is consistent allegiance to and advocacy for the perceiver's own side's principles that signal trustworthiness and positive moral character (see also McFall, 1987; Sitkin & Roth, 1993).

Third, it is not that our participants went out of their way to heap praise on their own party's members who issued FFs. Instead, partisan leniency was reflected in participants' middling acceptability ratings of their own politicians' FFs (seen most clearly in the unadjusted means reported in the Supplemental Materials). As Study 5 demonstrated, issuing FFs can have dual, opposing effects: diminishing moral evaluations due to its signal of diminished general trustworthiness and boosting (or diminishing) moral evaluations due to its signal of partisan trustworthiness. In other words, partisan leniency more discouraged like-minded partisans from condemning FFs instead of encouraging them to see the FFs as perfectly justifiable.

Psychology's interest in whether in-group loyalty is or is not core to morality did not begin with moral foundations theory. Kohlberg (1981) saw the prizing of unconditional loyalty to an in-group (e.g., one's family) as a relatively unsophisticated moral stage beyond which sophisticated reasoners would ultimately grow, but at which others would stagnate (Walker, 2004). Turiel (1983) characterized in-group loyalty as a wise convention instead of a basic component of moral character. More recently, Kugler et al. (2014) cautioned that moral foundations like loyalty not be legitimized as *moral* foundations unless they can be shown to predict prototypically moral behavior. Our own findings (in Studies 3 and 5) adopt this final standard to understand whether people perceive in-group loyalty to be a signal of moral character, and thus as a truly moral foundation. Given that perceptions of partisan trustworthiness (differential trustworthiness to an in-group and out-group) predicted forecasts that the target would behave in prototypically moral ways,

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

it seems that in-group loyalty is *seen* as a reflection of moral character. But, crucially, loyalty to the perceiver's in-group (vs. out-group) was deemed a foundation of good moral character.

Although we have operationalized a politician's partisan trustworthiness as the difference between how much he can be trusted by his in-group as opposed to his out-group, which is consistent with the inherent partiality that is part and parcel of what in-group loyalty reflects, note that we can also empirically distinguish these two components. We conducted an additional analysis (see Supplemental Materials Table SM25) that decomposed partisan trustworthiness into the politician's perceived in-group trustworthiness and perceived out-group trustworthiness. In short, the effect is largely symmetrical. That is, both a politician's perceived in-group trustworthiness, $B = 0.07$, $(2570.02) = 5.46$, $p < .001$, and a politician's perceived out-group trustworthiness, $B = -0.08$, $t(2570.02) = -5.59$, $p < .001$, were each (opposing) signals that had different moral meaning to politically likeminded and opposing perceivers. Partisan trustworthiness simply captures these two independent signals in a single index.

Amplifying Falsehoods Oneself

One natural question is whether partisans' relative comfort with their own politicians' FFs also implies a comfort with telling or spreading such falsehoods themselves. Social media websites like Twitter make passing on (or retweeting) such information as simple as the click of a button. And indeed, Pennycook et al. (2021) showed that people prefer to share politically concordant content, even when more careful reflection would have allowed them to identify that content as false and thereby be more reluctant to share it. Although such work suggests few people are interested in knowingly spreading misinformation, it also suggests that accuracy concerns may not be naturally salient. In that sense, our own studies—which focus all participants on content accuracy and then probe for partisan biases in FF acceptability—likely reflected an especially conservative context in which to explore responses to and evaluations of political misinformation.

But even when attention to accuracy is high, when might likeminded partisans be especially likely to pass on a statement they know to be untrue? A finding from the nonpolitical domain may be informative. When Hildreth and Anderson (2018) placed groups of participants in competition and made the importance of loyalty salient to them, the participants saw themselves as moral when they lied to benefit their own group financially. Outside observers did not view their lies so charitably. This work offers an initial suggestion that loyal partisans—who occupy a political battlefield marked by severe tribalism—may be willing to serve up such falsehoods themselves. Situational factors that heighten partisans' sense of loyalty to their political in-group—as being engaged in political dialog just may do—may only further encourage this tendency. Whether such comfort would apply more to policy falsehoods—as stronger displays of partisan loyalty and trustworthiness—is a topic for future exploration.

Finally, although we have focused on how people respond to flagged falsehoods, our findings may also speak to why political lying has been so prevalent. Of course, many liars are not caught. But when they are—if those lies are in the service of an ideological agenda—the response is likely to be politically polarized. Those who never would have supported such a candidate may protest, while those in ideological agreement may be inclined to tolerate (even if not explicitly endorse) it. That pattern has become an increasingly common storyline in the American political arena as of late.

Conclusion

In her 2021 Nobel Peace Prize lecture, Filipino-American journalist and antifake-news activist Maria Ressa cautioned, "Without facts, you can't have truth. Without truth, you can't have trust. Without trust, we have no shared reality, no democracy, and it becomes impossible to deal with the world's existential problems." The present work in part confirms Ressa's grave warning, as falsehoods do undermine a fibber's perceived general trustworthiness. But what her perspective misses is that falsehoods thrive not simply because people do not understand the facts, but because those trafficking in untruths are tolerated by those who still see the fibbers as trustworthy to their own side. Correcting lies and distortions of course helps. But unless the electorate can avoid evaluating political figures' moral character through the lens of what those politicians are likely to do for one's own political tribe, instead of for people more generally, false discourse will not receive the universal condemnation it requires to be more fully eradicated.

References

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *The Journal of Economic Perspectives*, *31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Bergstrom, V. N. Z., Plaks, J. E., & Chasteen, A. L. (2022). To believe or not to believe: Stereotypes about agnostics. *Psychology of Religion and Spirituality*, *14*(1), 21–30. https://doi.org/10.1037/rel0000419

Bernhard, H., Fehr, E., & Fischbacher, U. (2006). Group affiliation and altruistic norm enforcement. *The American Economic Review*, *96*(2), 217–221. https://doi.org/10.1257/000282806777212594

Bocian, K., & Wojciszke, B. (2014). Self-interest bias in moral judgments of others' actions. *Personality and Social Psychology Bulletin*, *40*(7), 898–909. https://doi.org/10.1177/0146167214529800

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin*, *86*(2), 307–324. https://doi.org/10.1037/0033-2909.86.2.307

Clifford, S. (2017). Individual differences in group loyalty predict partisan strength. *Political Behavior*, *39*(3), 531–552. https://doi.org/10.1007/s11109-016-9367-3

Clifford, S., Jewell, R. M., & Waggoner, P. D. (2015). Are samples drawn from Mechanical Turk valid for research on political ideology? *Research & Politics*, *2*(4), Article 205316801562207. https://doi.org/10.1177/2053168015622072

Cohen, G. L. (2003). Party over policy: The dominating impact of group influence on political beliefs. *Journal of Personality and Social Psychology*, *85*(5), 808–822. https://doi.org/10.1037/0022-3514.85.5.808

Collins Dictionary. (2017). *Word of the Year 2017*. https://www.collinsdictionary.com/us/woty

Colquitt, J. A., Scott, B. A., & LePine, J. A. (2007). Trust, trustworthiness, and trust propensity: A meta-analytic test of their unique relationships with risk taking and job performance. *Journal of Applied Psychology*, *92*(4), 909–927. https://doi.org/10.1037/0021-9010.92.4.909

Cortina, L. M., & Magley, V. J. (2003). Raising voice, risking retaliation: Events following interpersonal mistreatment in the workplace. *Journal of Occupational Health Psychology*, *8*(4), 247–265. https://doi.org/10.1037/1076-8998.8.4.247

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Crone, D. L., & Laham, S. M. (2015). Multiple moral foundations predict responses to sacrificial dilemmas. *Personality and Individual Differences*, *85*, 60–65. https://doi.org/10.1016/j.paid.2015.04.041

Croson, R., Boles, T., & Murnighan, J. K. (2003). Cheap talk in bargaining experiments: Lying and threats in ultimatum games. *Journal of Economic Behavior & Organization*, *51*(2), 143–159. https://doi.org/10.1016/S0167-2681(02)00092-6

Curry, O. S. (2016). Morality as cooperation: A problem-centered approach. In T. K. Shackelford & R. D. Hansen (Eds.), *The evolution of morality* (pp. 27–51). Springer International Publishing. https://doi.org/10.1007/978-3-319-19671-8_2

Curry, O. S., Chesters, M. J., & Van Lissa, C. J. (2019). Mapping morality with a compass: Testing the theory of 'morality-as-cooperation' with a new questionnaire. *Journal of Research in Personality*, *78*, 106–124. https://doi.org/10.1016/j.jrp.2018.10.008

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*(2), 353–380. https://doi.org/10.1016/j.cognition.2008.03.006

Delton, A. W., & Krasnow, M. M. (2015). Adaptionist approaches to moral psychology. In J. Decety & T. Wheatley (Eds.), *The moral brain: A multidisciplinary perspective* (pp. 19–34). Boston Review.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, *70*(5), 979–995. https://doi.org/10.1037/0022-3514.70.5.979

Doherty, D., & Wolak, J. (2012). When do the ends justify the means? Evaluating procedural fairness. *Political Behavior*, *34*(2), 301–323. https://doi.org/10.1007/s11109-011-9166-9

Druckman, J. N., & McGrath, M. C. (2019). The evidence for motivated reasoning in climate change preference formation. *Nature Climate Change*, *9*(2), 111–119. https://doi.org/10.1038/s41558-018-0360-1

Dunbar, N. E., Gangi, K., Coveleski, S., Adams, A., Bernhold, Q., & Giles, H. (2016). When is it acceptable to lie? Interpersonal and intergroup perspectives on deception. *Communication Studies*, *67*(2), 129–146. https://doi.org/10.1080/10510974.2016.1146911

Dunn, J. (1988). Trust and political agency. In D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations*. Oxford University Press.

Effron, D. A. (2018). It could have been true: How counterfactual thoughts reduce condemnation of falsehoods and increase political polarization. *Personality and Social Psychology Bulletin*, *44*(5), 729–745. https://doi.org/10.1177/0146167217746152

Everett, J. A., Pizarro, D. A., & Crockett, M. J. (2016). Inference of trustworthiness from intuitive moral judgments. *Journal of Experimental Psychology: General*, *145*(6), 772–787. https://doi.org/10.1037/xge0000165

Federico, C. M., Weber, C. R., Ergun, D., & Hunt, C. (2013). Mapping the connections between politics and morality: The multiple sociopolitical orientations involved in moral intuition. *Political Psychology*, *34*(4), 589–610. https://doi.org/10.1111/pops.12006

Fiske, A. P. (1991). *Structures of social life: The four elementary forms of human relations: Communal sharing, authority ranking, equality matching, market pricing*. The Free Press.

Fiske, A. P., & Haslam, N. (2005). The four basic social bonds: Structures for coordinating interaction. In M. M. Baldwin (Ed.), *Interpersonal cognition* (pp. 267–298). Guilford Press.

Frimer, J. A., Biesanz, J. C., Walker, L. J., & MacKinlay, C. W. (2013). Liberals and conservatives rely on common moral foundations when making moral judgments about influential people. *Journal of Personality and Social Psychology*, *104*(6), 1040–1059. https://doi.org/10.1037/a0032277

Gambetta, D. (1988). Can we trust? In D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations*. Blackwell.

Gosa, T. L. (2017). Hip hop, authenticity, and styleshifting in the 2016 presidential election. *Journal of Popular Music Studies*, *29*(3), Article e12263. https://doi.org/10.1111/jpms.12236

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*, *47*, 55–130. https://doi.org/10.1016/B978-0-12-407236-7.00002-4

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029–1046. https://doi.org/10.1037/a0015141

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, *101*(2), 366–385. https://doi.org/10.1037/a0021847

Green, D. P., Palmquist, B., & Schickler, E. (2004). *Political parties and the social identities of voters: Partisan hearts and minds*. Yale University Press.

Groenendyk, E. (2012). Justifying party identification: A case of identifying with the 'Lesser of Two Evils'. *Political Behavior*, *34*(3), 453–475. https://doi.org/10.1007/s11109-011-9170-0

Haidt, J. (2008). Morality. *Perspectives on Psychological Science*, *3*(1), 65–72. https://doi.org/10.1111/j.1745-6916.2008.00063.x

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, *20*(1), 98–116. https://doi.org/10.1007/s11211-007-0034-z

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Publications.

Heilman, M. E., Block, C. J., & Lucas, J. A. (1992). Presumed incompetent? Stigmatization and affirmative action efforts. *Journal of Applied Psychology*, *77*(4), 536–544. https://doi.org/10.1037/0021-9010.77.4.536

Heiphetz, L., Strohminger, N., & Young, L. L. (2017). The role of moral beliefs, memories, and preferences in representations of identity. *Cognitive Science*, *41*(3), 744–767. https://doi.org/10.1111/cogs.12354

Helzer, E. G., & Critcher, C. R. (2018). What do we evaluate when we evaluate moral character? In K. Gray & J. Graham (Eds.), *Atlas of moral psychology* (pp. 99–107). Guilford Press.

Hildreth, J. A., Gino, F., & Bazerman, M. (2016). Blind loyalty? When group loyalty makes us see evil or engage in it. *Organizational Behavior and Human Decision Processes*, *132*, 16–36. https://doi.org/10.1016/j.obhdp.2015.10.001

Hildreth, J. A. D., & Anderson, C. (2018). Does loyalty trump honesty? Moral judgments of loyalty-driven deceit. *Journal of Experimental Social Psychology*, *79*, 87–94. https://doi.org/10.1016/j.jesp.2018.06.001

Hirschman, A. O. (1970). *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states* (Vol. 25). Harvard University Press.

Holan, A. D. (2015, December 11). All politicians lie. Some lie more than others. *New York Times*. https://www.nytimes.com/2015/12/13/opinion/campaign-stops/all-politicians-lie-some-lie-more-than-others.html

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion*. Yale University Press.

Iwai, T., Carvalho, J. V., & Lalli, V. M. (2018). Explaining transgressions with moral disengagement strategies and their effects on trust repair. *Brazilian Administration Review*, *15*(4), Article e180016. https://doi.org/10.1590/1807-7692bar2018180016

Jacobson, L. (2015). *Ben Carson said raising the minimum wage will increase joblessness*. http://www.politifact.com/truth-o-meter/statements/2015/nov/10/ben-carson/ben-carson-said-raising-minimum-wage-will-increase/

Kashy, D. A., & DePaulo, B. M. (1996). Who lies? *Journal of Personality and Social Psychology*, *70*(5), 1037–1051. https://doi.org/10.1037/0022-3514.70.5.1037

Kessler, G., & Kelly, M. (2018, January 20). President Trump made 2,140 false or misleading claims in his first year. *The Washington Post*. https://www.washingtonpost.com/news/fact-checker/wp/2018/01/20/president-trump-made-2140-false-or-misleading-claims-in-his-first-year/?utm_term=.2a5f69041dc7per

Kim, P. H., Dirks, K. T., & Cooper, C. D. (2009). The repair of trust: A dynamic bilateral perspective and multilevel conceptualization. *Academy of Management Review*, *34*(3), 401–422. https://doi.org/10.5465/amr.2009.40631887

Kohlberg, L. (1981). *Essays on moral development: The psychology of moral development* (Vol. 2). Harper & Row.

Kugler, M., Jost, J. T., & Noorbaloochi, S. (2014). Another look at moral foundations theory: Do authoritarianism and social dominance orientation explain liberal-conservative differences in "moral" intuitions? *Social Justice Research*, *27*(4), 413–431. https://doi.org/10.1007/s11211-014-0223-5

Landy, J. F., & Uhlmann, E. L. (2018). Morality is personal. In K. Gray & J. Graham (Eds.), *Atlas of moral psychology* (pp. 121–132). The Guilford Press.

Leth, P. (2021). Utterance interpretation and actual intentions. *Axiomathes*, *31*(3), 279–298. https://doi.org/10.1007/s10516-019-09462-x

Levine, E. E., Bitterly, T. B., Cohen, T. R., & Schweitzer, M. E. (2018). Who is trustworthy? Predicting trustworthy intentions and behavior. *Journal of Personality and Social Psychology*, *115*(3), 468–494. https://doi.org/10.1037/pspi0000136

Levine, E. E., & Schweitzer, M. E. (2015). Prosocial lies: When deception breeds trust. *Organizational Behavior and Human Decision Processes*, *126*, 88–106. https://doi.org/10.1016/j.obhdp.2014.10.007

Levine, E. F., & Schweitzer, M. (2014). Are liars ethical? On the tension between benevolence and honesty. *Journal of Experimental Social Psychology*, *53*, 107–117. https://doi.org/10.1016/j.jesp.2014.03.005

Levine, J. M., & Moreland, R. L. (2002). Group reactions to loyalty and disloyalty. In E. Lawler & S. Thye (Eds.), *Group cohesion, trust, and solidarity: Advances in group processes* (Vol. 19, pp. 203–228). Elsevier Science. https://doi.org/10.1016/S0882-6145(02)19008-4

Lim, C. (2018). Checking how fact-checkers check. *Research & Politics*, *5*(3), Article 205316801878684. https://doi.org/10.1177/2053168018786848

Lindskold, S., & Walters, P. S. (1983). Categories for acceptability of lies. *The Journal of Social Psychology*, *120*(1), 129–136. https://doi.org/10.1080/00224545.1983.9712018

Locksley, A., Ortiz, V., & Hepburn, C. (1980). Social categorization and discriminatory behavior: Extinguishing the minimal intergroup discrimination effect. *Journal of Personality and Social Psychology*, *39*(5), 773–783. https://doi.org/10.1037/0022-3514.39.5.773

Mahon, J. E. (2008). Two definitions of lying. *The International Journal of Applied Philosophy*, *22*(2), 211–230. https://doi.org/10.5840/ijap200822216

Malle, B. F., Knobe, J. M., & Nelson, S. E. (2007). Actor-observer asymmetries in explanations of behavior: New answers to an old question. *Journal of Personality and Social Psychology*, *93*(4), 491–514. https://doi.org/10.1037/0022-3514.93.4.491

Mason, L. (2015). 'I disrespectfully agree': The differential effects of partisan sorting on social and issue polarization. *American Journal of Political Science*, *59*(1), 128–145. https://doi.org/10.1111/ajps.12089

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709–734. https://doi.org/10.2307/258792

McFall, L. (1987). Integrity. *Ethics*, *98*(1), 5–20. https://doi.org/10.1086/292912

McManus, J. (2021). Emotions and ethical decision making at work: Organizational norms, emotional dogs, and the rational tales they tell themselves and others. *Journal of Business Ethics*. Advance online publication. https://doi.org/10.1007/s10551-019-04286-6

Milojev, P., Osborne, D., Greaves, L. M., Bulbulia, J., Wilson, M. S., Davies, C. L., Liu, J. H., & Sibley, C. G. (2014). Rightwing authoritarianism and social dominance orientation predict different moral signatures. *Social Justice Research*, *27*(2), 149–174. https://doi.org/10.1007/s11211-014-0213-7

Misch, A., Over, H., & Carpenter, M. (2014). Stick with your group: Young children's attitudes about group loyalty. *Journal of Experimental Child Psychology*, *126*, 19–36. https://doi.org/10.1016/j.jecp.2014.02.008

Mooney, C. Z. (2001). The public clash of private values: The politics of morality policy. In C. Z. Mooney (Ed.), *The public clash of private values* (pp. 3–17). Chatham House Publishers.

Mucciaroni, G., Ferraiolo, K., & Rubado, M. E. (2019). Framing morality policy issues: State legislative debates on abortion restrictions. *Policy Sciences*, *52*(2), 171–189. https://doi.org/10.1007/s11077-018-9336-2

Oborne, P. (2014). *The rise of political lying*. Simon and Schuster.

Oliver, R. L. (1999). Whence consumer loyalty? *Journal of Marketing*, *63*(4 Suppl. 1), 33–44. https://doi.org/10.1177/00222429990634s105

Oxford Dictionary. (2016). *Word of the Year 2016*. https://en.oxforddictionaries.com/word-of-the-year/word-of-the-year-2016

Padgett, M. Y., & Morris, K. A. (2000). *The impact of sex-based preferential selection on subordinate perceptions of a new supervisor* [Conference session] 2000 Administrative Sciences Association of Canada Conference in Montreal, QC, Canada.

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, *147*(12), 1865–1880. https://doi.org/10.1037/xge0000465

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, *592*(7855), 590–595. https://doi.org/10.1038/s41586-021-03344-2

Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, *25*(5), 388–402. https://doi.org/10.1016/j.tics.2021.02.007

Permoser, J. M. (2019). What are morality policies? The politics of values in a post-secular world. *Political Studies Review*, *17*(3), 310–325. https://doi.org/10.1177/1478929918816538

Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review*, *118*(1), 57–75. https://doi.org/10.1037/a0021867

Resnick, G. (2015). *Ben Carson's stabbing story is full of holes*. https://www.thedailybeast.com/ben-carsons-stabbing-story-is-full-of-holes

Rokeach, M. (1973). *The nature of human values*. Free Press.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, *23*(3), 393–404. https://doi.org/10.5465/amr.1998.926617

Royce, J. (1908). *The philosophy of loyalty*. Vanderbilt University Press.

Sauer, K. A., Capps, D. K., Jackson, D. F., & Capps, K. A. (2021). Six minutes to promote change: People, not facts, alter students' perceptions on climate change. *Ecology and Evolution*, *11*(11), 5790–5802. https://doi.org/10.1002/ece3.7553

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology*, *25*, 1–65. https://doi.org/10.1016/S0065-2601(08)60281-6

Schweitzer, M. E., Hershey, J. C., & Bradlow, E. T. (2006). Promises and lies: Restoring violated trust. *Organizational Behavior and Human Decision Processes*, *101*(1), 1–19. https://doi.org/10.1016/j.obhdp.2006.05.005

Seiter, J. S., Bruschke, J., & Bai, C. (2002). The acceptability of deception as a function of perceivers' culture, deceiver's intention, and deceiver-deceived relationship. *Western Journal of Communication*, *66*(2), 158–180. https://doi.org/10.1080/10570310209374731

Shaver, K. G. (1985). *The attribution of blame: Causality, responsibility, and blameworthiness*. Springer. https://doi.org/10.1007/978-1-4612-5094-4_5

Shweder, R. A., Much, N. C., Mahapatra, M., & Park, L. (1997). The "big three" of morality (autonomy, community, divinity) and the "big three" explanations of suffering. In A. M. Brandt, & P. Rozin (Eds.), *Morality and health* (pp. 119–169). Routledge.

Simpson, J. A. (2007). Foundations of interpersonal trust. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (2nd ed., pp. 587–607). Guilford Press.

Sitkin, S. B., & Roth, N. L. (1993). Explaining the limited effectiveness of legalistic "remedies" for trust/distrust. *Organization Science*, *4*(3), 367–392. https://doi.org/10.1287/orsc.4.3.367

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Skitka, L. J. (2002). Do the means always justify the ends, or do the ends sometimes justify the means? A value protection model of justice reasoning. *Personality and Social Psychology Bulletin*, *28*(5), 588–597. https://doi.org/10.1177/0146167202288003

Suh, I., Sweeney, J. T., Linke, K., & Wall, J. M. (2020). Boiling the frog slowly: The immersion of C-suite financial executives into fraud. *Journal of Business Ethics*, *162*(3), 645–673. https://doi.org/10.1007/s10551-018-3982-3

Swider, B. W., Harris, T. B., & Gong, Q. (2022). First impression effects in organizational psychology. *Journal of Applied Psychology*, *107*(3), 346–369. https://doi.org/10.1037/apl0000921

Swire, B., Berinsky, A. J., Lewandowsky, S., & Ecker, U. K. H. (2017). Processing political misinformation: Comprehending the Trump phenomenon. *Royal Society Open Science*, *4*(3), Article 160802. https://doi.org/10.1098/rsos.160802

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin, & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–37). Brooks/Cole.

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. Austin (Eds.), *Psychology of intergroup relations* (pp. 7–24). Nelson-Hall.

Taylor, C. (1989). *Sources of the self: The making of the modern identity*. Harvard University Press.

Tepe, B., & Aydinli-Karakulak, A. (2019). Beyond harmfulness and impurity: Moral wrongness as a violation of relational motivations. *Journal of Personality and Social Psychology*, *117*(2), 310–337. https://doi.org/10.1037/pspi0000169

Thau, S., Derfler-Rozin, R., Pitesa, M., Mitchell, M. S., & Pillutla, M. M. (2015). Unethical for the sake of the group: Risk of social exclusion and pro-group unethical behavior. *Journal of Applied Psychology*, *100*(1), 98–113. https://doi.org/10.1037/a0036708

Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., & Herrmann, E. (2012). Two key steps in the evolution of human cooperation: The interdependence hypothesis. *Current Anthropology*, *53*(6), 673–692. https://doi.org/10.1086/668207

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge University Press.

Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science*, *10*(1), 72–81. https://doi.org/10.1177/1745691614556679

Umphress, E. E., Bingham, J. B., & Mitchell, M. S. (2010). Unethical behavior in the name of the company: The moderating effect of organizational identification and positive reciprocity beliefs on unethical pro-organizational behavior. *Journal of Applied Psychology*, *95*(4), 769–780. https://doi.org/10.1037/a0019214

Valdesolo, P., & DeSteno, D. (2007). Moral hypocrisy: Social groups and the flexibility of virtue. *Psychological Science*, *18*(8), 689–690. https://doi.org/10.1111/j.1467-9280.2007.01961.x

Van Vugt, M., & Hart, C. M. (2004). Social identity as social glue: The origins of group loyalty. *Journal of Personality and Social Psychology*, *86*(4), 585–598. https://doi.org/10.1037/0022-3514.86.4.585

Vosgerau, J., Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2019). 99% impossible: A valid, or falsifiable, internal meta-analysis. *Journal of Experimental Psychology: General*, *148*(9), 1628–1639. https://doi.org/10.1037/xge0000663

Walker, A. C., Turpin, M. H., Meyers, E. A., Stolz, J. A., Fugelsang, J. A., & Koehler, D. J. (2021). Controlling the narrative: Euphemistic language affects judgments of actions while avoiding perceptions of dishonesty. *Cognition*, *211*, Article 104633. https://doi.org/10.1016/j.cognition.2021.104633

Walker, L. J. (2004). Kohlberg and the structural–developmental approach to moral psychology. In D. C. Thomasma & D. N. Weisstub (Eds.), *The variables of moral capacity* (pp. 43–56). Kluwer Academic. https://doi.org/10.1007/978-1-4020-2552-5_3

Waytz, A., Dungan, J., & Young, L. (2013). The whistleblower's dilemma and the fairness–Loyalty tradeoff. *Journal of Experimental Social Psychology*, *49*(6), 1027–1033. https://doi.org/10.1016/j.jesp.2013.07.002

Weber, C. R., & Federico, C. M. (2013). Moral foundations and heterogeneity in ideological preferences. *Political Psychology*, *34*(1), 107–126. https://doi.org/10.1111/j.1467-9221.2012.00922.x

Weiner, B. (1995). Attribution theory in organizational behavior: A relationship of mutual benefit. In M. J. Martinko (Ed.), *Attribution theory in organizational perspective* (pp. 3–6). St. Lucie Press.

Wendell, D. G., & Tatalovich, R. (2021). Classifying public policies with moral foundations theory. *Policy Sciences*, *54*(1), 155–182. https://doi.org/10.1007/s11077-020-09399-8

Whitson, J. A., Kim, J., Wang, C. S., Menon, T., & Webster, B. D. (2019). Regulatory focus and conspiratorial perceptions: The importance of personal control. *Personality and Social Psychology Bulletin*, *45*(1), 3–15. https://doi.org/10.1177/0146167218775070

Young, L., & Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition*, *120*(2), 202–214. https://doi.org/10.1016/j.cognition.2011.04.005

(*Appendices follow*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix A

Language of Political Communication by Study

Study/Condition	Republican FF	Democratic FF
Study 1a (Tweet—Immigration)	The facts are clear: when immigrants move into your neighborhood, crime increases. #ImmigrationCausesCrime	The facts are clear: when immigrants move into your neighborhood, crime decreases. #ImmigrationStopsCrime
Study 1b (Tweet—Minimum wage)	The facts show that every single time other states raised their minimum wage, unemployment rose. NM should clearly avoid that mistake. #KeepTheMinimumWage	The facts show that every single time other states raised their minimum wage, unemployment fell. NM should clearly follow their lead. #RaiseTheMinimumWage
Study 1c (Speech—Minimum wage)		
Variant 1	Although studies show that the expense of raising the minimum wage keeps most working-class families from getting any pay raises at all, that isn't why I oppose increasing the minimum wage. My steadfast opposition is rooted in research that consistently shows that raising the minimum wage leads to increases in unemployment.	Although studies show that the benefit of raising the minimum wage leads to most working-class families getting larger pay raises than they would have, that isn't why I support increasing the minimum wage. My steadfast support is rooted in research that consistently shows that raising the minimum wage leads to decreases in the unemployment rate.
Variant 2	Although studies show that the expense of raising the minimum wage leads to increases in the unemployment rate, that isn't why I oppose increasing the minimum wage. My steadfast opposition is rooted in research that consistently shows that raising the minimum wage keeps most working-class families from getting any pay raises at all.	Although studies show that the benefit of raising the minimum wage leads to decreases in the unemployment rate, that isn't why I support increasing the minimum wage. My steadfast support is rooted in research that consistently shows that raising the minimum wage leads to most working-class families getting larger pay raises than they would have.
Study 1d (Tweet—School vouchers)	Support our kids, support vouchers! All the research is clear: Children who use vouchers to attend private/parochial schools see soaring test scores. #VouchersForNM	Support our kids, oppose vouchers! All the research is clear: Children who use vouchers to attend private/parochial schools see plummeting test scores. #NOVouchersForNM
Study 2a (Gun control)		
Policy FF	NM gun owners need their rights protected. U.S. crime stats show that states with more guns have less gun violence #GunOwnersRights	NM gun control laws need to be better. U.S. crime stats show that states with more guns have more gun violence #GunOwnersRights
Personal FF	Gun violence has touched us all. When I was a teenager, I was a customer at McDonalds and was shot at during a robbery attempt. #GunViolence.	Gun violence has touched us all. When I was a teenager, I was a customer at McDonalds and was shot at during a robbery attempt. #GunViolence.
Study 2b (Affirmative action)		
Policy FF	Even though affirmative action tries to help minorities, it hurts everyone. U.S. Census data show Whites and African Americans earn less in states with affirmative action #BanAffirmativeAction	Affirmative action helps everyone. U.S. Census data show both Whites and African Americans earn more in states with affirmative action #AffirmativeAction
Personal FF	College is where we learn about ourselves. I learned about the complex issues of race in America from my African American roommate. #CollegeLife	College is where we learn about ourselves. I learned about the complex issues of race in America from my African American roommate. #CollegeLife
Study 3 (Minimum wage)		
Policy FF	The facts show that every single time other states raised their minimum wage, unemployment rates rose. NM should clearly avoid that mistake. #KeepTheMinimumWage	The facts show that every single time other states raised their minimum wage, unemployment rates fell. NM should clearly follow their lead. #RaiseTheMinimumWage
Study 4 (Gun control/polling access)		
Policy FF	NM gun owners need their rights protected. U.S. crime stats show that states with more guns have less gun violence. #GunOwnersRights	NM gun control laws need to be better. U.S. Crime stats show that states with more guns have more gun violence. #GunControl
Electoral FF	Lines in Downtown Albuquerque voting precincts are 4 hr long. If you're not already in line, you won't get to vote because polls close at 7.	
Study 5 (Gun control)		
Falsehood	The research is clear: Higher rates of gun ownership DO produce less crime. But what matters to me is protecting NM gun owners' rights. #GunOwnersRights	The research is clear: Higher rates of gun ownership DO produce more crime. But what matters to me is getting guns off NM streets. #GunControl
Truth	The research is clear: Higher rates of gun ownership DON'T produce less crime (or have any effect on crime). But what matters to me is protecting NM gun owners' rights. #GunOwnersRights	The research is clear: Higher rates of gun ownership DON'T produce more crime (or have any effect on crime). But what matters to me is getting guns off NM streets. #GunControl

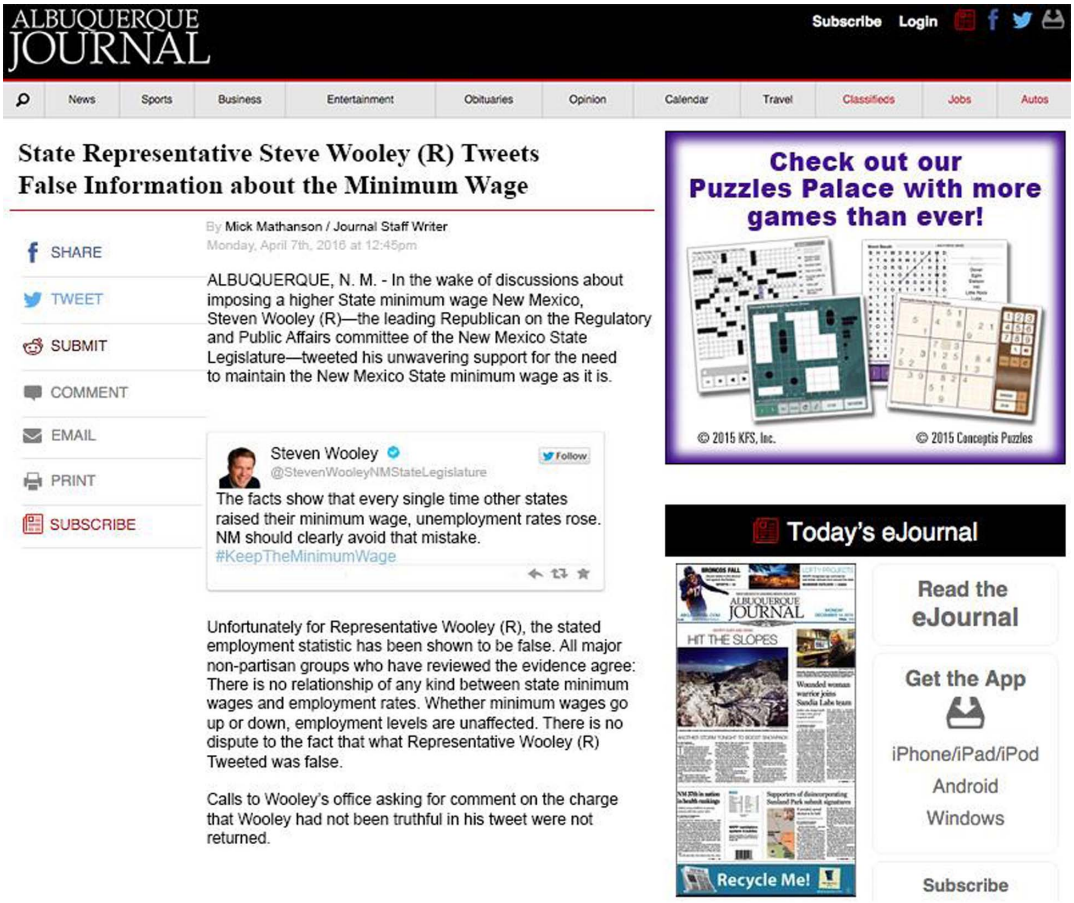
Note. FF = flagged falsehood.

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix B

Sample Stimulus From Study 1b (Minimum Wage), Republican Party Condition



Note. See the online article for the color version of this figure.

Received April 24, 2020
Revision received April 24, 2022
Accepted May 11, 2022 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.