Journal of Experimental Psychology: General

When and Why “Staying out of It” Backfires in Moral and Political Disagreements

Ike Silver and Alex Shaw

Online First Publication, April 21, 2022. http://dx.doi.org/10.1037/xge0001201

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

https://doi.org/10.1037/xge0001201

When and Why "Staying out of It" Backfires in Moral and Political Disagreements

Ike Silver[1, 2] and Alex Shaw[3]

[1] The Wharton School, University of Pennsylvania

[2] Kellogg School of Management, Northwestern University

[3] Department of Psychology, University of Chicago

People care where others around them stand on contentious moral and political issues. Yet when faced with the prospect of taking sides and the possibility of alienating observers with whom they might disagree, actors often try to "stay out of it"—communicating that they would rather not to take a side at all. We demonstrate that despite its intuitive appeal for reducing conflict, opting not to take sides over moralized issues can harm trust, even relative to siding against an observer's viewpoint outright. Across eleven experiments (N = 4,383) using controlled scenarios, real press video clips, and incentivized economic games, we find that attempts to stay out of the fray are often interpreted as deceptive and untrustworthy. When actors choose not to take sides, observers often ascribe concealed opposition, an attribution of strategic deception which provokes distrust and undermines real-stakes cooperation and partner choice. We further demonstrate that this effect arises only when staying out of it seems strategic: Actors who seem to hold genuine middle-ground beliefs or who lack incentives for impression management are not distrusted for avoiding conflict. People are often asked to take sides in moral and political disagreement. Our findings outline a reputational risk awaiting those who opt not to do so.

Keywords: politics, polarization, trust, impression management, nondisclosure

People dislike and distrust those who oppose them over contentious political issues like immigration policy, gun control, COVID-19 safety, or abortion rights (Finkel et al., 2020; Rogowski & Sutherland, 2016). Indeed, both liberals and conservatives see their own moral and political beliefs as objectively superior to alternative perspectives, avoid exposure to opposing viewpoints, and often construe disagreement over hot-button issues as a threat to their core values (Chambers & Melnyk, 2006; Frimer et al., 2017; Toner et al., 2013).[1] More broadly, across a variety of interpersonal settings, those who find themselves on "the wrong side" of moralized issues often encounter outrage, intolerance, and prejudice (Brandt et al., 2014; Chambers et al., 2013; Ditto et al., 2019). It is no surprise then, given the costs of moral disagreement, that many individuals feel discomfort at the prospect of sharing their views about hot-button issues in personal and professional settings (Pew Research Center, 2019).

One intuitive way to avoid hostility when asked to weigh in on such issues is simply not to take sides. Is expressing, for example, that one would prefer to "stay out of it" an effective interpersonal strategy? Although much is known about how observers respond to those who side with or against them in moral and political conflict, we know little about how people evaluate those who indicate that they would rather not take sides at all. In this paper, we propose and find that opting not to weigh in on contentious issues can backfire, often drawing stronger reproach than opposing an observer's ideological position outright.

At first blush, opting not to take sides seems sensible enough, particularly when staking out a clear position on a contentious issue might anger audiences who hold the opposing view. In such settings, a desire not to weigh in might reasonably seem intellectually humble or interpersonally courteous.[2] At a minimum, choosing not to take sides (vs. disagreeing with one's audience outright) may seem less

Ike Silver https://orcid.org/0000-0002-7206-8018

The authors thank Geoff Goodwin, Deborah Small, Barbara Mellers, Corey Cusimano, the Penn Moral Psychology Reading Group, the Wharton Decision Processes Lab, and the UChicago Developmental Investigation of Behavioral Strategies lab for helpful feedback on this work. The authors are also grateful to attendees at the Society for Consumer Psychology, Society for Personality and Social Psychology, and Association for Consumer Research annual conferences, who offered generative comments and questions on earlier iterations.

All pre-registrations, stimuli, data, and supplemental materials: https://researchbox.org/118&PEER_REVIEW_passcode=OJNSOY.

Correspondence concerning this article should be addressed to Ike Silver, Kellogg School of Management, Northwestern University, Kellogg Global Hub 2211 Campus Drive, Suite 5409, Evanston, IL 60208, United States. Email: Ike.M.Silver@gmail.com

[1] Such dynamics have powerful behavioral consequences: To name just a few, they impact where people work (Gandz & Murray, 1980) and what they buy, (Jost, 2017), as well as whom they choose to collaborate with (Marks et al., 2019), talk to (Chen & Rohla, 2018), or vote for (Parker & Isbell, 2010).

[2] Even on Gricean conversational frameworks, which often prescribe answering questions with adequate information, demurring when one's answer might cause discomfort or interpersonal conflict is seen as normative, polite, and considerate (Grice, 1975; Yoon et al., 2020).

likely to provoke moral outrage or alienate one from relevant social groups (Byrne, 1971; Crockett, 2017). As noted, outright opposition can be costly. In everyday settings, disputes over morality and politics can end friendships, alienate coworkers, and strain family ties. And for prominent individuals in the public eye, contradicting observers' moral beliefs can provoke reproach in the form of protests, boycotts, or public ridicule. To sidestep these costs, it may seem sensible to stay out of the fray altogether, avoiding a position offensive to either side by saying things like "I'd rather not get into that right now," "I'm neutral on this issue," or "Let's not talk politics over dinner."

By contrast, we submit that across a variety of situations, opting not to take sides can backfire. Why? In the context of charged discussions, observers may readily interpret such stances through the lens of social incentives. In general, people are attuned to the risks and rewards that others face in social contexts and will account for such incentives when interpreting behavior (Jones & Pittman, 1982; Schlenker, 1980; Sperber et al., 2010). For example, observers tend to discount bragging and self-promotion because they recognize these as deceptive impression management tactics intended to cultivate a positive reputation (Berman et al., 2015; Wosinska et al., 1996). Similarly, an actor who opts not to take sides when asked to respond to a hot-button issue may seem to be doing something strategic. More specifically, we propose that observers will draw on contextual information about an actor's intended audience to interpret relevant incentives, infer underlying beliefs, evaluate trustworthiness, and make decisions about cooperation and support.

Our contexts of interest are any in which an actor is asked to weigh in on a contentious moral or political issue in front of an audience with a known position. Such situations arise frequently. Consider, for example, a workplace conversation in which a new member of an office is asked for her opinion about gun control. Or imagine a prominent businesswoman asked in a TV interview for her perspective on COVID-19 policy or a politician called upon to indicate whether or not a recent election was fraudulently decided. Across these cases and many others like them, actors are put on the spot to provide an opinion: They can take a position aligned with or opposed to their audience's, or they can choose not to take a position at all. Importantly, when actors address divisive issues, observers likely have intuitions, if not common knowledge, about what the actor's audience believes (e.g., the ideological lean of an office, a friend group, or a news station's viewers), especially given that people often sort their personal and professional networks along partisan lines (e.g., Alford et al., 2011; Bonica et al., 2020). Here, we explore how observers use audience information in such contexts to interpret and evaluate those who try to stay above the fray.

First, we predict that choosing not to take sides will be interpreted as tacit opposition when addressing an audience with a known position. Take one of our examples from above. Imagine a new employee who says she "doesn't want express an opinion" about gun control in an office where most are politically liberal and support stricter gun laws. On our account, her coworkers will likely infer that their new colleague may be concealing underlying conservative, progun convictions. After all, if she agreed with the opinion held by most of her coworkers, saying so would be both truthful and aligned with social incentives. We posit that "not wanting to express an opinion" may instead signal that an actor's underlying beliefs actually oppose those of her audience. Thus, when addressing a predominantly liberal audience, such an assertion will signal more conservative underlying beliefs: When addressing a conservative audience, saying the same thing might signal more liberal underlying beliefs.

Second, we predict that if an actor seems to harbor private convictions that differ from what they say publicly (e.g., "I really cannot not get involved"), trying to stay out of it may seem like a strategic concealment tactic, and therefore elicit distrust. Converging evidence suggests that when actors send "false-signals" about their underlying beliefs or qualities, observers respond negatively (Jordan et al., 2017; Silver et al., 2021), often with distrust in particular. For example, individuals who choose not to disclose unsavory past behavior (like illegal drug use) on personal questionnaires are less attractive dates and employees because intentionally hiding past behavior seems untrustworthy (John et al., 2016). Similarly, when audience information is available as context, opting not to weigh in may seem like strategically concealing one's oppositional views. Indeed, even for powerfully divisive issues, opting not to get involved may elicit stronger distrust than opposing an observer's viewpoint outright.

Unlike more duplicitous ways of avoiding tough questions, such as "dodging" (answering unasked but related questions; Rogers & Norton, 2011), "paltering" (answering with truthful but misleading statements; Rogers et al., 2017), or "deflecting" (answering with further questions; Bitterly & Schweitzer, 2020), the strategy we investigate entails making a clear choice not to get involved. In this way, trying to stay out of it may be understood—at least by observers—as a form of active nondisclosure (Baum & Critcher, 2019), and distinguished both from covert attempts to change the subject and from efforts to avoid situations in which contentious topics might be discussed in the first place. Thus, unlike previous work, we do not focus on whether observers can detect an actor's desire not to take sides, but on how they will attribute underlying beliefs and intentions to actors who choose openly to "stay out of it" and on how these belief inferences impact downstream judgments and decisions.

Importantly, our account holds that people interpret attempts to stay out of it through the lens of social incentives, such that opting not to take sides should be judged more negatively to the extent that it seems strategically motivated. If, for example, an observer believes that an actor's choice not to take sides reflects genuinely moderate underlying beliefs (i.e., no underlying commitment one way or the other) or can attribute it to nonstrategic motives, we do not expect it to backfire. In this way, we do not predict a broad aversion to moderate beliefs or to nonresponsiveness per se but a more specific cynicism about the strategic motives of those who try to skirt contentious topics. This feature distinguishes our investigation from other accounts which might predict more general negative feelings toward moral apathy or uncooperativeness (e.g., Zlatev, 2019, Grice, 1975). Our effects depend on specific patterns of belief inference and motive attribution drawn from common and predictable social contexts.

Present Experiments

Eleven experiments conducted with online and student samples examined how observers respond to actors who try to stay out of moral issues. These employed several complementary methods, including scenario studies (Experiments 1a-b, 3, 5, 6a-b), judgments of real press video clips (Experiments 2a-b), and incentive compatible economic games (Experiments 4a-b). Across our stimuli, we presented participants with a variety of examples of real and hypothetical actors (e.g., other study participants, friends, family members,

professors, celebrities, politicians) opting not to take sides over a variety of divisive issues (e.g., abortion, immigration, gun rights, COVID-19 safety). We then gauged their reactions in terms of belief inferences, attitudinal trust, intended voting support, partner choice, and real-stakes cooperation. Importantly, we also conducted a follow-up study which confirms that actors put on the spot to take sides (a) see staying out of it as a strongly appealing option and (b) spontaneously generate expressions much like those we test in our main experiments, suggesting that speakers may not fully grasp the costs that come with choosing not to take sides. In summary, we find robust backfire effects in response to common tactics people use to avoid taking sides, and we also document theoretically informed boundaries. In the General Discussion, we consider more effective ways to stay out of moral and political disagreements.

Before reporting these experiments, we note two final points. First, it should be noted that we investigate participants' responses to what people say, which is readily observed, rather than to what they believe, which must be inferred. That is, although an actor's choice to stay out of it may be consistent with a variety of underlying goals and dispositions (e.g., a concealed opinion, a moderate or mixed opinion, no opinion), observers typically rely on the charged social contexts in which such communication is embedded when making inferences. We contend that, whatever one's true motives, opting not to weigh in often *resembles* strategic concealment. Second, our experiments pay special attention to the conservative comparison between staying out of it and opposing an observer's view outright, but we note that avoiding a position also entails an additional opportunity cost: that of failing to side with either side. Thus, an actor who avoids taking sides may fail both to placate those who would oppose their underlying position and to woo those who would support it. We document this latter (and somewhat more obvious) cost in Experiment 3.

Experiments 1b and 4a-6b, and our follow-up study were all preregistered at AsPredicted.Org. We report all manipulations and measures. All sample sizes were determined in advance, and any reported exclusions were preregistered[3]. Study procedures were approved by the IRB at the University of Chicago. All preregistrations, materials, and data can be accessed at: https://researchbox.org/118&PEER_REVIEW_passcode=OJNSOY

Experiments 1a-1b: Attributing Beliefs to Those Who Choose Not to Take Sides

Experiments 1a and 1b tested the prediction that opting not to take sides in front of a primarily conservative audience would signal more liberal beliefs while doing so in front of a primarily liberal audience would signal more conservative beliefs. Experiment 1A used scenarios about judges, politicians, and business owners asked for their views in front of large public audiences. Experiment 1B used scenarios about a college professor speaking at a faculty meeting and a cousin interacting with family at a barbeque.

Experiment 1a

Method

Three hundred and one MTurk participants (mean age = 36.8, *SD* = 11.8, 49.8% female) read one of three short stimulus-sampled scenarios. Each described a case in which an actor is asked for their position on a contentious issue in a public forum. The first scenario described a county judge who is asked in an internet forum whether abortion should be legal or illegal. The second described a prominent businessman who is asked at a press conference whether NFL players should be required to stand or allowed to kneel during the national anthem. The third described an elected representative who is asked during a town hall whether she thinks confederate statues in her district should be taken down or left standing. In each case, the public figure opts not to take sides, saying that the issue is "very complex" and that they "cannot take a side at this time." Note that across our studies, we vary how actors verbalize their choice to stay out of it (see Table 1). We confirm that our stimuli resemble what people are actually inclined to say in our follow-up study (which appears after Experiment 6b).

In each scenario, we manipulated the prevailing viewpoint of the actor's audience. For example, in the confederate statue vignette, the representative's constituents predominantly believe either (a) that the statues should be taken down or (b) that the statues should remain up. We refer to conditions in which the neutral actor's audience holds a stereotypically conservative position (i.e., that statues should remain up, that abortion should be illegal, or that NFL players should stand during the national anthem) as "conservative audience" conditions (vs. "liberal audience" conditions when the audience holds the opposite views)[4]. We therefore randomly assigned participants to one of six cells in a 3 (vignette) $\times$ 2 (audience view) between-subjects design.

Participants then made an inference about the actor's underlying personal beliefs using a 7-point scale from -3 to +3, with -3 referring to a strongly held liberal view and +3 referring to a strongly held conservative view. For example, in the confederate statue scenario, -3 indicated that the representative "Strongly believes that the statues should be taken down" whereas +3 indicated "Strongly believes that the statues should be left up." In all cases, the midpoint of the scale indicated "is neutral on this issue." We expected to observe a medium effect of audience in each vignette ($d \sim .5$) and so set our target sample size to 50 per cell for this first experiment.

For this and all subsequent studies, age and gender demographics were collected at the end. Short multiple-choice attention checks (e.g., "what issue did you read about?") were also collected at the end of all studies to audit comprehension of basic details in our stimuli.

Results

A two-way ANOVA detected a significant main effect of audience condition ($F(1, 295) = 241.05$, $p < .001$, $\eta^2_G = .45$, 90% CI [.38, .50]). Participants interpreted an actor's choice not to take sides as signaling more conservative beliefs when addressing a liberal audience ($M = +1.11$, $SD = 1.09$) and more liberal beliefs when addressing a conservative audience ($M = -.89$, $SD = 1.19$). This effect of audience condition on belief inference held separately in all three vignettes (abortion: $t(95) = 11.96$, $p < .001$, $d = 2.4$,

[3] We initially sought to avoid exclusions altogether; however, we preregistered exclusions based on attention in studies conducted during the revision process in response to worries about data quality online in the wake of the COVID-19 pandemic. All reported results are robust to exclusions.

[4] Condition names were chosen for ease of exposition, rather than to test anything about liberal or conservative ideology.

Table 1
How Actors Opted to Stay Out of It in Our Study Stimuli (Scenarios: 1a-1b, 3, 5, 6a-b; Video Clips: 2a-2b; Incentivized Economic Games: 4a-b)

Exp	Actor	Issue	Staying out of it stimulus
1a	Businessman	Kneeling for national anthem	"Well, I think the issue is quite complex, and I really can't take a side on it at this time."
	State representative	Confederate statues	
	Judge	Abortion	
1b	English professor	Protesting the police	"Gosh there are so many perspectives on this issue. Truthfully, I'd rather not get into it right now."
	Cousin	COVID-19 mask mandates	"You know, this whole COVID mask thing is so complicated, and I really don't like talking politics with family."
2a	NFL owner	Kneeling for national anthem	"We aren't doing anything on that . . . there's really nothing to talk about."
2b	Former celebrity	Presidential politics	"We're neutral. I'm not taking either side. It's just uncomfortable."
3	State representative	Kneeling for national anthem	"Well, I think the issue is quite complex, and I can't really take a side at this time."
4a	Anonymous Prolific worker	Gun control	(Computer message): "Your partner declined to take a side on this issue"
4b	Anonymous MTurk worker	Gun control	(Computer message): "Your partner declined to take a side on this issue"
5	New friend	Gun control	"You know, I'd really rather not take sides on that issue right now."
			"You know, I'd really rather not take sides on that issue right now. I just don't know enough about guns or gun policy to have an opinion."
			"You know, I'd really rather not take sides on that issue right now. I think there are good arguments on both sides of the gun debate."
			"You know, I'd really rather not take sides on that issue right now. I typically try not to talk about political issues with friends."
6a	County judge	Gun control	"This is a very important and complex issue, and I'm not sure what I think. Consider me neutral."
6b	State representative	Immigration/border security	"You know, I think this is a very important issue, and personally, I just can't take a side on it at this time."

95% CI [1.90, 2.96]; anthem-kneeling: $t(100) = 8.40$, $p < .001$, $d = 1.66$, 95% CI [1.21, 2.12]; confederate statues: $t(100) = 7.1$, $p < .001$, $d = 1.41$, 95% CI [.97, 1.84]; see Figure 1). Moreover, all six scenario conditions differed from the neutral midpoint (all ps $< .001$). Specifically, participants always attributed to neutral public figures beliefs which opposed the prevailing views of their audiences (e.g., when in front of a liberal audience, staying out of a discussion about abortion rights signaled prolife beliefs, and vice versa). We also found a main effect of vignette ($F(2, 295) = 3.14$, $p = .045$, $\eta^2_G = .021$, 90% CI [.00, .050]) and a marginally significant vignette $\times$ condition interaction ($F(2, 295) = 2.77$, $p = .065$, $\eta^2_G = .018$, 90% CI [.00, .047]), which each reflect variance in the predicted effect across qualitatively different scenarios. These results do not qualify our interpretation of the main effect of audience.

Experiment 1b

Method

Experiment 1b was a preregistered replication of Experiment 1a with a larger sample in two new contexts: a department meeting of college professors and a socially-distant barbeque among extended family. Each represented an interesting potential extension of our 1a result. The college professor scenario was designed to test a case in which the public figure's profession (teaching English) does not directly intersect with the issue in question (protests against the police). The extended family scenario was designed to generalize our effect beyond public figures.

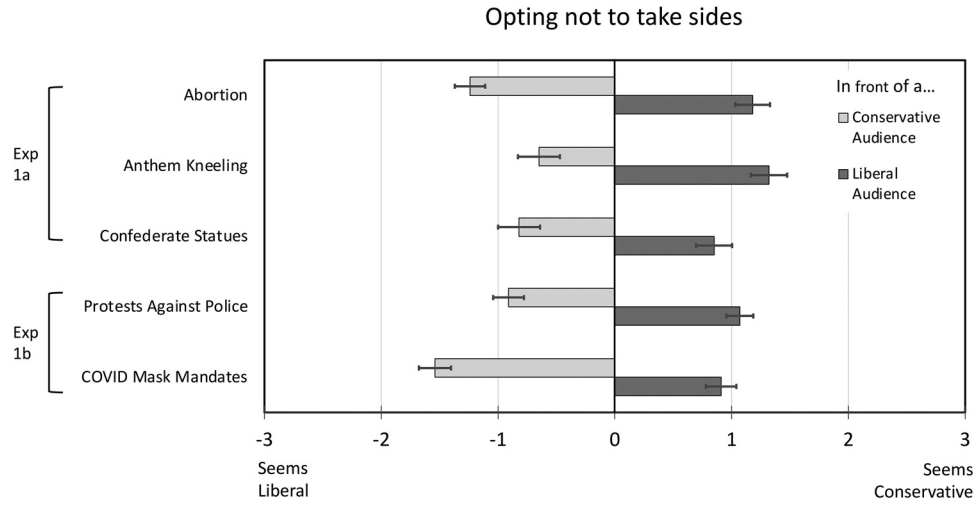
Four hundred and thirty MTurkers were recruited (mean age = 38.9, $SD = 13.3$, 46.3% female) and followed the same procedure from Experiment 1a. Following our preregistration, nine additional participants were excluded for failing a simple multiple-choice attention check. We aimed to collect 100 subjects per cell in all scenario studies conducted after 1a. Participants read one of two vignettes. One described a college English professor who is asked at a department meeting what he thinks of protests against the police. The professor responds, "Gosh there are so many perspectives on this issue. Truthfully, I'd rather not get into it right now." The other described a cousin at a family barbeque who is asked for her views on mask-wearing requirements at the local supermarket. She responds, "You know, this whole COVID mask thing is so complicated, and I really do not like talking politics with family." Audience beliefs were once again randomly assigned. In conservative audience conditions, the audience described in the scenarios (the professor's department colleagues, other family members at the barbeque) were depicted as holding stereotypically conservative beliefs (supporting police over protesters, opposing mask mandates). In liberal audience conditions, these observers held the opposite views. Participants again made inferences about the actor's personal beliefs using similar 7-point scales from -3 to +3, with -3 referring to a strongly held liberal view and +3 referring to a strongly held conservative view. 0 again indicated neutrality.

In addition to age and gender demographics, participants in Experiment 1b (and all remaining studies) also indicated their overarching political attitude on a 7-point scale from -3 "*Strongly Liberal*" to +3 "*Strongly Conservative*;" 0 indicated "In the middle" ($M = -.5$, $SD = 1.8$).[5]

[5] Across studies, we did not detect any consistent patterns of interaction between either participants' general political attitudes or their views on specific issues and their inferences about the beliefs of actors who opted not to take sides.

Figure 1

Experiments 1A and 1B: Belief Inferences by Vignette and Audience Condition



Note. Error bars represent standard errors.

Results

A two-way ANOVA detected a main effect of audience condition ($F(1, 426) = 289.51$, $p < .001$, $\eta^2_G = .40$, 90% CI [.35, .45]). Participants again interpreted an actor's choice not to take sides as signaling more conservative beliefs when addressing a liberal audience ($M = .99$, $SD = 1.28$) and more liberal beliefs when addressing a conservative audience ($M = -1.22$, $SD = 1.43$). Effects of audience held separately in each vignette (protests: $t(213) = 11.29$, $p < .001$, $d = 1.54$, 95% CI [1.23, 1.85]; COVID masks: $t(213) = 12.74$, $p < .001$, $d = 1.74$, 95% CI [1.42, 2.05]; see Figure 1). Here too, all four scenario conditions differed from the scale's midpoint (all ps $< .001$), such that participants always inferred opposition to the actor's audience from their choice not to take sides. We also found a weaker main effect of vignette ($F(1, 426) = 9.17$, $p = .003$, $\eta^2_G = .021$, 90% CI [.004, .048]) and a marginal interaction ($F(1, 426) = 2.99$, $p = .085$, $\eta^2_G = .007$, 90% CI [.00, .026]) which do not qualify our interpretation of audience effects on belief inference.

Discussion

In Experiments 1a-b, opting to stay out of it telegraphed diametrically opposed underlying views when communicated to different audiences. Actors who opt not to take sides seem liberal in front of a conservative audience but conservative in front of a liberal audience. This pattern held across actors from elected officials to academics to family members and across five different political issues.

Experiments 2a and 2b: Replications With Video Stimuli

Experiments 2a-b generalized Experiment 1a-b's belief inference results to more naturalistic stimuli: Press video clips of actors opting not to take sides in response to questions from reporters about their moral and political beliefs. In other words, participants observed actors staying out of it in the same fashion they might in the real world, via short news clips of public figures declining to respond to contentious topics. This time, participants were also asked to imagine that instead of staying out of it, the actor had sided against their viewpoint explicitly and to predict whether they would, in that case, find them more or less trustworthy.

Experiment 2a

Method

One hundred and eighty-seven participants (mean age 24.7, SD = 9.5, 73.2% female) from a university behavioral lab (i.e., one session of sign-ups) watched a ~30s clip of the owner of an NFL team (the Kansas City Chiefs) responding to political controversy about whether players should be allowed to kneel during the national anthem (https://www.youtube.com/watch?v=XHe1w-7aRno). In the clip, a reporter at a press event asks: "Can you tell us where you are with the NFL's and the Chiefs' stance on the national anthem?" The owner avoids taking a side, saying "We aren't doing anything on that," and, "there's really nothing to talk about." Between conditions, we manipulated what participants were told about the press event's audience: that the fans of the team and viewers of the news station were either "mostly conservative" (conservative audience condition) or "mostly liberal" (liberal audience condition).

Participants completed the 7-point belief inference measure used previously (-3 "the owner strongly believes players should be allowed to kneel" to +3 "the owner strongly believes players should be required to stand"). On a separate page, participants were then asked for their own personal view on the issue in a binary forced choice: "I mostly believe players should be allowed to kneel" or "I mostly believe players should be required to stand" (86% and 14% picked each option, respectively). This binary variable allowed us to assign each participant to an appropriate counterfactual for the following question.

We next asked participants to imagine that instead of opting to stay out of it, the owner had made a statement that opposed their viewpoint directly (either that he believes "players should be

allowed to kneel" or "players should be required to stand," based on their answer to the previous question). For example, a participant who thought players should be allowed to kneel considered how they would feel if the owner had said, instead of avoiding a side, that he thought players should be required to stand. Participants predicted how much more or less sincere, trustworthy, and honest (random order, $\alpha = .77$) they would find the owner on a scale from -3: "Much less" to +3: "Much more." Our two key dependent variables were belief inferences and this composite predicted trust measure. The specific trust items—trustworthiness, sincerity, honesty—were chosen to tap broader perceptions of trust as well as morally-relevant subcomponents, benevolence (sincerity of intention) and integrity (honesty/ethicality; Levine & Schweitzer, 2015). Participants also answered the same overarching political attitudes question from Experiment 1b ($M = -1.19$, $SD = 1.40$).

Results

Belief Inferences

Replicating Experiment 1, participants who witnessed the football team owner choose not to take sides while speaking to a liberal audience attributed to him more conservative underlying beliefs ($M = +.24$, $SD = 1.06$) than those who learned he was speaking to a conservative audience ($M = -.04$, $SD = .97$; $t(185) = 1.91$, $p = .058$, $d = .28$, 95% CI [−.01, .57]). One-sample t-tests comparing mean belief inferences in each condition to the midpoint of the scale detected a significant effect only in the liberal audience condition ($t(95) = 2.21$, $p = .030$, $d = .23$, 95% CI [.022, .43]) but not in the conservative audience condition ($t(90) = .43$, $p = .67$, $d = -.05$, 95% CI [−.25, .16]), which may reflect participants' prior assumption that wealthy NFL owners lean conservative at baseline. In any case, these results suggest that the actor's implied beliefs seemed relatively more conservative when addressing a liberal audience and relatively more liberal when addressing a conservative audience.

Trust

We next tested whether participants might distrust someone who implicitly opposed them by opting not to take sides more strongly than someone who explicitly opposed them. To do this, we examined the subset of participants whose personal position on the kneeling issue happened to align with that of the actor's (randomly-assigned) audience ($n = 97$). For these participants, we expected that staying out of it would seem like implicit opposition. A one-sample t test on the composite trust measure for this subset revealed a significant effect: Participants predicted that they would trust the actor more if he had espoused a view on the NFL kneeling controversy that they themselves eschewed ($M = +.33$, $SD = 1.23$; $t(96) = 2.67$, $p = .009$, $d = .27$, 95% CI [.069, .48]).

For participants who happened to disagree with the actor's audience ($n = 90$), remaining neutral signaled tacit support for the participants' viewpoint. The predicted trust question therefore invoked a different comparison; namely, between implicitly agreeing with the participants' position and explicitly opposing it. For example, would a liberal participant respond more favorably to someone who tries to stay out of it but seems, in reality, to agree with them or to someone who explicitly opposes their position? We did not have any predictions about this comparison, but interestingly, we found that even in this case, people did not trust the actor who avoided a side any more than someone who opposed them outright ($M = +.04$, $SD = 1.20$; $t(89) = .29$, $p = .77$, $d = .031$, 95% CI [−.18, .24]).

Experiment 2b

Method

Experiment 2b replicated Experiment 2a with a larger sample viewing a different video clip. In 2a, our sample size was restricted to one sitting of participants in a behavioral lab session. In 2b, we recruited 300 MTurkers (mean age = 37.1, $SD = 11.9$, 39.7% female), increasing our target to 150 per cell in light of smaller effects observed in 2a. Participants this time viewed a clip of a celebrity musician (Backstreet Boy AJ Mclean; https://www.youtube.com/watch?v=Qukk3xQM1Ac). In the clip, Mclean is approached in an airport baggage carousel by a reporter and asked for his personal political views. He responds, "We are neutral. I'm not taking either side. It's just uncomfortable." We again manipulated whether the reporter was said to represent a conservative channel with a primarily conservative audience or a liberal channel with a primarily liberal audience. Participants then inferred the celebrity's personal political beliefs on a 7-point scale from -3 "*strongly liberal*" to +3 "*strongly conservative*." 0 indicated "*politically neutral*." Participants next indicated their own general political attitudes on a 6-point scale. In this experiment only, we eliminated the midpoint from this general political attitude measure, so as to separate participants by their predominant personal viewpoint for the predicted trust question (64% leaned liberal, 36% leaned conservative). As in Experiment 2a, liberals were asked to imagine that instead of saying that he would not take either side, the celebrity had said he leans conservative. Conservatives were asked to imagine that instead of saying that he would not take either side, the celebrity had said he leans liberal. Participants predicted how much more or less sincere, trustworthy, and honest they would find the actor in this imagined alternative scenario ($\alpha = .90$).

Results

Belief Inferences

Participants who believed that the celebrity was speaking to a liberal news station attributed more conservative beliefs ($M = +.38$, $SD = .96$), and the opposite was true for those who believed he was speaking to a conservative news station ($M = -.34$, $SD = 1.09$; $t(298) = 6.15$, $p < .001$, $d = .71$, 95% CI [.48, .94]). One-sample t-tests comparing mean belief inferences in each condition to the neutral midpoint of the scale detected significant effects in both the conservative audience ($t(147) = 3.86$, $p < .001$, $d = -.32$, 95% CI [−.48, −.15]) and liberal audience conditions ($t(151) = 4.92$, $p < .001$, $d = .40$, 95% CI [.23, .57]).

Trust

We again examined first the subset of participants who shared the political view of the celebrity's audience ($n = 147$). A one-sample t test on the predicted trust measure for this subset revealed a marginally significant effect, indicating that participants would trust

the celebrity more on average if he had espoused a general political attitude that they themselves opposed instead of saying he would not take either side ($M = +.20$, $SD = 1.26$; $t(146) = 1.92$, $p = .057$, $d = .16$, 95% CI [−.005, .32]).

Participants who did not agree with the audience they read about ($n = 153$) again judged an actor whose implied position aligned with their own as no more trustworthy than one who sided against them explicitly ($M = -.03$, $SD = 1.11$; $t(152) = .32$, $p = .75$, $d = -.02$, 95% CI [−.18, .13]).

Discussion

In Experiments 2a-b, using more realistic stimuli, audience information again dictated participants' attributions of belief to actors who opt not to take sides. Specifically, in three of our four conditions, staying out of it was interpreted as tacit disagreement. In addition, among these actors' intended audience, staying out of it was generally expected to be less trustworthy than direct opposition. Meanwhile, when participants inferred that trying not to take sides meant tacit agreement with their view, they did not respond any more positively to it than to siding against their view outright.

Experiment 3: Comparing Staying Out of It to Outright Opposition

In Experiments 2a-b, participants predicted that they would trust an actor who opted not to take sides less than one opposed their viewpoint outright. However, these studies asked participants to compare staying out of it to an imagined alternative and predict trust. To eliminate potential demand characteristics and more directly examine whether staying out of it backfires relative to stating an opposing position, Experiment 3 randomized participants to read about either an actor who opts not to take sides or one who expresses outright opposition. Participants then evaluated trustworthiness and reported voting intentions.

Method

Four hundred and one MTurkers (mean age = 37.8, $SD = 12.9$, 41.1% female) read a short scenario adapted from Experiment 1a. In it, a businessman considering a run for political office holds a press conference to get to know his constituents. He is asked by a reporter whether he thinks NFL players should be allowed to kneel or required to stand during the national anthem. We experimentally manipulated both what the businessman's audience mostly believed about the kneeling issue (conservative vs. liberal audience conditions) and also whether the businessman stays out of it or sides against his audience explicitly (not-taking-sides vs. opposition conditions). In both the not-taking-sides and opposition conditions, the businessman notes that "the issue is quite complex." In the not-taking-sides conditions, he continues, "and I really cannot take a side at this time." In the opposition conditions, he continues, "but I believe that players should be (allowed to kneel/required to stand)," espousing whichever viewpoint *opposed* his audience's prevailing position. Participants were thus sorted into one cell in a 2 audience (conservative vs. liberal) × 2 response (not-taking-sides vs. opposition) between-subjects experiment.

Next, participants completed the belief inference measure from Experiment 1. They then rated trust in the businessman on 7-point agreement scales ("Strongly disagree" to "Strongly agree") with three items: "The businessman is sincere," "The businessman is trustworthy," "The businessman is honest," ($\alpha = .96$). We also included an exploratory measure meant to assess participants' willingness to vote for the businessman for political office: "If I were a voter in this district, I would consider voting for the businessman."

Finally, participants provided their personal view on the kneeling issue on a 6-point scale from "believe strongly that players should be required to stand" to "believe strongly that players should be allowed to kneel." We binary coded this variable to capture each participant's prevailing view on the issue (64% believed NFL players should be allowed to kneel; 36% believed the opposite). Participants then completed age and gender demographics and answered the same broader political attitude measure used previously. Participants leaned liberal in their overarching political attitudes ($M = -.54$, $SD = 1.84$).

Results

Belief Inferences

A two-way ANOVA with audience condition (liberal vs. conservative) and response condition (not-taking-sides vs. opposition) as factors detected a main effect of audience ($F(1, 397) = 468.81$, $p < .001$, $\eta^2_G = .54$, 90% CI [.49, .58]). This effect held separately for planned comparisons within each response condition. Unsurprisingly, when the businessman opposed a conservative audience outright by saying outright that he thought NFL players should be allowed to kneel during the national anthem, he was believed to hold more liberal views on the issue ($M = -1.62$, $SD = 1.38$) than when he opposed a liberal audience by saying the opposite ($M = +1.95$, $SD = 1.08$; $t(197) = 20.36$, $p < .001$, $d = 2.88$, 95% CI [2.49, 3.28]). Replicating our previous results, saying that he "couldn't take sides" telegraphed more liberal beliefs when addressing a conservative audience ($M = -.68$, $SD = 1.24$) and more conservative beliefs when addressing a liberal audience ($M = +1.08$, $SD = 1.21$; $t(200) = 10.18$, $p < .001$, $d = 1.43$, 95% CI [1.12, 1.74]). Although there was no main effect of actor response ($F(1, 397) = .076.$, $p = .78$, $\eta^2_G = .00$), we detected an interaction ($F(1, 397) = 54.45$, $p < .001$, $\eta^2_G = .12$, 90% CI [.074, .17]), indicating that choosing not to take sides led to less extreme belief attributions than outright opposition in both audience conditions (both ps $< .001$). In other words, as would be expected, explicit statements of disagreement signaled stronger opposition than the tacit opposition implied by staying out of it.

Trust

In this design, we again expected that saying that one "couldn't take sides" would be received differently based on whether the participant broadly agreed or disagreed with the actor's audience. For instance, participants who think that NFL players should be free to kneel during the national anthem should prefer someone who expresses support for that view in the face of an audience who thinks the opposite, as compared to someone who takes no side. In this case, taking a side clearly seems more praiseworthy than staying out of it. A more interesting and strict test concerns whether participants who *agree* with the actor's audience would respond to outright opposition more favorably than to staying out of it. For this reason, we

first examined the results for trust and voting intentions aggregated over all participants ($n = 401$), and then focused on the subset of participants ($n = 181$) who happened to share the viewpoint of the audience in their randomly assigned condition.

We did not detect an interaction between audience condition and actor response ($F(1, 397) = .0059$; $p = .93$, $\eta^2_G = .00$) and so collapse across audience conditions for ease of exposition. Among the full sample, participants saw the public figure as substantially less trustworthy when he said he could not take sides ($M = 3.48$, $SD = 1.56$) than when he opposed his audience outright ($M = 5.39$, $SD = 1.37$; $t(399) = 13.05$, $p < .001$, $d = 1.30$, 95% CI [1.09, 1.52]). Among the target subsample (i.e., those participants who agreed with the group addressed in the scenario), not taking sides was also seen as much less trustworthy than outright opposition ($M = 3.25$, $SD = 1.57$ vs. $M = 5.01$, $SD = 1.29$; $t(179) = 8.03$, $p < .001$, $d = 1.20$, 95% CI [.88, 1.52]).

Voting Intentions

We again detected no interaction ($F(1, 397) = 3.32$ $p = .069$, $\eta^2_G = .01$), and so again collapsed across audience conditions. Among the full sample, the businessman received more intended voting support for opposing his audience ($M = 4.62$, $SD = 1.92$) than for staying out of it ($M = 3.28$, $SD = 1.67$; $t(399) = 7.44$, $p < .001$, $d = .74$, 95% CI [.54, .95]). Among participants who agreed with the audience they had read about, the same trend emerged, albeit nonsignificantly ($M = 3.33$, $SD = 1.80$ vs. $M = 2.91$, $SD = 1.60$; $t(179) = 1.66$, $p = .10$, $d = .25$, 95% CI = [−.047, .54]). People were no more likely to vote for a businessman who took no side than for one who opposed their view outright: If anything, they were less likely to do so.

Discussion

In Experiment 3, observers again used audience information to attribute underlying political commitments to actors who choose not to sides. Unsurprisingly, taking no side signals weaker opposition than outright disagreement: When the businessman directly sided against his audience, he was believed to oppose them more strongly than when he declined to take sides. Nevertheless, participants were more likely to trust and no less likely to consider voting for the businessman when he sided against their viewpoint outright than when he took neither side. Indeed, even among those whom such a strategy seems tailored to placate, staying out of it may backfire. Given that doing so also incurs the opportunity cost of failing to side with the supporters of either outright position, the aggregate costs of staying out of it can be steep.

Experiments 4a-4b: Behavioral Effects on Cooperation and Partner Choice

We have argued that when staying out of it resembles strategically concealed opposition, it can provoke distrust and backfire. Experiments 4a-4b tested for effects of distrust on incentivized behavioral measures (4a: cooperation and 4b: partner choice) in an economic game. In Experiment 4a, participants signaled their beliefs on a contentious political issue to an anonymous partner and then learned that their partner had decided either to signal agreement, signal opposition, or decline to take sides. In Experiment 4b, participants signaled their beliefs and then picked between two potential partners, one who had responded with opposition and one who had declined to take sides. If staying out of it truly harms trust, it should harm cooperation and partner-choice when real money is on the line.

Notably, these experiments further generalize our findings. Our studies thus far have explored observer responses to politicians, businesspeople, celebrities, college professors, and family members who opt to stay out of it. Experiments 4a-b sought to show that our effects can arise in any interpersonal setting where moral and political issues are up for discussion and there exists some incentive to conceal a controversial perspective.

Experiment 4a

Method

Six hundred American citizens (mean age = 35.7, $SD = 12.4$, 49.7% female) were recruited from Prolific.com to participate in a study about political beliefs and cooperation. All were informed that they would be partnered with another worker, signal beliefs about an important issue, and then play a game with their partner for real bonus money.

Participants first learned the rules of a Prisoner's Dilemma game (Axelrod, 1980) and answered two multiple-choice comprehension check questions to ensure that they understood them. Participants were given two chances to answer these questions correctly before being removed for inattentiveness (prior to random assignment)[6]. We referred to this game within the study as "The Reliability Game" and to the choice options (i.e., cooperate or defect) as "Rely" and "Avoid."

Participants were then instructed that they had been partnered with another Prolific worker and that they would participate in a belief signaling exercise before playing the Reliability Game for real bonus money. In actuality, all participants were assigned to signal their beliefs first, and we randomly assigned how their partners responded. To avoid deception, partners were recruited from a separate sample ($n = 150$), and their responses were randomly matched with our main study participants for payment purposes.

Participants selected between the following two choice options: "I believe ordinary citizens should be allowed to own guns" and "I believe ordinary citizens should NOT be allowed to own guns" in a forced choice (69% and 31% chose each option, respectively). Participants were then told that their beliefs had been shared with their partner and that their partner had been given the option either to signal beliefs back or to decline to take a side on the issue of gun ownership. Note that participants were fully aware that their partners were given the freedom to decline to take sides by the experimenter. Partner responses were randomly assigned to be either agreement (responding by selecting the same statement as the participant), opposition (responding by selecting the opposite statement as the participant), or not-taking-sides (responding by selecting the option "decline to take sides"). After seeing their partner's response, participants reported trust on the same scale used in Experiment 3 ($\alpha = .92$) and selected whether to cooperate or not in an incentive-compatible Prisoner's Dilemma. Participants were also asked to make belief inferences, attributing to their partner convictions about

[6] We stopped recruiting when we reached 600 participants who passed the attention check and completed the entire study. We set our target sample to 200 per cell for Experiments 4a-b to increase power and account for binary DVs.

gun ownership on a scale from -3 "My partner definitely believes that ordinary people should be allowed to own guns" to +3 "My partner definitely believes that ordinary people should NOT be allowed to own guns"; 0 indicated "My partner's beliefs on this issue are neutral." In Experiments 4a-4b, we collected an additional exploratory item capturing the extent to which participants thought reasonable people could disagree about the issue of gun control but did not find any consistent interactions. Participants leaned liberal in their overarching political beliefs ($M = -.65$, $SD = 1.72$).

Results

Belief Inferences

Following our preregistered plan, this time we recoded participants' belief inferences onto a -3 to +3 "perceived opposition" scale, such that numbers less than 0 always indicated inferred agreement and numbers greater than 0 always indicated inferred opposition. To do this, we simply reverse-coded responses from participants who indicated that ordinary people should be allowed to own guns.

In this experiment, participants were randomly assigned to see their partner indicate either outright agreement, outright opposition, or opt not to take sides. Unsurprisingly, a one-way ANOVA detected a significant main effect of condition on perceived opposition ($F(2, 597) = 659.92$, $p < .001$, $\eta^2_G = .69$, 90% CI [.66, .71]). Participants in the agreement condition inferred that their partner agreed with their viewpoint ($M = -2.30$, $SD = 1.16$), while those in the opposition condition inferred that their partner opposed it ($M = +2.42$, $SD = 1.21$). As predicted, participants in the not-taking-sides condition also inferred that their partner opposed their viewpoint ($M = +.60$, $SD = 1.54$; one-sample $t(199) = 5.46$, $p < .001$, $d = .39$, 95% CI [.24, .53], vs. the scale's midpoint) albeit much less strongly than in the opposition condition ($t(396) = 13.18$, $p < .001$, $d = 1.32$, 95% CI [1.10, 1.54]).

Trust

We preregistered predictions that participants would trust their partner most in the agreement condition, that trust would be weaker in both the not-taking-sides and opposition conditions, and that not taking sides would garner no more trust than outright opposition. The results corroborated these predictions. A one-way ANOVA detected a significant omnibus effect of condition ($F(2, 597) = 44.88$, $p < .001$, $\eta^2_G = .13$, 90% CI [.09, .17]). Participants trusted their partner more in the agreement condition ($M = 5.27$, $SD = 1.10$) than in the not-taking-sides condition ($M = 4.20$, $SD = 1.28$; $t(400) = 8.96$, $p < .001$, $d = .89$, 95% CI [.69, 1.10]). They also trusted their partner more in the opposition condition ($M = 5.07$, $SD = 1.22$) than in the not-taking-sides condition ($t(396) = 6.98$, $p < .001$, $d = .70$, 95% CI [.50, .90]). Interestingly, in terms of attitudinal trust, agreement and opposition differed only marginally ($t(398) = 1.67$, $p = .097$, $d = .17$, 95% CI [−.030, .36]; See Figure 2).

Cooperation

We preregistered predictions that cooperation would be strongest in the agreement condition, that both not taking sides and opposition would see less cooperation, and that not taking sides would engender no more cooperation than outright opposition. The results again corroborated our predictions.

A Chi-Square test of independence detected an omnibus effect of condition on cooperation ($\chi^2(2\ df) = 26.84$, $p < .001$). Rates of cooperation were highest in the agreement condition (90.1%), followed by the opposition condition (72.7%), followed by the not-taking-sides condition (70.5%). Differences in cooperation between the agreement condition and both the opposition ($\chi^2(1\ df) = 20.01$, $p < .001$, $OR = 3.41$, 95% CI [1.95, 5.96]) and not-taking-sides conditions ($\chi^2(1\ df) = 24.45$, $p < .001$, $OR = 3.81$, 95% CI [2.19, 6.62]) were significant. There was no difference in cooperation between the not-taking-sides and opposition conditions ($\chi^2(1\ df) = .24$, $p = .62$). If anything, participants were less likely to cooperate with a partner who declined to take sides on a hot-button political issue than with one who opposed them outright ($OR = .90$, 95% CI [.58, 1.39])[7]. See Figure 3.

Experiment 4b

Method

Experiment 4a found that both declining to take sides and outright opposition harmed trust and cooperation relative to agreement. We next examined how such behaviors impact partner choice: How would participants respond when given the opportunity to choose between a partner who opposed them outright and one opted not to take sides?

Experiment 4b followed a broadly similar design. 402 MTurkers (mean age = 39.8, $SD = 12.4$, 43.8% female) learned the rules of a Prisoner's Dilemma (again, "The Reliability Game") and completed the same comprehension check questions with the same exclusion criteria. This time, participants learned that they had been randomly grouped with two other MTurkers, and that they would select which of the two to partner with for The Reliability Game.

To inform their choice, participants again engaged in a belief exchange exercise. Participants signaled their beliefs about gun control, communicating either that ordinary citizens should or should not be allowed to own guns (69% and 31% chose each option, respectively).

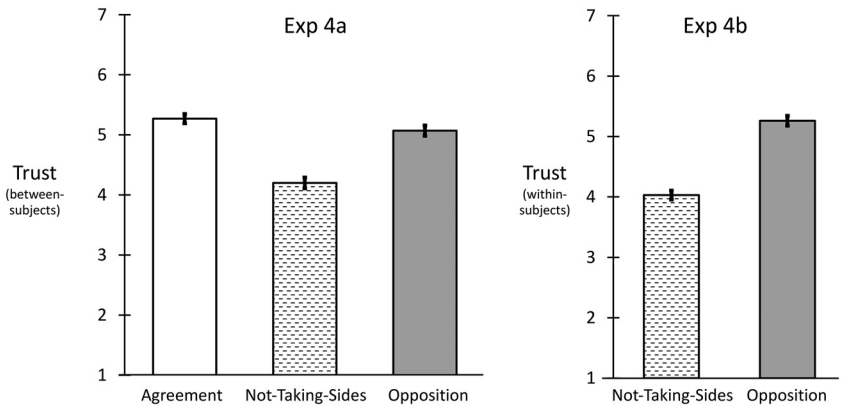
One potential partner responded with outright opposition (selecting the opposite statement from the participant) while the other stayed out of it (by declining to take sides). The answer choices and language used for this exercise were identical to that used in 4a.

Participants rated the same trust items used previously for both partners ($\alpha_{\text{opposition-partner}} = .89$, $\alpha_{\text{neutral-partner}} = .95$) and, critically, selected which of the two partners to play the Reliability Game with. Partner choice served as our primary dependent variable. Participants then chose whether to cooperate or defect with their selected partner and completed the usual demographic questions. Participants leaned liberal in the overarching political beliefs ($M = -.51$, $SD = 1.81$).[8]

[7] Interestingly, rates of cooperation were relatively high across all conditions. *Post hoc*, we believe that this reflects the specific payout amounts we chose for the game. This pattern does not impact our interpretation of the results.

[8] We omitted belief inferences in this study only to ensure that making belief inferences salient by measuring them is not a necessary ingredient for our trust effects. We suspect that participants make the same general attributions whether we measure them or not.

Figure 2
Experiments 4A and 4B: Attitudinal Trust Between- and Within- Subjects, in Response to Signals of Agreement (4a Only), Opposition, or Not-Taking-Sides



Note. Error bars represent standard errors.

Results

Trust

Once again, participants indicated greater trust toward the partner who opposed their view on gun control outright (M = 5.26, SD = 1.17) than the one who did not take sides (M = 4.03, SD = 1.46; paired $t(401) = 13.17$, $p < .001$, $d = .66$, 95% CI [.55, .77]; See Figure 2).

Partner Choice

61.2% of participants opted to play the prisoner's dilemma game with the partner who opposed their views on gun control outright, while only 38.8% opted to play with the partner who declined to take sides. A proportion test comparing this preference for outright opposition to chance revealed a significant effect ($\chi^2(df = 1) = 19.7$, $p < .001$). Moreover, logistic regression revealed that the preference for opposition (over not-taking-sides) was predicted by participant-level differences in attitudinal trust between the two responses ($B = .92$, $SE = .11$, Wald $Z = 8.51$, $p < .001$; see Figure 3)[9].

Discussion

Experiments 4a and 4b provide further evidence for our belief inference and trust effects, and extend them in two ways. First, they show that our effects on trust are not restricted to public figures giving public interviews: When an anonymous partner's decision not to take sides signaled tacit disagreement with a participant's viewpoint, trust suffered. Second, they extend our results to an incentive-compatible cooperation paradigm. Declining to take sides provokes distrust that influences behavior, not just attitudes and intentions.

Note that in 4a, although participants trusted partners who opposed their views outright more than those who did not take sides, cooperation rates were similar across the two conditions. Post hoc, we suspect that cooperation decisions in these two conditions may be driven by different judgment processes: Whereas actors who stay out of it may see lower rates of cooperation because they are distrusted, ideological opponents may see lower rates of cooperation because they are disliked. Still, the results from 4b paint a more definitive picture. When asked to decide which they would rather have as a cooperative partner, participants preferred those who endorse views that they themselves eschew to those who opt not to take sides.

Experiment 5: The Role of Justifications for Staying Out of It

Our experiments focus on how observers respond to actors who opt not to take sides, but as we have noted, such public stances may reflect a variety of underlying dispositions. For example, an actor may choose not to take sides because they wish to avoid opposing their audience, because they feel sympathetic to both sides of an argument, or because they lack any opinion at all. Experiment 5 tested how observers would respond to different potential justifications for staying out of it. Broadly, we expected our prior results to be robust but that the size of our effect might be influenced by the type of justification offered. In discussions of heated political issues, the presence of impression management incentives might activate suspicions about the motives behind an avoidant stance, regardless of how it is justified. Nevertheless, we were curious whether certain justifications would soften distrust more than others.

Experiment 5 also sought to further generalize our effects beyond high stakes declarations made by public figures: Participants made judgments about the side-taking behavior of a new friend at an informal social gathering.

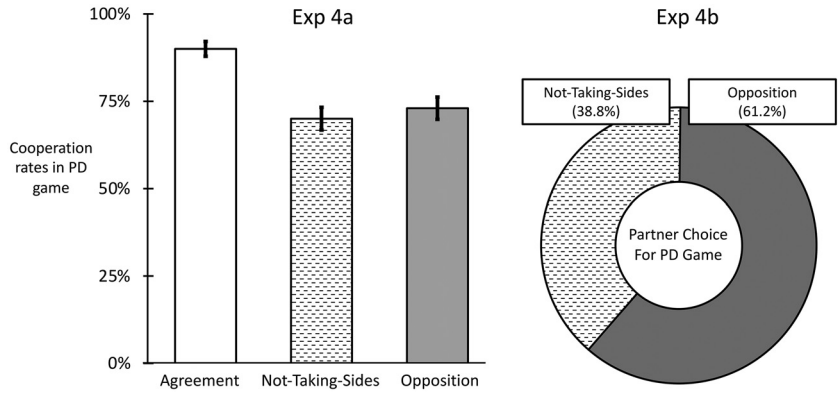
Method

Five hundred forty-eight MTurkers (mean age = 39.6, SD = 13.1, 48.7% female) were recruited for a short scenario study.

[9] Conditional on their choice of partner, participants' subsequent cooperation rates were similar whether playing the PD game with the opposition partner (71%) or the not-taking-sides partner (72%; $\chi^2(df = 1) = .001$, $p = .98$).

Figure 3

Experiments 4A and 4B: Incentivized Cooperation and Partner Choice for a Prisoner's Dilemma Game, In Response to Signals of Agreement (4a Only), Opposition, or Not-Taking-Sides



Note. Error bars represent standard errors.

Participants began by rating their beliefs about gun control by selecting between two statements: "I am typically AGAINST most gun control efforts" or "I am typically IN FAVOR of most gun control efforts" (26% and 74% chose each option, respectively). We used this measure to ensure participants were presented with a scenario in which the actor's audience shared *their* beliefs. By doing so, we could more easily compare staying out of it to outright opposition.

Participants were then randomly assigned to one of five versions of a scenario about a social gathering among friends. Participants imagined meeting a new acquaintance at a small group hang-out. In the scenario, the group begins discussing the state of U.S. politics and the issue of gun control specifically. All participants imagined that their friends by and large shared their views about gun control but that they did not yet know what the new acquaintance believed.

We manipulated how the acquaintance responded when asked to weigh in. Four conditions depicted staying out of it and the fifth depicted outright opposition for comparison. In the not-taking-sides-baseline condition, the acquaintance simply says: "You know, I'd really rather not take sides on that issue right now." In three additional conditions, the acquaintance says the same thing but also provides further justification for staying out of it (not-taking-sides-ignorant condition: "I just do not know enough about guns or gun policy to have an opinion"; not-taking-sides-ambivalent condition: "I think there are good arguments on both sides of the gun debate"; not-taking-sides-principled: "I try not to talk about politics with friends"). We note that in our follow-up study (which appears after study 6b), these were the four most common categories of response from real people staying out of it in social scenarios, with the plurality providing no justification at all. A fifth condition depicted outright opposition to the participant and their friend group: "You know, all things considered, I personally think (do not think) more gun control would be a good thing for our country."

After reading, participants completed our usual belief inference measure (-3: "*Strongly supports gun control efforts*"; 0: "*Is neutral on this issue*"; +3: "*Strongly opposes gun control efforts*") and the same trust items used previously ($\alpha = .91$). In this study, we also included two measures of perceived competence as exploratory measures ($r = .93$); "This person is informed about the issue of gun control" and "This person is knowledgeable about the issue of gun control." Participants then answered a basic comprehension check by selecting whether the scenario was about gun control, police protests, or green energy. One additional participant failed and was excluded. Participants leaned liberal in their overarching attitudes ($M = -.55$, $SD = 1.8$).

Results

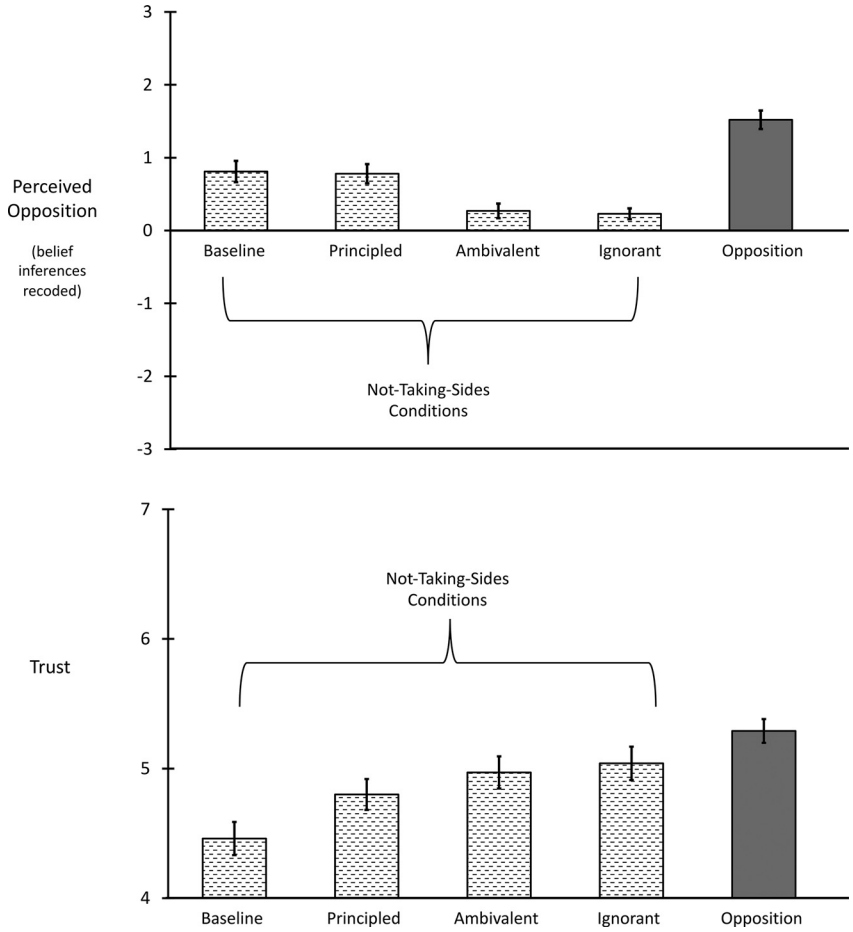
Results for belief inferences and trust across all conditions from Experiment 5 are displayed in Figure 4 In the interest of brevity, we focus here on the most critical tests of our theory (primarily comparing different justifications for staying out of it to outright opposition), but significance tests for all comparisons are available in the supplement.

Belief Inferences

Following our preregistered plan, we recoded participants' belief inferences onto a perceived opposition scale (-3 →+3), with higher numbers indicating ascriptions of oppositional beliefs. A one-way ANOVA predicting perceived opposition detected a significant omnibus effect of condition ($F(4, 543) = 19.58$, $p < .001$, $\eta^2_G = .13$, 90% CI [.081, .16]). As previously, the outright opposition condition provoked the strongest attributions of opposition ($M = 1.52$, $SD = 1.33$) compared to all other conditions. It was followed by the not-taking-sides-baseline ($M = .81$, $SD = 1.53$; $t(218) = 3.67$, $p < .001$, $d = .50$, 95% CI [.23, .76]) and -principled ($M = .78$, $SD = 1.38$; $t(217) = 4.04$, $p < .001$, $d = .55$, 95% CI [.28, .82]) conditions, which differed significantly from opposition but not from one another; and then by the -ambivalent ($M = .27$, $SD = 1.05$; $t(216) = 7.17$, $p < .001$, $d = 1.04$, 95% CI [.76, 1.33]) and -ignorant ($M = .23$, $SD = .78$; $t(219) = 8.8$, $p < .001$, $d = 1.19$, 95% CI [.90, 1.47]) conditions, which differed from all other conditions but not from each other. In line with our theory and prior results, participants inferred opposition from all not-taking-sides conditions relative to the scale's neutral midpoint ($ps < .001$).

Figure 4

Experiment 5: Perceptions of Opposition (−3 → +3; Top Panel) and Attitudinal Trust (1–7; Bottom Panel) From Potential Justifications for Staying Out of it



Note. Participants continue to perceive opposition and discount trust (relative to outright opposition) in the presence of explicit reasons for not taking sides. Error bars represent standard errors.

Trust

A one-way ANOVA predicting trust detected a significant omnibus effect of condition ($F(4, 543) = 6.80$, $p < .001$, $\eta^2_G = .048$, 90% CI [.018, .075]). Participants trusted outright opposition the most ($M = 5.29$, $SD = .96$), with opposition differing from three of the four not-taking-sides conditions: baseline ($M = 4.46$, $SD = 1.34$; $t(218) = 5.31$, $p < .001$, $d = .72$, 95% CI [.44, .99]), principled ($M = 4.80$, $SD = 1.25$; $t(217) = 3.31$, $p = .001$, $d = .45$, 95% CI [.18, .72]), and ambivalent ($M = 4.97$, $SD = 1.30$; $t(216) = 2.10$, $p < .037$, $d = .28$, 95% CI [.017, .55]). Ignorance ($M = 5.04$, $SD = 1.35$; $t(219) = 1.62$, $p = .11$, $d = .22$, 95% CI [−.047, .48]) differed directionally but nonsignificantly from outright opposition. Both the ambivalent and ignorance conditions saw significantly more trust than baseline where no justification was provided. These results suggest that some forms of additional justification may shrink, but not necessarily close, the gap in trust between not taking sides and outright moral opposition. Even pleading ignorance earns no more trust than outright opposition.

Competence

A one-way ANOVA predicting perceived competence detected a significant omnibus effect of condition ($F(4, 543) = 37.66$, $p < .001$, $\eta^2_G = .22$, 90% CI [.16, .26]). Participants viewed the new friend as similarly knowledgeable and informed in the opposition condition ($M = 4.19$, $SD = 1.3$) relative to the principled ($M = 4.10$, $SD = .94$; $t(217) = .55$, $p = .58$) and baseline ($M = 4.33$, $SD = 1.14$; $t(218) = .88$, $p = .38$) conditions. Unsurprisingly, the ignorance condition, in which distrust was weakest, also entailed a rather large competence penalty relative to opposition ($M = 2.79$, $SD = 1.44$; $t(219) = 7.57$, $p < .001$, $d = −1.02$, 95% CI [−1.30, −.74]) and to all other conditions ($ps < .001$). Unexpectedly, the friend who justified their opting not to take a position in terms of sympathy for both sides was seen as more informed and knowledgeable relative to outright opposition ($M = 4.69$, $SD = 1.31$; $t(216) = 2.87$, $p = .004$, $d = .39$, 95% CI [.12, .66]), albeit less trustworthy.

Discussion

Experiment 5 generalizes our prior effects across a host of potential justifications in the context of a meeting a new friend: Outright opposition earned stronger trust than any of the justifications for staying out of it we tested. Appealing to ignorance as a reason for not taking sides came closest to eliminating the trust penalty documented in prior studies, but it also exacted a substantial hit to perceived competence.

Still, the results of this experiment provide reason to hope that adequately explaining one's reasons for staying out of it might mitigate its costs to some extent. For example, although simply noting the merits of both sides was still seen as less trustworthy than outright opposition, perhaps one could more articulately explain the nuances of one's moderate position in a way that might diminish distrust and signal knowledgeability. However, it may prove difficult to explain one's views effectively in the heat of a charged conversation about a contentious topic. Moreover, such a strategy would likely require the actor to engage in precisely the discussion they may have hoped to avoid by staying out of it in the first place. As shown in Studies 2a-2b, even for public figures trained in public relations, actual attempts to stay out of it are often glib, off-the-cuff declinations to take sides, and such expressions, we demonstrate, often backfire.

Experiments 6a-6b: Strategic Attributions Drive Negative Responses to Those Who Try Not to Take Sides

We have argued that opting not to take sides can undermine trust and harm cooperation when it resembles strategically concealed opposition. However, not all contexts will lead to strategic attributions. Experiments 6a and 6b sought to demonstrate boundary conditions which contextualize our theory and rule out multiple potential alternative explanations. Experiment 6a investigated whether perceptions of opposition and associated distrust would be lessened if the actor lacks incentives for impression management. Experiment 6b provided observers with an explicit private signal of the actor's moderate beliefs and tested whether distrust would be attenuated as a result.

These moderations would be consistent with our theory but inconsistent with other reasons people may dislike and distrust attempts to stay out of it. For example, if people dislike not taking sides because they find nonresponsiveness uncooperative (e.g., Grice, 1975), orthogonal manipulations of actor incentive or private signals of moderate belief should have no effect across cases in which what the actor actually says is held constant. If people dislike staying out of it because it signals moral conflict or apathy (Critcher et al., 2013; Zlatev 2019), a truly moderate position should, if anything, provoke *greater* distrust than seemingly concealed convictions. By contrast, we predicted that removing social incentives (6a) or wiping out the inference of concealed opposition (6b) should attenuate the costs of staying out of it.

Experiment 6A

Method

Experiment 6a was designed to test whether distrust in an actor who opts not to take sides would be lessened if the actor lacks incentives for impression management. 499 MTurkers (mean age = 39.3, *SD* = 12.8, 46.1% female) participated in a scenario study. At the outset, we asked participants for their beliefs about gun control, specifically whether ordinary citizens should be allowed to own assault weapons (40% said yes, 60% said no). As in Experiment 5, we used this question to ensure that participants always read a scenario in which the neutral actor's audience shared their overarching view on the issue.

Participants then read a short vignette about an elected county judge approaching the end of his term. In a public interview with either a predominantly liberal or conservative audience (according to the participant's own view), the judge is asked about his opinion on gun control. In all conditions, the judge says, "This is a very important and complex issue, and I'm not sure what I think. Consider me neutral." Participants were assigned to one of two versions of this scenario. In the high-incentives condition, the judge has decided to seek reelection and is giving an interview two weeks before the election. In the low-incentives condition, the judge has decided not to seek reelection and is giving an interview two weeks before retirement.

Participants again attributed beliefs on a 7-point scale, from -3: "*Believes strongly that ordinary citizens should be allowed to own assault weapons*," to +3 indicating, "*Believes strongly that ordinary citizens should NOT be allowed to own assault weapons*;" 0 indicated "*is neutral on this issue*." Participants also completed our usual trust items ($\alpha = .96$). We predicted that participants would see the judge in the high-incentives condition, who is seeking reelection, as more strategically motivated than the judge in the low-incentives condition, who is planning to retire. Consequently, we predicted that the former would seem to be strategically concealing oppositional beliefs, but the latter might not, and that the former would seem less trustworthy as a result. Participants leaned liberal in their overarching political attitudes ($M = -.39$, $SD = 1.77$).

Results

Belief Inferences

We again recoded the belief inference measure onto a 7-point perceived opposition scale, which captured the extent to which the neutral actor's beliefs seemed to oppose those of the participant (and those of the audience in the scenario that they read). We predicted that, in the eyes of observers, opting not to take sides would more strongly resemble strategically concealed opposition in the presence of incentives for impression management versus in the absence of such incentives. In line with this prediction, participants treated the judge's statement as a stronger signal of opposition when he had high versus low incentives for impression management ($M = +.45$, $SD = 1.50$ vs. $M = -.05$, $SD = 1.34$; $t(497) = 3.92$, $p < .001$, $d = .35$, 95% CI [.17, .53]).

Trust

Moreover, not taking sides seemed less trustworthy in the high versus the low incentives condition ($M = 3.19$, $SD = 1.70$ vs. $M = 3.86$, $SD = 1.70$; $t(497) = 4.42$, $p < .001$, $d = .40$, 95% CI [.22, .57]).

Process Evidence

Finally, we predicted that the effect of incentive-condition on trust would be mediated by perceived opposition. This, too, was substantiated. Mediation analysis with 10,000 bootstrapped samples revealed a significant indirect effect of condition (0 = low incentives, 1 = high incentives) on trust via perceived opposition ($B = -.17$, $SE = .05$, 95% CI [−.28, −.08]).

Experiment 6B

Method

On our account, if an actor's public decision not to take sides seems to align with a privately held moderate or neutral position (i.e., if it signals no underlying commitment one way or the other), it should seem neither strategic nor untrustworthy. Experiment 6b was designed to test this prediction.

Five hundred and twenty-five MTurkers (mean age = 36.1, *SD* = 11.6, 41.9% female) read a short scenario about a prominent businesswoman running for office in a large diverse state. On the campaign trail, the businesswoman is asked to weigh in as to whether she supports increased spending on security at the U.S./Mexico border. We again wanted to compare to outright opposition, so, as in Experiments 5 and 6a, participants always read about a campaign rally in a district where voters predominantly shared *their* viewpoint. Participants indicated their personal view on a 6-point scale from "*Strongly support increased spending on border security*" to "*Strongly oppose increased spending on border security*." Dichotomizing this variable, 46% supported and 54% opposed increased spending on border security. Participants who supported increased spending then read a scenario in which the potential supporters at the rally were predominantly conservative, while participants who opposed increased spending read a scenario in which the potential supporters at the rally were predominantly liberal.

Participants were assigned to one of three conditions. In all cases, the businesswoman is asked for her view on border security. She always responds by saying that the issue of border security is "very important." In the opposition condition, she opposes her audience (and the participants' personal viewpoint), by also saying either "I oppose increased spending on border security" (when addressing conservatives) or "I support increased spending on border security" (when addressing liberals). In the not-taking-sides condition, she instead says that she "cannot take a side on it at this time." To tease apart whether participants object to middle-ground positions per se or to the attribution of strategic concealment, we also included a not-taking-sides neutral beliefs condition, which was identical to the other not-taking-sides condition, except that participants were also given private information from a conversation that the businesswoman later had with her husband where she says, "you've known me for years, and you know that I've always been neutral in the debate over border security and immigration." Although providing observers with an unambiguous private signal of truly neutral beliefs (i.e., no commitment one way or the other, a middle-ground position) is rarely possible in the real world, it is theoretically illuminating to know whether participants would display a similar distrust if they did not attribute to them strategically concealed attitudes.

Participants were then asked to make inferences about the businesswoman's beliefs on a scale from -3 "*strongly opposes increased spending on border security*" to +3 "*strongly supports increased spending on border security*;" 0 again indicated "*is neutral on this issue*."

Participants answered the same trust items (α = .97) and the same voting intentions items from Experiment 3. Participants leaned liberal in their overarching attitudes ($M = -.44$, $SD = 1.73$).

Results

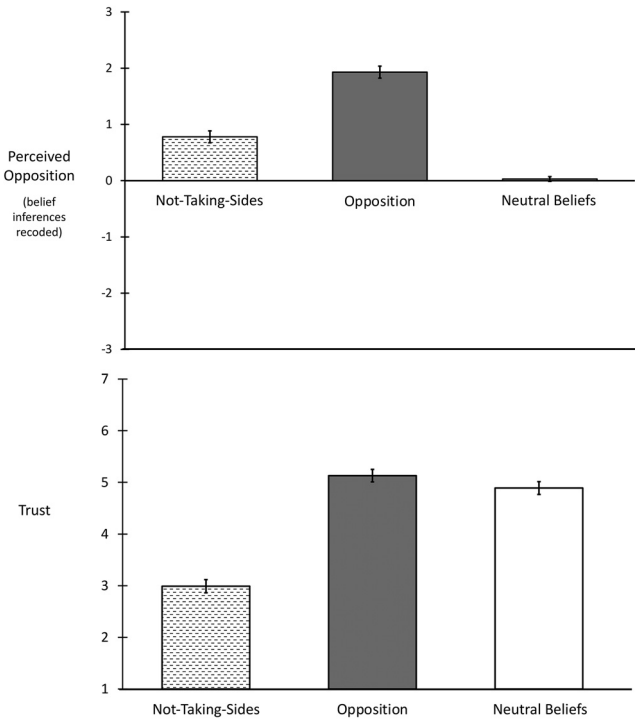
Belief Inferences

Following our preregistered plan, we again recoded participants' belief inferences onto a -3 to +3 "perceived opposition" scale, such that larger numbers always indicated greater inferred opposition. A one-way ANOVA predicting perceived opposition detected a significant omnibus effect of condition ($F(2, 522) = 114.44$, $p < .001$, $\eta^2_G = .30$, 90% CI [.25, .35]). Unsurprisingly, the businesswoman was believed to oppose her audience more strongly in the outright opposition condition ($M = 1.93$, $SD = 1.41$) than in either of the other conditions (not-taking-sides $M = .78$, $SD = 1.40$; $t(347) = 7.66$, $p < .001$, $d = .82$, 95% CI [.60, 1.04]; neutral beliefs $M = .03$, $SD = .55$; $t(350) = 16.70$, $p < .001$, $d = 1.78$, 95% CI [1.53, 2.03]). Comparing those two conditions, the businesswoman seemed to oppose her audience more strongly in the not-taking-sides condition than in the neutral beliefs condition ($t(347) = 6.62$, $p < .001$, $d = .71$, 95% CI [.49, .92]). In both the outright opposition condition and the not-taking-sides condition, perceived opposition differed from the neutral midpoint (ps $< .001$). As we predicted, a private signal of neutral beliefs wiped out the inference that not taking sides concealed underlying opposition. We next examined whether appearing to hold genuinely neutral beliefs would eliminate the negative downstream effects of not taking sides (see Figure 5).

Trust

A one-way ANOVA detected an omnibus effect of condition ($F(2, 522) = 93.45$, $p < .001$, $\eta^2_G = .26$, 90% CI [.21, .31]). As

Figure 5
Experiment 6B: Perceptions of Opposition (−3 to +3; Top Panel) and Attitudinal Trust (1 to 7; Bottom Panel) by Condition



Note. Error bars represent standard errors.

previously, the businesswoman was seen as more trustworthy for siding against her audience (and the participant) than for staying out of it (M = 5.21, SD = 1.57 vs. M = 3.04, SD = 1.71; $t(347)$ = 12.32, $p < .001$, d = 1.31, 95% CI [1.09, 1.55]). As predicted, adding a signal of privately neutral beliefs increased trust (M = 5.02, SD = 1.63; $t(347)$ = 11.05, $p < .001$, d = 1.18, 95% CI [.95, 1.41]). The difference between opposition and neutral beliefs conditions was not significant ($t(350)$ = 1.13, p = .26, d = .12, 95% CI [−.09, .33]). Both the opposition and neutral beliefs conditions saw greater trust than the not-taking-sides condition (in which participants attributed concealed opposition; see Figure 5).

Voting Intentions

A one-way ANOVA detected a significant omnibus effect of condition ($F(2, 522)$ = 21.0, $p < .001$, η^2_G = .075, 90% CI [.041, .11]). In this case, the businesswoman was more likely to receive voting support in the neutral beliefs condition (M = 3.97, SD = 1.57) than in either the not-taking-sides condition (M = 2.86, SD = 1.64; $t(347)$ = 6.43, $p < .001$, d = .69, 95% CI [.47, .90]) or the opposition condition (M = 3.07, SD = 1.85; $t(350)$ = 4.87, $p < .001$, d = .52, 95% CI [.31, .73]). We did not detect a difference between the not-taking-sides condition and outright opposition to the participant's view ($t(347)$ = 1.14, p = .26, d = .12, 95% CI [-.09, .33]). In the absence of a private signal of truly neutral beliefs, staying out of it garnered no more voting support than opposing the participant's view outright.

Process Evidence

To examine our proposed process further, we focused on the comparison between not-taking-sides and neutral beliefs conditions and fit an additional, exploratory serial mediation model. This model treated neutral beliefs versus not-taking-sides as the independent variable (coded 0 and 1, respectively), inferred opposition as mediator 1, trust as mediator 2, and voting intentions as the outcome variable. This indirect pathway was significant (B = −.25, SE = .06, 95% CI [−.37, −.15]), consistent with our theorizing that staying out of it backfires specifically when it resembles concealed opposition, which in turn harms trust and erodes voting support.

Discussion

In Experiment 6a, when incentives for impression management were weakened, and in Experiment 6b, when not taking sides seemed to align with private neutral beliefs, distrust of those who opt to stay out of it was diminished. These moderations are consistent with our claim that opting not to take sides draws particular scorn when it seems to conceal hidden opposition. However, they are inconsistent with alternative accounts of our effects based on dislike of moral ambivalence or of nonresponsive statements per se.

Although these boundary cases are theoretically important, we note that in the real world, actors who try to stay above the fray may struggle to effectively signal an absence of underlying beliefs, or to communicate an absence of reputational incentives, especially in high-stakes social contexts. Indeed, in practice, actors may find it difficult to send unambiguous private signals of "true" neutrality to observers, and they may be often assumed to have strategic motives and judged more negatively as a result.

Follow-up Study: Actors Misunderstand the Costs of Staying Out of It

We have suggested that, despite its costs, opting not to take sides seems intuitively appealing. To substantiate this claim and to provide evidence that people may misunderstand the reputational consequences of staying out of it, we conducted a follow-up study. Shifting from the observer's perspective to the actor's, we sought to show that when facing ideologically hostile audiences, people often prefer not to share their opinions and, more specifically, that they expect simply expressing a preference not to take sides to provoke less distrust than outright opposition. This follow-up study was also designed to explore what sorts of expressions people spontaneously generate when trying to avoid taking sides and to ensure that the stimuli used in our prior experiments resemble what people actually say when they attempt to stay out of it.

Method

We ran two identical versions of this study: One with participants from a business school's behavioral lab (n = 203, mean age = 23.5, 67.5% female) and the other with MTurk workers (n = 292, mean age = 40.4, 42.5% female). Each survey was separately preregistered but with identical analysis plans. As results were similar across these two populations, we report analyses pooling across them (total N = 495). Separate analyses can be found in our supplemental materials and yield the same conclusions.

At the outset of the survey, participants were asked to list a hot-button social issue currently up for debate in America (e.g., on the news, on social media, in the workplace). Answers covered a variety of contemporary issues including COVID-19 policy, racial justice, abortion rights, and so forth. Participants then read a short workplace scenario designed to probe their intuitions about taking (or not taking) sides in a contentious discussion, given the explicit goal of building trust. The text of the scenario read:

> *You are being considered for a promotion to a position of leadership at work. In order to be selected, you need the people you work with to like and trust you. But you also know that you disagree with your coworkers about a hot-button political topic (the one you listed on the prior page).*
>
> *One day, you are sitting at lunch with your coworkers and the issue comes up. Your coworkers are talking about their opinions, and although you haven't said anything, you know you disagree with them.*
>
> *At some point, one of your coworkers turns to you and says, "Well, what do you think about all of this?"*

After reading, participants were reminded to consider that they needed others in this setting to trust them, and then asked what they would be most likely to do in this situation: (a) Disagree with the group, (b) Try to stay out of it, (c) Agree with the group. Participants were also asked to report exactly what they would say. The multiple-choice question probed people's intuitive preferences for taking (or not taking) sides, while the written-response question explored how people might verbalize preferences to stay out of it. Next, and on a separate page, we asked all participants to imagine that they had in fact decided to try to stay out of it and to select which of the following two strategies would be more

effective for building liking and trust: (a) Try to keep it short: Just say I prefer to stay out of it, or (b) Try to explain more deeply why I prefer to stay out of it. This third question captured people's general sense of whether staying out of it requires deeper justification to be considered socially acceptable.

Finally, participants reported their age, gender, and political attitudes ($M = -.51$, $SD = 1.69$). Three additional participants were excluded for giving nonsensical answers to the written-response question, as determined by hypothesis-blind RAs.

Results

Looking first at answers to the "what would you do" question, 62.8% of participants indicated that they would try to stay out of it, a larger proportion than both other choice options combined (disagree with the group 32.3%, agree with the group 4.8%; $\chi^2(df = 1) = 32.1$, $p < .001$). People seem to see staying out of it as an intuitively appealing strategy for building trust with an audience hostile to their viewpoint.

When people opt to "stay out of it," what might they say? To investigate, we turned next to participants' written reports, enlisting the help of two hypothesis-blind RAs to code data and a third to arbitrate disagreements. As an initial check for data quality, RAs coded whether each participant wrote a response which matched their selection on the prior multiple-choice question. Did those who indicated a preference to take (or not to take) sides write responses which shared (or withheld) their opinion? Ninety-four percent matched. Further coding focused on written responses which tried to stay out of it, including both responses from participants who indicated an explicit preference to stay out of it and wrote responses to match ($n = 290$), as well as responses from participants who indicated a preference to take sides but wrote responses which clearly avoided doing so ($n = 9$).

Next, RAs sorted all such not-taking-sides responses ($N = 299$) into one of four categories identified a priori as possible ways people might express and justify their choice to stay out of it. These categories matched the conditions used in Experiment 5. The first captured simple expressions of preference not to take sides without further justification (e.g., "I'd rather not get involved in this issue"). The second captured expressions justified by a general principle or rule (e.g., "I'd rather not get involved in this issue because as a rule I do not talk politics at work"). The third captured expressions justified by ignorance or insufficient information (e.g., "I'd rather not get involved in this issue because I do not know enough about it to have an opinion"). The fourth captured expressions justified by a belief that both sides have merit (e.g., "I'd rather not get involved in this issue because on one hand I believe X, but on the other hand I also believe Y"). Finally, RAs could choose to indicate that no category fit the spirit of the response in question by selecting a fifth "other" category to be explored informally. Results of this coding are displayed in Table 2 below.

The most common strategy by far for those who preferred not to take sides was to simply say so without further justification (49.2%), followed by appealing to a general principle of public neutrality (23.1%), ignorance of relevant information (14.7%), or the merits of both sides (6.0%). Only 7.0% of responses were identified as fitting none of these categories. This coding suggests that when given the goal of maintaining trust, many people may view simply stating that they would rather not to get involved (vs. providing deeper justification) as an effective strategy.

Table 2
Participant-Generated Responses Opting Not to Take Sides, Categorized by Type of Justification Provided

Justification for staying out of it	Count	% of all "staying out of it" responses
None	147	49.2%
General principle	69	23.1%
Ignorance of issue	44	14.7%
Both sides have merit	18	6.0%
Other	21	7.0%

Importantly, this preference for short and simple expressions does not seem attributable to participants trying to speed through our study. Indeed, although participants were free to use curt expressions like "no comment" or "leave me alone," very few actually did, with over 80% of not-taking-sides responses being 10 words or longer. At the same time, not-taking-sides responses ($M = 17.6$) were 32% shorter on average compared to taking-sides responses ($M = 26.0$ words; $t(493) = 6.21$, $p < .001$), which likely reflects a more general intuition that politely staying out of it requires less explaining than sharing one's position. Finally, turning to the last question in our survey, when asked explicitly whether providing a deeper explanation for staying out of it would be helpful for building liking and trust, 74.7% indicated that a short answer ("just say I'd rather stay out of it") would prove more effective. That is, even when endorsing the longer expression took no additional time, participants still had the intuition that short and simple was the way to go.

Discussion

Our follow-up study tested people's intuitions about (not) taking sides from the perspective of the actor. The data suggest that when given the explicit goal of building trust and faced with the prospect of siding against their audience, people prefer not to take sides, suggesting that actors may miscalculate the costs of staying out of it. Moreover, participants who indicated a desire to stay out of it typically used short, straightforward expressions, often saying things like "I prefer to keep my political opinions to myself" or "You know, I just do not like to talk politics." In other words, participants put on the spot to take sides often generated just the sorts of not-taking-sides responses we have previously demonstrated backfire.

General Discussion

Taking the wrong side on a hot-button issue in public can have serious interpersonal consequences. Yet our results suggest that refusing to take sides carries its own risks. Across our experiments, we find that choosing not to take sides is often interpreted as strategically concealed opposition to the audience's prevailing position. As a result, staying out of it often seems less trustworthy than outright opposition; it makes one a less desirable cooperative partner; and it fails to win additional voting support or engender increased cooperation even among those it seems most likely to placate. Indeed, because staying out of it also entails failing to side

with either side, staying out of it can substantially undermine trust and support overall. Importantly, these effects persist across a host of paradigms, contexts, actors, issues, and ways to articulate one's reticence to take sides.

However, if opting not to take sides can be attributed to nonstrategic motives, these patterns of inference and distrust attenuate, suggesting that our results do not reflect distrust of moderate positions or nonresponsiveness per se, but concerns that staying above the fray may represent a deceptive attempt at impression management. Accordingly, we find that observers respond less negatively to attempts to stay out of it when they seem genuine, either because they are divorced from reputational incentives or accompanied by a private assurance that the speaker does not harbor partisan convictions one way or the other. These boundary conditions corroborate and clarify our account, and they also differentiate it from multiple potential alternatives.

For example, prior work on political apathy and moral indecision (Critcher et al., 2013; Zlatev, 2019) might predict distrust in some of our cases, but such accounts should, if anything, predict stronger distrust of actors who seem genuinely uninformed or torn between sides (as compared to those who say they prefer not to take sides but seem to harbor private convictions). This is the opposite of what we find in Experiments 5 and 6b: When observers can attribute an actor's decision not to take sides to ignorance or genuine long-term indecision, distrust is weakened. Relatedly, our results cannot be explained by people's general distrust of those who hold abhorrent views or who abet moral injustice (Baron & Ritov, 2004). Such accounts would predict that actively endorsing the "wrong" position (as in our opposition conditions) should seem worse than tacitly supporting them (as in our not-taking-sides conditions, where perceived opposition was present but weaker). Here, too, our studies find the opposite: Outright opposition was more trustworthy than taking no side at all.

One might reasonably wonder how our results relate to work on conversational norms (e.g., Grice, 1975). In our estimation, such work would not necessarily predict our effects, nor could it obviously explain them *ex post.* For example, although conversational norms prescribe that speakers should be responsive to direct questions where possible, they also prioritize politeness, dictating that information which might offend or cause conflict should often be kept private (Yoon et al., 2020). Thus, it is not clear that trying to stay out of it and seeming unresponsive should seem a worse violation than opposing one's audience outright. But even if nonresponsiveness were the greater sin, conversational norms might not obviously predict attenuations of our effect across cases which hold conversational behavior constant (e.g., Experiments 6a-b) or which entail no conversation at all (Experiments 4a-4b). An account based on negative responses to perceived impression management affords us sharper predictions and clearer explanations. Stepping back, while Gricean norms are typically thought to govern utilitarian exchanges of facts and information (Grice, 1975), discussions of political opinion may serve other social purposes (coalition building, impression formation, persuasion), which complicate the situation and merit more targeted study.

Here, we focused on how observers respond to actors who use simple and straightforward language to express their choice to "stay out of it," as compared to actors who oppose observers' moral beliefs outright. This approach is not without realism. As we show in our follow-up study, when facing audiences with whom they disagree, people spontaneously offer statements like "I'd rather not talk about my political views right now" and endorse them as effective for building trust specifically. In fact, when asked explicitly whether providing deeper justifications for choosing to stay out of it would help make a better impression, people did not seem to think that it would. Even public figures, who are presumably well-versed in dealing with media relations and public perception, sometimes employ similar tactics, and we replicate our results with two such naturalistic cases in Experiments 2a and 2b.

Of course, our studies do not investigate every manner in which the choice to stay out of it might be expressed or justified, and some conversational strategies for avoiding a strong position may prove more effective than others. Building on our initial exploration of justification tactics in Experiment 5, future work can broaden the range of conversational approaches tested as a means of identifying boundary strategies that might work better. At one end of the spectrum, it is not hard to imagine glibber responses like "no comment" faring even worse than those we tested, as these might seem not only deceptive and untrustworthy per our account, but also ruder and more abrupt. At the other, it seems possible that, especially for those whose beliefs really do fall somewhere in the middle, a more in-depth discussion of one's moderate or pragmatic preferences may help to soften the penalties associated with staying out of it. Indeed, emerging work suggests that effectively balancing the interests of multiple moral perspectives in pursuit of solutions that work for all sides can signal authentic moral character and garner respect (Puryear et al., 2022). And, more broadly, deeper and more connective conversations about personal experience can sometimes lesson interpersonal political hostility and bring people together (e.g., Kardas et al., 2021; Kubin et al., 2021). Still, while intriguing, such strategies are also more effortful, and they may not come readily to mind for those put on the spot to take sides and worried about saying the wrong thing. Moreover, the potential benefits of these strategies come with an ironic cost: If to avoid distrust one needs to carefully explain the nuanced beliefs behind one's choice not to take sides, there is an important sense in which one is not really free to stay out of it.

Zooming out, we hope future scholarship will continue to explore how observers respond to strategies for dealing with nuance and finding common ground in discussions of polarized issues. What sorts of inferences do people make about someone who plays devil's advocate, or who surfaces ideologically-inconvenient evidence in the name of impartial fact-finding, or who admits that their group's perspective may sometimes be biased? Unfortunately, we suspect that for touchy two-sided issues, such strategies, however well-intended, may also elicit skepticism and provoke distrust. In a polarized environment with clear social incentives, sending a credible signal that one harbors no strategic agenda may prove difficult.

Theoretical Implications and Future Directions

Our findings contribute to emerging research on the psychology of side-taking (DeScioli & Kurzban, 2013; Shaw et al., 2017) and specifically highlight the nuanced inferences people make about

those who try to avoid two-sided issues in the political sphere. Although interest in political conflict and affective polarization has exploded in recent years (Finkel et al., 2020; Van Prooijen & Krouwel, 2019; Westfall et al., 2015), little is known about people's judgments of those who try to cut a middle path by trying not to get involved. This gap is important given that discussions of contentious political issues ensnare friends, family members, coworkers, businesspeople, celebrities, and politicians, with serious interpersonal and societal consequences. We provide the first demonstration that observers make sophisticated belief attributions and character judgments from ostensibly signal-less choices not to take sides. As this was the first investigation into this issue, we chose to focus on expressions directed at relatively homogenous audiences (i.e., groups holding a common prevailing opinion, as is often the case in polarized environments). However, future research can explore how our effects play out with undecided or mixed audiences. Broadly, we suspect that observers may interpret staying out of it as strategic concealment in such cases, too, although they may struggle to pinpoint exactly which beliefs and opinions actors are trying to conceal.

Furthermore, this research advances a practical understanding of how actors deal with difficult questions in public contexts and how strategies for doing so are interpreted by observers. Previous work has explored a slew of evasive rhetorical tactics which allow actors to respond to direct questions without offering any substantive answers (Bitterly & Schweitzer, 2020; Rogers et al., 2017; Rogers & Norton, 2011). Although we agree that strategies like dodging and paltering are prevalent and fascinating, attempts to earnestly avoid take sides have received scant attention. Our central result—that such strategies provoke stronger distrust than outright opposition—may seem surprising since disagreement over divisive issues is known to provoke anger, prejudice, and even violence (Skitka, 2010), and maintaining impartiality is often a virtue (Shaw et al., 2019), particularly for those in positions leadership (Everett et al., 2018). Yet, at least in the contexts we examined, penalties associated with staying out of it were often steeper.

Nevertheless, our follow-up study suggests that when faced with the prospect of fragmenting support by taking a controversial position, staying out of it seems intuitively attractive. Why might this be so? Perhaps in context, actors underweight the indirect cost of seeming deceptive against the salient risk of directly angering observers, or perhaps they fail to realize they might inadvertently portray themselves as evasive or inauthentic at all. Another possibility is that actors might choose to stay out of it because they assume that doing so will garner less attention (e.g., social gossip, news coverage) than taking a strong stand one way or another. Yet another possible explanation is that staying out of it provides some cover for actors to change their position if public opinion later shifts (i.e., without seeming hypocritical; Effron et al., 2018). Building on the results from our follow-up study, future authors can investigate further the intuitive pull of staying out of it by asking why actors choose to stay out of it and whether any specific expectations about doing so are warranted. Our data thus far seem to align with the "impression mismanagement" thesis (Steinmetz et al., 2017), that people sometimes adopt self-presentation strategies which actually portray them more negatively.

We certainly do not mean to imply that not taking a side is always the wrong choice. Penalties for standing on the wrong side of a contentious political issue in public are well-documented and sometimes severe. In our studies, opting to stay out of it softened perceived opposition relative to outright disagreement and, in some cases, this alone may be worth the costs of seeming less trustworthy. More generally, there remain open questions about other dimensions of evaluation which may prove relevant. How might staying out of it impact perceptions of confidence, moral conviction, or fitness for specific roles? More data here will offer a broader lens on the costs and benefits of (not) taking sides.

A further question concerns whether certain leadership roles in society are protected from the patterns of inference and judgment documented here. For example, judges, high-level bureaucrats, and even journalists are sometimes expected to maintain political impartiality in fulfilling their professional responsibilities. We suspect that for these sorts of actors, choosing not to take sides may not harm trust to the same extent because observers may attribute it to role-specific norms of conduct rather than to impression management motives. Roles that explicitly require neutrality may thus represent an interesting potential boundary to our effects, and they may serve an important social function—allowing actors to occupy middle-ground positions for the purposes of impartial information gathering and arbitrating disagreement.

For actors who do anticipate and fear the repercussions of conspicuously staying out of it, the backfire effects we document here seem to incentivize taking sides. In certain cases, this dynamic may compel actors to endorse positions on contentious issues they know little about or to express convictions they actually lack. Without clear and convincing communication about nuanced or moderate positions, the psychology we document may lead observers to sort those who try to avoid conflict according to a "with me or against me" mindset, leaving little room to surface nuance or deescalate conflict. Importantly, such effects may not be limited to the domain of political discourse: Staying out of it may provoke distrust and dislike across myriad issues in family feuds, workplace disputes, disagreements on social media, negotiations, or even legal proceedings—anywhere reputational incentives are on the line and avoiding the conversation might appear strategic. Future researchers should look to generalize our effects to other domains of public disagreement and to incorporate the social costs of staying out of it into explanatory theories of polarization and intergroup conflict.

Conclusion

Open-minded discussion of complex and consequential issues is a hallmark of a well-functioning society. But a healthy public debate also affords its participants the freedom to avoid taking sides. Our results suggest that, in practice, trying to exercise that freedom can backfire. We present evidence that choosing to stay out of it is often imbued by observers with predictable patterns of social meaning and that it can provoke skepticism and distrust as a result. These findings advance our understanding of social incentives for side-taking in moral conflict, and we hope they open avenues for mitigating ideological disagreement more broadly.

Context of the Research

Both authors are interested in the tactics people use to construct and protect their reputations; and in how others interpret and

evaluate the use of such tactics in return. In some sense, we are studying the 'epistemic toolkit' observers use to suss out deceptive impression management strategies and the judgments they levy against those who deploy them. This psychology, it turns out, serves as an important input to moral judgment across contexts from charity and philanthropy (Silver et al., 2021), to intellectual property (Silver & Shaw, 2018), to resource allocation (Shaw et al., 2018). Here, we extend this work into the domain of political discourse, demonstrating that people penalize those who try to protect their reputation by avoiding contentious topics like immigration, abortion, or gun control—sometimes even more so than those who oppose them on such issues outright.

References

Alford, J. R., Hatemi, P. K., Hibbing, J. R., Martin, N. G., & Eaves, L. J., (2011). The politics of mate choice. *The Journal of Politics*, *73*(2), 362–379. https://doi.org/10.1017/S0022381611000016

Axelrod, R. (1980). Effective choice in the prisoner's dilemma. *Journal of Conflict Resolution*, *24*(1), 3–25. https://doi.org/10.1177/002200278002400101

Baron, J., & Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes*, *94*(2), 74–85. https://doi.org/10.1016/j.obhdp.2004.03.003

Baum, S. M., & Critcher, C. R. (2020). The costs of not disclosing. *Current Opinion in Psychology*, *31*, 72-75.

Berman, J. Z., Levine, E. E., Barasch, A., & Small, D. A. (2015). The braggart's dilemma: On the social rewards and penalties of advertising prosocial behavior. *Journal of Marketing Research*, *52*(1), 90–104. https://doi.org/10.1509/jmr.14.0002

Bitterly, T. B., & Schweitzer, M. E. (2020). The economic and interpersonal consequences of deflecting direct questions. *Journal of Personality and Social Psychology*, *118*(5), 945–990. https://doi.org/10.1037/pspi0000200

Bonica, A., Rosenthal, H., Blackwood, K., & Rothman, D. J., (2020). Ideological sorting of physicians in both geography and the workplace. *Journal of Health Politics, Policy and Law*, *45*(6), 1023–1057. https://doi.org/10.1215/03616878-8641555

Brandt, M. J., Reyna, C., Chambers, J. R., Crawford, J. T., & Wetherell, G. (2014). The ideological-conflict hypothesis: Intolerance among both liberals and conservatives. *Current Directions in Psychological Science*, *23*(1), 27–34. https://doi.org/10.1177/0963721413510932

Byrne, D. (1971). *The attraction paradigm*. Academic Press.

Chambers, J. R., & Melnyk, D. (2006). Why do I hate thee? Conflict misperceptions and intergroup mistrust. *Personality and Social Psychology Bulletin*, *32*(10), 1295–1311. https://doi.org/10.1177/0146167206289979

Chambers, J. R., Schlenker, B. R., & Collisson, B., (2013). Ideology and prejudice: The role of value conflicts. *Psychological Science*, *24*(2), 140–149. https://doi.org/10.1177/0956797612447820

Chen, M. K., & Rohla, R. (2018). The effect of partisanship and political advertising on close family ties. *Science*, *360*(6392), 1020–1024. https://doi.org/10.1126/science.aaq1433

Critcher, C. R., Inbar, Y., & Pizarro, D. A. (2013). How quick decisions illuminate moral character. *Social Psychological and Personality Science*, *4*(3), 308–315. https://doi.org/10.1177/1948550612457688

Crockett, M. J. (2017). Moral outrage in the digital age. *Nature Human Behaviour*, *1*(11), 769–771. https://doi.org/10.1038/s41562-017-0213-3

DeScioli, P., & Kurzban, R. (2013). A solution to the mysteries of morality. *Psychological Bulletin*, *139*(2), 477–496. https://doi.org/10.1037/a0029065

Ditto, P. H., Liu, B. S., Clark, C. J., Wojcik, S. P., Chen, E. E., Grady, R. H., Celniker, J. B., & Zinger, J. F. (2019). At least bias is bipartisan: A meta-analytic comparison of partisan bias in liberals and conservatives. *Perspectives on Psychological Science*, *14*(2), 273–291. https://doi.org/10.1177/1745691617746796

Effron, D. A., O'Connor, K., Leroy, H., & Lucas, B. J. (2018). From inconsistency to hypocrisy: When does "saying one thing but doing another" invite condemnation? *Research in Organizational Behavior*, *38*, 61–75. https://doi.org/10.1016/j.riob.2018.10.003

Everett, J. A. C., Faber, N. S., Savulescu, J., & Crockett, M. J. (2018). The costs of being consequentialist: Social inference from instrumental harm and impartial beneficence. *Journal of Experimental Social Psychology*, *79*, 200–216. https://doi.org/10.1016/j.jesp.2018.07.004

Finkel, E. J., Bail, C. A., Cikara, M., Ditto, P. H., Iyengar, S., Klar, S., Mason, L., McGrath, M. C., Nyhan, B., Rand, D. G., Skitka, L. J., Tucker, J. A., Van Bavel, J. J., Wang, C. S., & Druckman, J. N. (2020). Political sectarianism in America. *Science*, *370*(6516), 533–536. https://doi.org/10.1126/science.abe1715

Frimer, J. A., Skitka, L. J., & Motyl, M. (2017). Liberals and conservatives are similarly motivated to avoid exposure to one another's opinions. *Journal of Experimental Social Psychology*, *72*, 1–12. https://doi.org/10.1016/j.jesp.2017.04.003

Gandz, J., & Murray, V. V. (1980). The experience of workplace politics. *Academy of Management Journal*, *23*, 237–251.

Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. L. Morgan (Eds.), *Syntax and semantics, Vol. 3, speech acts* (pp. 41–58). Academic Press.

John, L. K., Barasz, K., & Norton, M. I. (2016). Hiding personal information reveals the worst. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(4), 954–959. https://doi.org/10.1073/pnas.1516868113

Jones, E. E., & Pittman, T. S. (1982). Toward a general theory of strategic self-presentation. *Psychological Perspectives on the Self*, *1*, 231–262.

Jordan, J. J., Sommers, R., Bloom, P., & Rand, D. G. (2017). Why do we hate hypocrites? Evidence for a theory of false signaling. *Psychological Science*, *28*(3), 356–368. https://doi.org/10.1177/0956797616685771

Jost, J. T. (2017). The marketplace of ideology: "Elective affinities" in political psychology and their implications for consumer behavior. *Journal of Consumer Psychology*, *27*(4), 502–520. https://doi.org/10.1016/j.jcps.2017.07.003

Kardas, M., Kumar, A., & Epley, N. (2021). Overly shallow?: Miscalibrated expectations create a barrier to deeper conversation. *Journal of Personality and Social Psychology*. Advance Online Publication, https://doi.org/10.1037/pspa0000281

Kubin, E., Puryear, C., Schein, C., & Gray, K. (2021). Personal experiences bridge moral and political divides better than facts. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(6), Article e2008389118. https://doi.org/10.1073/pnas.2008389118

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

Marks, J., Copland, E., Loh, E., Sunstein, C. R., & Sharot, T. (2019). Epistemic spillovers: Learning others' political views reduces the ability to assess and use their expertise in nonpolitical domains. *Cognition*, *188*, 74–84. https://doi.org/10.1016/j.cognition.2018.10.003

Parker, M. T., & Isbell, L. M. (2010). How I vote depends on how I feel: The differential impact of anger and fear on political information processing. *Psychological Science*, *21*(4), 548–550. https://doi.org/10.1177/0956797610364006

Pew Research Center. (2019). *The public's level of comfort talking politics and Trump*. Public highly critical of state of political discourse in the U.S. https://www.pewresearch.org/politics/2019/06/19/the-publics-level-of-comfort-talking-politics-and-trump/

Puryear, C., & Gray, K. (2022). *Using "balanced pragmatism" in political discussions increases cross-partisan respect*. Pre-print.

Rogers, T., & Norton, M. I. (2011). The artful dodger: Answering the wrong question the right way. *Journal of Experimental Psychology: Applied*, *17*(2), 139–147. https://doi.org/10.1037/a0023439

Rogers, T., Zeckhauser, R., Gino, F., Norton, M. I., & Schweitzer, M. E. (2017). Artful paltering: The risks and rewards of using truthful statements to mislead others. *Journal of Personality and Social Psychology*, *112*(3), 456–473. https://doi.org/10.1037/pspi0000081

Rogowski, J. C., & Sutherland, J. L. (2016). How ideology fuels affective polarization. *Political Behavior*, *38*(2), 485–508. https://doi.org/10.1007/s11109-015-9323-7

Schlenker, B. R. (1980). *Impression management*. Brooks/Cole Publishing Company.

Shaw, A., Barakzai, A., & Keysar, B. (2019). When and why people evaluate negative reciprocity as more fair than positive reciprocity. *Cognitive Science*, *43*(8), Article e12773. https://doi.org/10.1111/cogs.12773

Shaw, A., Choshen-Hillel, S., & Caruso, E. M. (2018). Being biased against friends to appear unbiased. *Journal of Experimental Social Psychology*, *78*, 104–115.

Shaw, A., DeScioli, P., Barakzai, A., & Kurzban, R. (2017). Whoever is not with me is against me: The costs of neutrality among friends. *Journal of Experimental Social Psychology*, *71*, 96–104. https://doi.org/10.1016/j.jesp.2017.03.002

Silver, I., & Shaw, A. (2018). No harm, still foul: Concerns about reputation drive dislike of harmless plagiarizers. *Cognitive Science*, *42*(Suppl 1), 213–240. https://doi.org/10.1111/cogs.12500

Silver, I., Newman, G., & Small, D. A. (2021). Inauthenticity aversion: Moral reactance toward tainted actors, actions, and objects. *Counselling Psychology Review*, *4*, 70–82. https://doi.org/10.1002/arcp.1064

Skitka, L. J. (2010). The psychology of moral conviction. *Social and Personality Psychology Compass*, *4*(4), 267–281. https://doi.org/10.1111/j.1751-9004.2010.00254.x

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language*, *25*(4), 359–393. https://doi.org/10.1111/j.1468-0017.2010.01394.x

Steinmetz, J., Sezer, O., & Sedikides, C. (2017). Impression mismanagement: People as inept self-presenters. *Social and Personality Psychology Compass*, *11*(6), Article e12321. https://doi.org/10.1111/spc3.12321

Toner, K., Leary, M. R., Asher, M. W., & Jongman-Sereno, K. P. (2013). Feeling superior is a bipartisan issue: Extremity (not direction) of political views predicts perceived belief superiority. *Psychological Science*, *24*(12), 2454–2462. https://doi.org/10.1177/0956797613494848

van Prooijen, J. W., & Krouwel, A. P. (2019). Psychological features of extreme political ideologies. *Current Directions in Psychological Science*, *28*(2), 159–163. https://doi.org/10.1177/0963721418817755

Westfall, J., Van Boven, L., Chambers, J. R., & Judd, C. M. (2015). Perceiving political polarization in the United States: Party identity strength and attitude extremity exacerbate the perceived partisan divide. *Perspectives on Psychological Science*, *10*(2), 145–158. https://doi.org/10.1177/1745691615569849

Wosinska, W., Dabul, A. J., Whetstone-Dion, R., & Cialdini, R. B. (1996). Self-presentational responses to success in the organization: The costs and benefits of modesty. *Basic and Applied Social Psychology*, *18*, 229–242. https://doi.org/10.1207/s15324834basp1802_8

Yoon, E. J., Tessler, M. H., Goodman, N. D., & Frank, M. C. (2020). Polite speech emerges from competing social goals. *Open Mind: Discoveries in Cognitive Science*, *4*, 71–87. https://doi.org/10.1162/opmi_a_00035

Zlatev, J. J. (2019). I may not agree with you, but I trust you: Caring about social issues signals integrity. *Psychological Science*, *30*(6), 880–892. https://doi.org/10.1177/0956797619837948

Received December 6, 2019
Revision received January 14, 2022
Accepted January 27, 2022 ■