Journal of Experimental Social Psychology 98 (2022) 104231

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

The partisan trade-off bias: When political polarization meets policy trade-offs☆

Daniela Goya-Tocchetto [a,*], Aaron C. Kay [a], Heidi Vuletich [b], Andrew Vonasch [c], Keith Payne [d]

[a] Department of Management & Organizations, The Fuqua School of Business, Duke University, United States
[b] Department of Psychology, Indiana University, United States
[c] Department of Psychology, University of Canterbury, New Zealand
[d] Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, United States

ARTICLE INFO

Editor: Jarret Crawford

Keywords:
Trade-offs' reasoning
Political polarization
Side-effect effect
Knobe effect
Partisan bias

ABSTRACT

Liberals and conservatives currently struggle to reach political agreement on policy proposals. While political polarization is closely associated with this phenomenon, the precise psychological mechanisms via which polarization works to affect political compromise remain to be fully explored. Across five studies ($N = 1236$; 2126 total individual observations), we uncover one such mechanism by exploring a novel and robust bias that emerges at the crossroads of policy trade-offs and partisanship. We call it the *Partisan Trade-off Bias*. When interpreting policy trade-offs, both Democrats and Republicans view the unintended but unavoidable side effects of policies proposed by contrapartisans as wanted and intended. Yet they do not attribute intentionality to the very same types of side effects of policies proposed by copartisans. We provide evidence for this bias across four types of policy trade-offs, including taxes, environmental regulation, gun control, and voting rights. Importantly, we show that the partisan trade-off bias is a unique contributor to decreased willingness to accept policy deals from contrapartisans, thus reducing the chances of reaching political agreement. Our studies suggest that the partisan trade-off bias is a product of the lack of trust in contrapartisans. In an experimental study, we manipulate trust and decrease the magnitude of this bias, showing evidence for our proposed mechanism and revealing a potential intervention to foster political compromise.

Political progress is rife with debates over policy trade-offs in which people are asked to accept undesired side effects in order to bring about desired primary effects. While policies with only desired effects (i.e., "win-win" policies) are implemented with little controversy, those with only undesired effects (i.e., "lose-lose" policies) are obviously avoided. Hence it is only the trade-offs that become matters of discussion at all. Yet trade-offs create natural obstacles to human reasoning (Tetlock, 2000). In the policy domain, reasoning through trade-offs can be particularly difficult because we cannot see the true desires and intentions of policy makers, we can only infer them, and these inferences are heavily reliant on how much we trust them. In a political environment in which negative partisanship and distrust are at historically high levels (Pew Research Center, 2016, 2019), is political polarization leading people to doubt the intentionality behind contrapartisan policy trade-offs in a way that they do not doubt the intentionality behind identical copartisan policies?

Consider the following example of a liberal policy trade-off. In order to protect the environment for future generations, policy makers support a plan to increase environmental regulations claiming it would protect the environment, but knowing it also has the short-term potential to cut jobs in the fossil-fuels industry. The primary goal of this policy is clear, namely, to protect the environment. Nonetheless, this policy comes with an undesirable short-term side effect which the policy makers were themselves aware of but did not wish to bring about. A charitable interpretation of the proponents' motives and desires would conclude that the intention or purpose of the policy was not to potentially cut jobs in the fossil fuel industry. That is, if they could somehow cut fossil fuels without threatening jobs in the short-term, they surely would. However, we predict that while Democrats will evaluate the potential decrease in the number of available jobs as *unintentional*, Republicans will perceive this negative effect as an *intentional* and wanted outcome of the liberal proposal. What is more, Republicans are expected to downplay the

☆ This paper has been recommended for acceptance by Dr Jarret Crawford.

* Corresponding author at: Duke University, 100 Fuqua Drive, Durham, NC 27708, United States.
E-mail address: dg217@duke.edu (D. Goya-Tocchetto).

https://doi.org/10.1016/j.jesp.2021.104231
Received 29 March 2021; Received in revised form 5 September 2021; Accepted 24 September 2021
Available online 14 October 2021

sincerity behind the primary effect of this policy. These partisan perceptions of intended and unintended effects may have important consequences for policymaking and compromise. Why support a policy to protect the environment, for example, if it is seen as just a ploy to harm big businesses and workers? We predict that this partisan bias will be symmetric, such that Democrats will also make biased attributions of intentionality for Republican policy trade-offs.

Across five studies ($N = 1236$; 2126 total individual observations including nested data) we provide evidence for this phenomenon, showing that negative partisanship biases attributions of intentionality in the evaluation of policy trade-offs. People malign the motives of contrapartisans in a way that they do not malign the motives of copartisans, even in contexts where the policy trade-off is identical. Instead of charitably acknowledging that members of the opposite political party neither want nor intend to bring about the negative side effects that might accompany their primary aims, people instead impugn their motives and assume the worst about their intentions. We call this the "Partisan Trade-off Bias."

This new bias is particularly worrisome insofar as people can be unwilling to negotiate if they believe that policy makers are being deceitful about their true purpose. We know that political polarization leads to political sectarianism. As a result, people hold contrapartisans in contempt solely on the basis of their political identity. This behavior has been hypothesized to undermine the development of creative cross-party solutions and mutually beneficial compromises (Finkel, Bail, Cikara, Ditto, Iyengar, Klar, Mason, McGrath, Nyhan, Rand, Skitka, Tucker, Van Bavel, Wang, & Druckman, 2020). While generalized contempt and opposition to policies proposed by contrapartisans have been associated with the issue of cross-party solutions, the precise psychological mechanisms via which polarization works to affect the possibility of political compromise remain to be fully explored. Our studies uncover one such mechanism, showing that the partisan trade-off bias leads to decreased willingness to accept policy deals from contrapartisans, thus reducing the chances of reaching an agreement. It is possible that people have become so distrustful of contrapartisan politicians that they are questioning the primary goals of policy trade-offs, perhaps believing that they are being used to veil the true purposes of contrapartisan policies—which would be the apparent "side effects." The partisan trade-off bias may thus pose a threat to political compromise.

Fortunately, understanding the partisan trade-off bias also reveals a potential solution to this problem. In our studies, we examine the causal mechanism driving this bias and design an intervention capable of decreasing it for members of both political parties. Below, we review the relevant literature, elaborate on why this type of bias might exist, and explain how we test for this important but unstudied potential effect of political polarization.

1.1. The psychology of trade-offs

Reasoning through trade-offs is inherently hard, as trade-offs involve comparisons between different dimensions that people may not be cognitively equipped to properly handle (Tetlock, 2000). For example, consider a trade-off that is at the core of several public policies: "what loss of liberty would I accept to achieve an increment in public safety?" (Tetlock, 2000, p. 242). Deriving an exact answer to this type of comparison is not something people can easily or naturally do. Trading-off values also leads to emotional costs associated with the dissonance of acknowledging that an important value had to be sacrificed. What is more, a person's commitment to other people and to cultural values may prevent them from making certain trade-offs in the first place (e.g., taboo trade-offs).

While economists in the utilitarian tradition have full confidence in people's ability to properly understand and reason their way through all kinds of trade-offs (Becker, 1981), many psychologists uphold a more pessimistic perspective on human cognitive abilities regarding trade-off reasoning (Kahneman, Slovic, & Tversky, 1982). Yet people are not at a complete loss when it comes to reasoning through trade-offs and they are more likely to make competent decisions involving trade-offs in domains that are of great importance for them (Tetlock, 1986; Tetlock, 2000; Tetlock, Hannum, & Micheletti, 1984).

A feature of trade-offs that makes reasoning through them particularly challenging is their intentionality structure. Using policy trade-offs as our paradigmatic case, we can see that these policies have a very specific structure: A given policy is adopted in order to bring about a *desired and intended primary effect 'x'* even though doing so will bring about a *foreseen but unintended side effect 'y'*. In this kind of cost-benefit analysis, *'x'* is *traded off* against *'y'* because the benefits of the former cannot be achieved without paying the costs of the latter. What makes trade-off reasoning challenging when it comes to intentionality is that both the primary and the side effects are foreseen, but only the former is desired. The latter is brought about solely because it is necessary to obtain the primary effect.

1.1.1. Trade-offs, side-effects, and perceived intent

The particular structure of a trade-off makes it possible for people to judge that an outcome was intentionally brought about by an agent even if the agent did not specifically desire, try, or even intend to bring about that outcome (Knobe, 2003; Machery, 2008; Vonasch & Baumeister, 2017). In the context of actions that have side effects, this phenomenon has been labeled the *Knobe Effect* or the *Side-Effect Effect* (hereafter, SE Effect). More specifically, the SE Effect describes an asymmetry in judgments about the intentionality of good versus bad side effects that are unintended but foreseen. In Knobe's original vignettes, a chairman of the board was told that adopting a program would increase profits but also harm/help the environment. He stated that he didn't care at all about the environment, he just wanted to increase profits. So, he adopted the program, thereby increasing profits and harming/helping the environment. In the harm condition, 82% of the participants judged that the chairman harmed the environment intentionally. In the help condition—where everything was identical except for the valence of the side effect—roughly 23% of the participants judged that the chairman helped the environment intentionally. This asymmetry poses problems for traditional models of the folk concept of intentional action whereby undesired and unintended, yet foreseen side effects are *unintentional* (Knobe, 2006).[1]

Our work extends the understanding and applications of the SE Effect. In the case of the SE effect, the moral wrongness of the outcome leads people to infer that the action was intentional (Guglielmo & Malle, 2010; Sripada, 2010, 2012). We predict that the partisan trade-off bias depends upon *who* advocates for a policy (rather than the morality of the policy outcome) and is driven primarily by a particular psychological consequence of political polarization (i.e., distrust). In the next section, we discuss how the current context of hyper political polarization lends itself to the emergence of the partisan trade-off bias and describe the hypotheses that are tested in the studies.

1.2. Political polarization and negative partisanship

Partisan animosity has reached historically high levels in the United States (Carlin & Love, 2018; Iyengar & Westwood, 2015). Hostility along partisan lines is running so deep among Americans that people not only dislike contrapartisans, but more than ever distrust them and believe that they are "more immoral" when compared with other Americans (Pew Research Center, 2019). One of the main consequences of this negative partisanship is an increase in partisan bias, i.e., the tendency to think or act in ways that favors one's own political group or denigrates one's opposite political group. To this point, much research shows that support

[1] For a review of the earlier work on the SE Effect, see Nadelhoffer (2010).

for policies is influenced by political partisanship (Bolsen, Druckman, & Cook, 2014; Bullock, 2011; Crawford, Kay, & Duke, 2015; Druckman, Peterson, & Slothuus, 2013; Kahan, Jenkins-Smith, & Braman, 2011; Lord, Ross, & Lepper, 1979; Taber & Lodge, 2006). Moreover, voters make use of partisan cues when they evaluate different policy proposals. For example, categorizing policies as either Democratic or Republican influences the judgments of self-identified independent voters based on their level of implicit partisan bias (Hawkins & Nosek, 2012) and support for policy "nudges" is heavily influenced by whether they are framed as supporting the goals of the Democratic or Republican party (Tannenbaum, Fox, & Rogers, 2017). In a similar vein, Republicans are less supportive of policy solutions to climate change when they are presented by Democrats (Van Boven, Ehret, & Sherman, 2018).

1.2.1. Contrapartisan distrust and attributions of intentionality

Another major consequence of heightened negative partisanship is the increasing polarization of political trust along partisan lines. Republicans and Democrats are significantly less likely to trust the government when it is controlled by contrapartisans (Bartels, 2002; Hetherington & Rudolph, 2014; Hetherington & Rudolph, 2018). This decreased willingness to trust the government is worrisome, as political trust is a crucial feature of functioning democratic societies. While political trust helps to reduce the ideological gap between policies proposed by the ruling party and those proposed by the opposite party, "polarized trust inhibits the formation of public consensus on public policy because it reduces the willingness of citizens to sacrifice their ideological proclivities for the common good" (Hetherington & Rudolph, p. 580, 2015).

The Affordable Care Act provides an example of the consequential effects of polarized trust on political compromise and policy support. Before President Obama signed the Affordable Care Act ("Obamacare") into law in 2010, then Governor Mitt Romney signed the Massachusetts Health Care Reform bill ("Romneycare") into law in 2006. Both plans were based on a conservative proposal by the Heritage Foundation, the goal of which was to increase how many citizens are covered by private insurance companies. A key component of both plans is the individual mandate whereby all individuals are required to purchase private insurance (which is the only way of incentivizing insurance companies both to cover individuals with preexisting conditions and to offer a wider variety of more affordable plans). Before President Obama implemented the individual mandate as part of the Affordable Care Act, conservatives were broadly supportive of the approach originally adopted by Romney—with prominent Republicans such as Rick Santorum, President G.H. Bush, Newt Gingrich, Bob Dole, Chuck Grassley, and others voicing their approval. Yet, once the individual mandate became part of the Affordable Care Act, conservatives did an about-face and charged that it was an unconstitutional assault on American freedom. Popular opinion about the individual mandate is unsurprisingly split along partisan lines—with Democrats largely in favor and Republicans largely opposed. How people view the intentions, goals, and side effects of the Affordable Care Act appears to be often colored by a partisan lens.

2. Overview of hypotheses

The psychology of trade-offs reasoning in this context of heightened negative partisanship leads us to test four hypotheses. The first and second hypotheses specify what we have coined the partisan trade-off bias. We call them *The Partisan Side Effect Bias* and *The Partisan Primary Effect Bias* hypotheses, respectively. The third hypothesis specifies a significant negative consequence of this bias—we call this *The Policy Deals* hypothesis.

Hypothesis 1. (*The Partisan Side Effect Bias*): People will evaluate the side effects of policy trade-offs proposed by contrapartisans as more intentionally brought about than the side effects of policies proposed by copartisans. Importantly, this bias will be symmetric for Democrats and Republicans.

Hypothesis 2. (*The Partisan Primary Effect Bias*): People will evaluate the primary effects of policy trade-offs proposed by contrapartisans as less intentionally brought about than the primary effects of policies proposed by copartisans. Importantly, this bias will be symmetric for Democrats and Republicans.

Hypothesis 3. (*The Policy Deals Hypothesis*): The Partisan Trade-off Bias will uniquely contribute to decreased willingness to accept policy deals from contrapartisans.

Taken together, *The Partisan Side Effect Bias* and *The Partisan Primary Effect Bias* hypotheses constitute what we have termed *The Partisan Trade-Off Bias*. A crucial aspect of these hypotheses is that we expect this bias to occur across a variety of policies, such that partisans will display the partisan trade-off bias even when evaluating content-free policies that are identical across the political spectrum. Another important feature is that we predict that the partisan trade-off bias will be symmetrical across the political spectrum, such that both Democrats and Republicans will be susceptible to the biased evaluation of policy trade-offs.

We expect the partisan trade-off bias to be a result of negative partisanship. Hence in the first two studies we measured negative attitudes towards contrapartisans using questions from the Pew Research Center polls on partisanship. These questions allowed us to draw connections between the present research and widely discussed trends towards affective polarization. However, these questions are not precise about psychological constructs. They include measures of emotions, general evaluations, traits, perceived morality, and other attributes of each party. To more precisely identify the psychological processes mediating the partisan trade-off bias, in Study 4 we created items to measure two specific constructs: liking and trust. We predict that lack of trust in members of the opposing party is the main mechanism driving this bias.

There is extensive evidence showing that people tend to distrust members of the opposing party (Iyengar & Krupenkin, 2018; Iyengar, Sood, & Lelkes, 2012; Iyengar & Westwood, 2015; Pew Research Center, 2019). And much research shows that trust is one of the main predictors of perceptions of others' motives and intentions. If we have reason to distrust a group, we have reason to doubt their intentions. Furthermore, the more we distrust an outgroup relative to our ingroup, the more we will doubt their intentions relative to the intentions of our own group. We predict that this gap in trust in contrapartisans relative to trust in copartisans is the main driver of the partisan trade-off bias, leading people to perceive the side effects of contrapartisan policies as more intentional and their primary effects as less intentional—relative to copartisan policies. Hence, we reasoned that the effect of political partisanship on intentionality attribution to the primary and side effects of policy trade-offs would be mediated, specifically, by how much people distrust contrapartisans relative to copartisans.

If a lack of trust drives the partisan trade-off bias, can we increase trust in contrapartisans and reduce this bias? We draw inspiration from work done in the field of organizational behavior to come up with an intervention that is capable of increasing trust in contrapartisan policy makers and reducing the partisan trade-off bias. The work on trust in leaders recognizes two important aspects in building trust: the presence of shared goals between leaders and different stakeholders, and the ability of stakeholders to exercise voice (Lyu & Ferrin, 2018; Cooper, Dirks, Ferrin, & Kim, P., 2006; Dirks & Ferrin, 2002; Holland, Cooper, & Sheehan, 2017). The former refers to the presence of shared values such as fostering justice and improving human well-being, while the latter refers to people's ability to express their views and influence decision making processes that impact lives. Recent research has explored the effects of listening on building trust and decreasing defensive processing of information in situations of conflict (Itzchakov & Kluger, 2017; Stine,

Table 1
Policies presented to participants by condition.

	Party Proposing Each Type of Policy Trade-off			
Conditions	Taxes	Gun Control	Environment	Voting
1	Republican	Democratic	Republican	Democratic
2	Democratic	Republican	Democratic	Republican
3	Republican	Republican	Democratic	Democratic
4	Democratic	Democratic	Republican	Republican
5	Republican	Democratic	Democratic	Republican
6	Democratic	Republican	Republican	Democratic

Thompson, & Cusella, 1995). For instance, experiencing high quality listening has been associated with benefits such as increased openness to change and decreased prejudice in people's attitudes (Itzchakov, Weinstein, Legate, & Amar, 2020). Our final hypothesis builds on this literature on the role of trust and listening on judgments of intentionality. We call it *The Trust Hypothesis*:

Hypothesis 4. (*The Trust Hypothesis*): The relative distrust in contrapartisans compared with copartisans will mediate the effect of political partisanship on attribution of intentionality; hence, increasing trust in contrapartisan policy makers should attenuate the partisan trade-off bias.

We expected that both dislike and distrust may contribute to the effect of political partisanship on attributions of intentionality. Yet our primary hypothesis was that distrust in contrapartisans would be the stronger mediator of the partisan trade-off bias.

3. Overview of studies

We test our four hypotheses across five studies ($N = 1236$; 2126 total individual observations). In the first study, we document the partisan trade-off bias across four types of policies: voting rights, environmental protection, taxation, and gun control. In the second study, we provide a stronger test of the partisan trade-off bias by replicating it with content-free policy proposals that maintain the structure of a trade-off but lack any type of political content. Taken together, the first two studies provide evidence for our first two hypotheses: *The Partisan Side Effect Bias* (H1) and *The Partisan Primary Effect Bias* (H2). In the third study, we show that the partisan trade-off bias leads to decreased willingness to accept policy deals from contrapartisans—providing evidence for *The Policy Deals* hypothesis (H3). In the fourth study, we provide support for the specific proposed mechanism underlying the partisan trade-off bias: lack of trust in contrapartisans. In the fifth and final study, we manipulate trust and, as a result, significantly decrease the magnitude of the bias. Taken together, the final two studies provide evidence for *The Trust* hypothesis (H4), in addition to a replication of the partisan trade-of bias across both studies.

We report all measures and manipulations in our studies and all materials, data, and code for all studies are made available in the project's OSF page.[2] No data was collected after reaching our target samples and sample sizes were all determined in advance of data collection. All data exclusions (if any) are reported for all studies. For each study, we report a sensitivity analysis of the minimum effect size each of our samples had power to detect.

4. Study 1

In the first study we investigated the perceived intentionality of both the primary effects and the side effects of Democratic and Republican policies. We tested the first two hypotheses (*The Partisan Side Effect Bias* and *The Partisan Primary Effect Bias*), predicting that people would evaluate the side effects of policy trade-offs proposed by contrapartisans as more intentional, and the primary effects as less intentional, than equivalent side and primary effects of the policies proposed by copartisans. Additionally, we measured general attitudes towards each party given that we expected the partisan trade-off bias to be associated with negative partisanship.[3]

4.1. Participants

Participants were 332 Mechanical Turk (mTurk) workers located in the United States (50% male; $M_{age} = 37.3$, $SD_{age} = 12.3$). Given that we used repeated measures, the long format data contained an $N = 664$. Forty-five percent of participants reported being Democratic, 24% labeled themselves as neither Democratic nor Republican, and 31% reported being Republican. Participants' reports of their political ideology matched their political partisanship closely: 46% of participants reported being liberal or slightly liberal, 21% reported being neither liberal nor conservative, and 32% rated themselves conservative or slightly conservative. Sensitivity power analyses with linear multiple regression using three predictors and standard criteria $alpha = 0.05$ and $beta = 0.80$ revealed a minimum effect size that could be detected by our sample equal to $f^2 = 0.01$ (Critical $F(3, 660) = 2.61$).

4.2. Procedures

After providing informed consent, participants reported their political partisanship and political orientation. Then, they read descriptions of two Republican policy proposals and two Democratic proposals, all involving trade-offs: each policy included an expected primary outcome and a side effect. For example, one of the Republican policy scenarios reads:

> *"Republicans say they want to reform the tax code. They support a tax break claiming it will help the rich have more money to create jobs, but knowing it also has the potential of cutting benefits for the poor. People have different views about the intentions of the law. What is your opinion?"*

Participants rated the extent to which they believed Democrats or Republicans intended the primary outcome and the side effect of each of their policies. For this study, we designed a total of eight policy proposals involving a policy trade-off. Policy proposals covered four different types of trade-offs (taxes, gun control, environmental regulation, and voting rights). For each type of trade-off, we developed a Republican and a Democratic version of the policy proposal, leaving us with a total of eight policy proposals (see Appendix A for full text of all policies). Each participant was presented with four policy proposals, one of each type of trade-off used in this study (taxes, gun control, environmental regulation, and voting rights). The policies were counterbalanced between participants and the order randomized within participant such that participants always saw two policy trade-offs proposed by Republicans, and two policy trade-offs proposed by Democrats. In total, participants we assigned to one of six conditions—see Table 1 below.

Next, participants answered a series of questions about their views on Democrats and Republicans. Finally, participants answered demographic questions.

[2] https://osf.io/qy23x/?view_only=1661466f2a024a609e81c3105bdc93e9

[3] For transparency, we would like to note that at the point in our research program when data was collected for the first two studies, we had not yet devised the hypothesis that the operating mechanism underlying the partisan trade-off bias was primarily trust—as opposed to more general negative attitudes. Therefore, in these studies we did not measure trust specifically and focused only on the relationship between this new bias and general negative partisanship. In studies 4 and 5, we use a multi-item measure of trust to try to assess its unique contribution from general dislike.

4.3. Measures

Political Partisanship. Participants rated their political partisanship on a 5-point scale from (1) Very Democratic to (5) Very Republican.[4] Higher scores on this variable indicate greater identification with the Republican party.

Perceived Intentionality of Side Effects (SE). For each policy, participants rated how intentional they perceived the respective side effect to be on a 5-point scale (from 1-Not at all to 5-Extremely). Responses regarding Democratic policies were averaged to create a single *Democratic SE Intention* score. *Republican SE Intention* scores were calculated similarly. Higher scores on these variables indicate higher perceived intentionality of side effects.

Perceived Intentionality of Primary Effects (PE). For each policy, participants rated how intentional they perceived the respective primary policy outcome to be on a 5-point scale (from 1-Not at all to 5-Extremely). Responses regarding Democratic policies were averaged to create a single *Democratic PE Intention* score. *Republican PE Intention* scores were calculated similarly. Higher scores on these variables indicate higher perceived intentionality of primary effects.

Negative Attitudes. To capture affective political polarization, we broadly assessed negative attitudes towards Republicans and Democrats using a variety of measures from the Pew Research Center. Importantly, at this point we had not devised the hypothesis that it is trust specifically that primarily drives the bias. Hence, we were not attempting to capture trust in this study. For our measure of negative attitudes, participants rated their feelings towards Democrats and Republicans on a 1–100 "feeling thermometer" scale, from (1) Very cold/unfavorable to (100) Very warm/favorable (Weisberg & Rusk, 1970). They also rated Democrats and Republicans on a variety of traits[5] (e.g., close-minded, lazy) using a 5-point scale (from 1-Not at all to 5-Extremely). Ratings of positive traits were reverse coded. Participants also rated the extent to which Democrats and Republicans made them feel frustrated, angry, afraid, and disgusted (5-point scale from 1-Not at all to 5-Extremely). Finally, participants rated how difficult it would be to get along with a new neighbor if they were Republican/Democratic (5-point scale from 1-Not difficult at all to 5-Extremely difficult), how they would feel if an immediate family member married a Republican/Democratic (7-point scale from 1-Extremely unhappy to 7-Extremely happy), and the extent to which they agreed that Republican/Democratic policies are so misguided that they threaten the nation's well-being (7-point Likert scale from 1-Strongly disagree to 7-Strongly agree). All Republican and Democratic items were standardized and averaged to make a *Republican Negative Attitudes* ($\alpha = 0.95$) score and a *Democratic Negative Attitudes* ($\alpha = 0.94$) score, respectively.

[4] In the first two studies, we also measured participants' political ideology on a 5-point scale from (1) Very Liberal to (5) Very Conservative. We would like to note that we use political partisanship in all our analyses across all studies, given that in our experiments participants are always reacting to target partisanship. But we would also like to note that, expectedly, responses to political partisanship and political ideology were highly correlated ($r(330) = 0.80$) and results were nearly identical independently of the measure used—see SM. The fact that our results replicate with both measures is likely a product of ideological alignment along party lines. 85% of Americans identify as Democrats and Republicans and, while liberals and conservatives used to be distributed more equally between the two parties, today the former are mostly Democrats and the latter are mostly Republicans (Finkel, Bail, Cikara, Ditto, Iyengar, Klar, Mason, McGrath, Nyhan, Rand, Skitka, Tucker, Van Bavel, Wang, & Druckman, 2020).

[5] See Appendix B for all items.

Table 2
Correlations between perceived intentionality of side effects of different policies by condition, and perceived intentionality of primary effects of different policies by condition. Note: $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

	Perceived Intentionality Correlations			
	Republican Policies		Democratic Policies	
Conditions	Side-Effects	Primary Effects	Side-Effects	Primary Effects
1	0.58***	0.47***	0.67***	0.24*
2	0.20	0.65***	0.62***	0.52***
3	0.71***	0.40***	0.17	0.33**
4	0.57***	0.30*	0.63***	0.59***
5	0.57***	0.48***	0.29*	0.67***
6	0.56***	0.20	0.54***	0.74***

4.4. Results[6]

We started our analyses by looking at the correlations—for each of the six conditions—between perceived intentionality of the side effects of Republican and Democratic policy trade-offs, and perceived intentionality of the primary effects of Republican and Democratic policy trade-offs. With very few exceptions—see Table 2 below—the correlations between these measures were sufficiently high to justify our next step in the analyses: the generation of average scores for intentionality ratings of side effects and primary effects by party who proposed the policy for each condition. More specifically, we created the following four average scores, per condition: (1) perceived intentionality of the side effects of Republican policies, (2) perceived intentionality of the primary effects of Republican policies, (3) perceived intentionality of the side effects of Democratic policies, and (4) perceived intentionality of the primary effects of Democratic policies.

After calculating these average scores, we transformed the data into long form by creating a dummy variable for the party proposing the policy trade-off (1 = Republican Policy; 0 = Democratic Policy). In this way, the dataset in long form included only two measures of intentionality per participant: (1) perceived intentionality for side effects, and (2) perceived intentionality for primary effects. Lastly, with this dataset, we used bivariate correlations to investigate the relationship between main variables and regression analyses to test the effect of political partisanship, party proposing the policy (1 = Republican Policy; 0 = Democratic Policy), and the interaction between these two variables on perceived intentionality of side effects and perceived intentionality of primary effects.

All results were in the predicted direction. Correlations among main variables are presented in Table 3. The more individuals identified with the Republican Party, the more intentional they perceived the side effects of Democratic policies to be ($r(330) = 0.44$, $p < .001$). In contrast, the more individuals identified with the Democratic Party, the more intentional they perceived the side effects of Republican policies to be ($r(330) = -0.57$, $p < .001$).

The results of regression analysis showed a significant two-way interaction of political partisanship by policy trade-off proponent (Republicans versus Democrats) on perceived intentionality of side effects ($b = -0.97$, $SE = 0.06$, $p < .001$, 95% CI [−1.10, −0.84], $F(3, 660) = 77.45$, $R^2 = 0.26$, *Cohen's* $f = 0.59$)—see Fig. 1. The more participants identified with the Democratic Party, the more they attributed intentionality to the side effects of Republican policies and the less they attributed intentionality to the side effects of Democratic policies. Republicans showed the same pattern, such that the more participants identified with the Republican Party the more they perceived the side effects of Democratic policies as intentional and the less they perceived

[6] See SM for supplemental analyses using linear mixed effect models with participant as a random factor. All results replicated using this alternative method.

Table 3
Bivariate correlations among Political Partisanship, Republican Primary Effect Intention, Republican Side Effect Intention, Republican Negative Attitudes, Democratic Primary Effect Intention, Democratic Side Effect Intention, and Democratic Negative Attitudes; *p = .05, **p < .01, *** p < .001.

	Political Partisanship	Rep. Side Effect Intention	Rep. Primary Effect Intention	Rep. Negative Attitudes	Dem. Side Effect Intention	Dem. Primary Effect Intention
Rep. Side Effect Intention	−0.57***	1				
Rep. Primary Effect Intention	0.21***	−0.20***	1			
Republican Negative Attitudes	−0.68***	0.66***	−0.10*	1		
Dem. Side Effect Intention	0.44***	−0.19***	0.21***	−0.34***	1	
Dem. Primary Effect Intention	−0.35***	0.24***	0.21***	0.37***	−0.26***	1
Democratic Negative Attitudes	0.64***	−0.40***	0.17**	−0.45***	0.54***	−0.39***

the side effects of Republican policies as intentional.

The results of regression analysis also revealed a significant two-way interaction of political partisanship by policy trade-off proponent (Republicans versus Democrats) on perceived intentionality of the primary effects of the policies ($b = 0.49$, $SE = 0.06$, $p < .001$, 95% CI [0.36, 0.61], $F(3, 660) = 25.10$, $R^2 = 0.10$)—see Fig. 2. The more participants identified with the Democratic Party, the more they attributed intentionality to the primary effects of Democratic policies and the less they attributed intentionality to the primary effects of Republican policies. Republicans again showed the same pattern, such that the more participants identified with the Republican Party the more they perceived the primary effects of Republican policies as intentional and the less they perceived the primary effects of Democratic policies as intentional.

Lastly, we expected negative attitudes to be associated with biased attributions of intentionality to the side and primary effects of policy trade-offs. Results of correlational analyses showed that greater negative attitudes towards Democrats (*Democratic Negative Attitudes*) were positively correlated with perceived intentionality of Democratic side effects ($r(330) = 0.54$, $p < .001$) and greater negative attitudes towards Republicans (*Republican Negative Attitudes*) were positively correlated with perceived intentionality of Republican side effects ($r(330) = 0.66$, $p < .001$). Additionally, participants higher in *Democratic Negative Attitudes* perceived the primary effects of Democratic policies as less intentional ($r(330) = -0.39$, $p < .001$), and those higher in *Republican Negative Attitudes* perceived the primary effects of Republican policies as less intentional ($r(330) = -0.10$, $p = .05$).

4.5. Discussion

In this first study, we have documented the partisan trade-off bias. That is, we have provided evidence for a symmetrical bias in participants' attribution of intentionality to the side effects and the primary effects of trade-off policies across the political spectrum. For example, when Democrats proposed a gun control policy with a primary goal of reducing the number of gun deaths and an unavoidable side effect related with the reduction of people's liberties, Republicans did not charitably interpret the intentions behind this policy. Quite the contrary, Republicans attributed comparatively less intentionality to the primary goal of this policy and more intentionality to the side effect of this policy. The partisan trade-off bias was symmetric. Both Republicans and

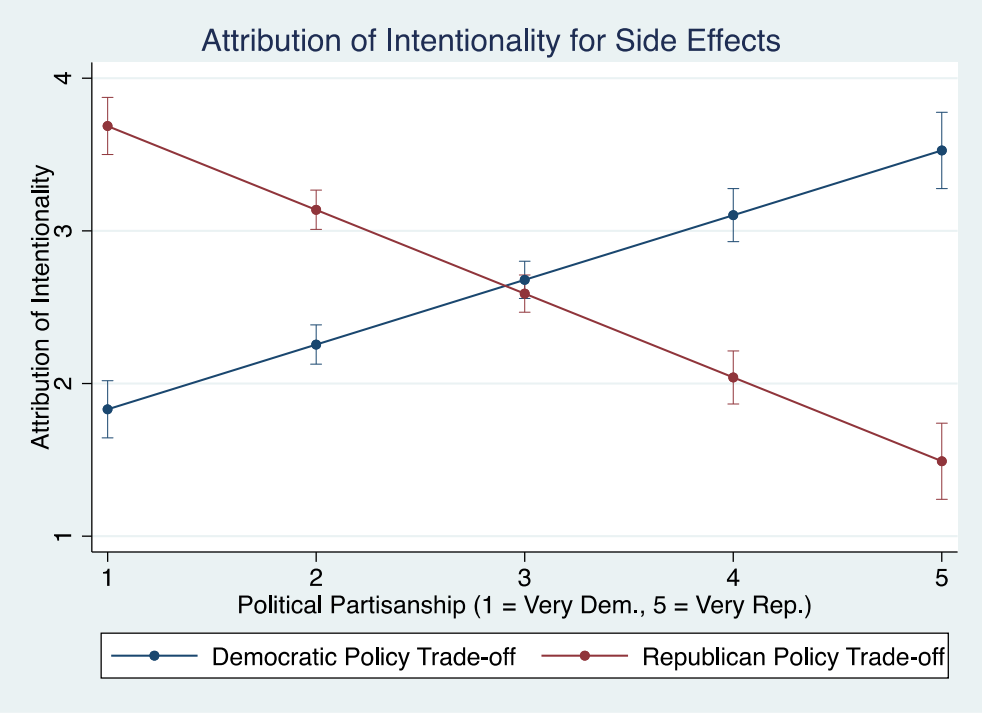

Fig. 1. Two-way interaction of political partisanship by policy trade-off proponent (Republican Policy Trade-off or Democratic Policy Trade-off) on perceived intentionality of side effects. Error bars represent 95% confidence intervals. (Study 1).

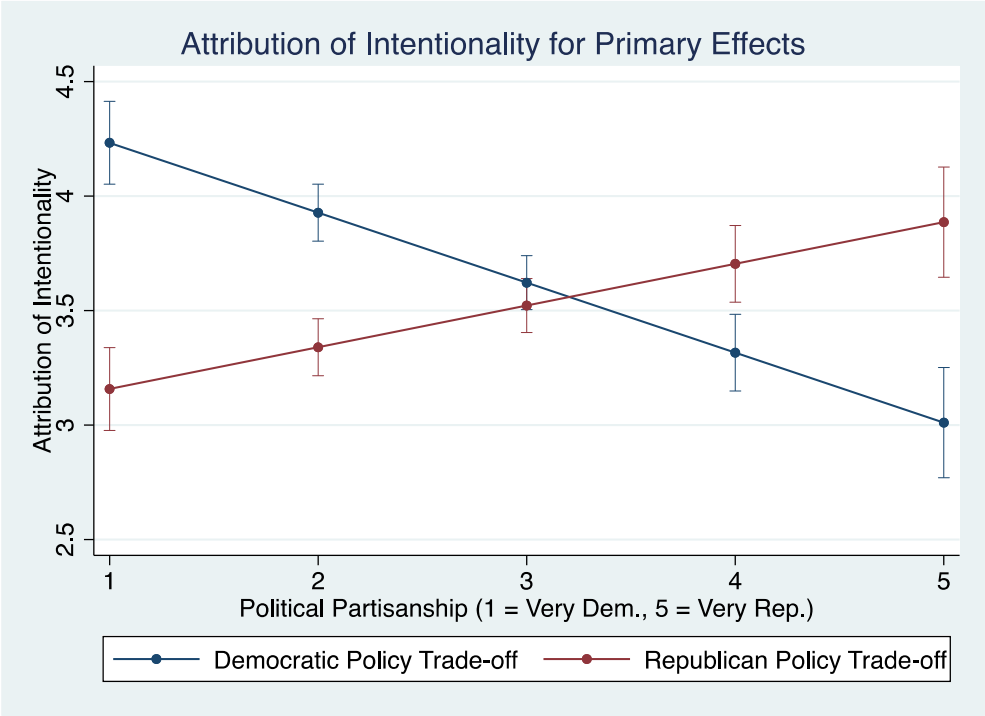

Fig. 2. Two-way interaction of political partisanship by policy trade-off proponent (Republican Policy Trade-off or Democratic Policy Trade-off) on perceived intentionality of primary effects. Error bars represent 95% confidence intervals. (Study 1).

Democrats have systematically maligned the intentions behind the policy trade-offs of contrapartisans.

In addition to documenting a new bias in the sphere of policy evaluation, we have shown that negative partisanship is highly correlated with the partisan trade-off bias in the expected directions. Both Democrats and Republicans that held stronger negative attitudes towards contrapartisans also more strongly displayed this bias in their evaluation of trade-off policies. These findings suggest that this bias is associated with political polarization. In the fourth and fifth studies we delve more deeply into the mechanism driving the partisan trade-off bias.

While in this first study we were able to document the partisan trade-off bias across a variety of policies (taxes, environment, gun control, and voting), it is possible—and expected—that people already have pre-existing attitudes regarding these specific policy trade-offs. If that is the case, people may also have pre-existing beliefs about the intentionality behind these policy trade-offs. Given the highly polarized current political environment, it would be reasonable to expect the content of specific policy trade-offs to potentially trigger associations with either the Republican or the Democratic parties. These associations can encompass assumptions about the intentionality of political members of these parties that are not directly related with the structure of policy trade-offs. In an attempt to gather evidence – to the best of our ability – that speaks directly to the structure of policy trade-offs (a desired primary effect that can only be achieved at the cost of an undesired side effect), we will use content-free policy trade-offs in the next study. These content-free policies will express the structure of a trade-off without reference to any specific type of policies that could be reasonably associated with other intentionality aspects. In this way, we will be able to focus on the processes related with thinking about any policies that involve trade-offs.

5. Study 2

In the second study we tested whether the content of the policies drove the results of the first study. Study 2 completes the test of the first two hypotheses by testing the independence of the partisan trade-off bias from the specific content of the policy trade-offs used in the first study. To do so, we developed content-free policy cases that replicate the structure of a policy trade-off but that lack any type of specific policy content.

5.1. Participants

We recruited 181 participants via Turk Prime and excluded 8 for failing open ended questions that functioned as attention checks. Our final sample was 173 participants (53.2% male; $M_{age} = 40.72$, $SD_{age} = 12.61$). Fifty percent of participants reported being Democratic and 34% reported being Republican. Participants' reports of their political ideology matched their political partisanship closely: 46.8% of participants reported being liberal or slightly liberal and 34.7% rated themselves conservative or slightly conservative (18.5% reported being neither liberal nor conservative). Sensitivity power analyses with linear multiple regression using three predictors and standard criteria, *alpha* = 0.05 and *beta* = 0.80, revealed a minimum effect size that could be detected by our sample equal to $f^2 = 0.064$ or $R^2 = 0.004$ (Critical $F(3, 169) = 2.66$).

5.2. Procedures

We designed a between-subjects design with two conditions: one condition in which participants were presented with a content-free policy trade-off proposed by members of the Republican Party, and another condition in which they were presented with an identical content-free policy trade-off proposed by members of the Democratic Party. After providing informed consent, participants reported their political partisanship and political ideology. They were then randomly assigned to one of these two conditions: Democratic Policy Condition or Republican Policy Condition. Participants in each condition read a content-free description of a policy trade-off, and the only different between conditions was the party proposing the policy. Here is what participants read:

Republicans (Democrats) say they want to implement Policy A. They support Policy A claiming it will help some people, but knowing it also has the potential of hurting some other people.
People have different views about the intentions of this policy. What is your opinion?

After reading this brief and completely content-free description of a policy trade-off, all participants completed the following measures: perceived intentionality of the primary effect, perceived intentionality of the side effect, and measures of political polarization plus demographics.

5.3. Measures

Political Partisanship. This is the same measure used in the first study (5-point scale from 1-Very Democratic to 5-Very Republican). Higher scores on this variable indicate greater identification with the Republican Party.

Perceived Intentionality of Side Effect (SE). Each participant only read about one policy trade-off in this study (Policy A Republican or Democratic) and rated how intentional they perceived the side effect to be on a 5-point scale (from 1-Not at all to 5-Extremely).

Perceived Intentionality of Primary Effect (PE). Each participant also rated how intentional they perceived the primary effect of the policy to be on a 5-point scale (from 1-Not at all to 5-Extremely).

Negative Attitudes. This measure was calculated just as in the first study. Again, Republican and Democratic items were standardized and averaged to create a *Republican Negative Attitudes* score ($\alpha = 0.95$) and a *Democratic Negative Attitudes* score ($\alpha = 0.96$), respectively.

5.4. Results

We replicated all the results from the first study with this stricter experimental paradigm—correlations among main variables are presented in Table 4. The more individuals identified with the Republican Party, the more intentional they perceived the side effects of Democratic policies to be ($r\,(80) = 0.59$, $p < .001$). In contrast, the more individuals identified with the Democratic Party, the more intentional they perceived the side effects of Republican policies to be ($r\,(89) = -0.28$, $p < .01$).

The perceived intentionality of the side effects and the primary effects of policy trade-offs across the political spectrum are depicted in Fig. 3a and 3b. In our analyses, we dummy coded conditions (Republican versus Democratic policy proponent) and used regression analyses to investigate the effect of condition by political partisanship on perceived intentionality of side effects and perceived intentionality of primary effects. The results of regression analysis showed a significant two-way interaction of political partisanship by condition (Republican versus Democratic policy trade-off) on perceived intentionality of the side effect ($b = -0.77$, $SE = 0.12$, $p < .001$, 95% CI [−1.02, −0.52], $F(3, 169) = 15.39$, $R^2 = 0.21$). As can be seen in Fig. 3a, the more participants identified with the Democratic Party, the more they attributed intentionality to the side effect of the Republican policy and the less they attributed intentionality to the side effect of the Democratic policy. Republicans showed the same pattern, such that the more they identified with the Republican Party the more they perceived the side effect of the Democratic policy as intentional and the less they perceived the side effect of the Republican policy as intentional.

We found similar results for attribution of intentionality to the primary effects of the policy trade-offs (Fig. 3b). The results of regression analysis revealed a significant two-way interaction of political partisanship by condition (Republican versus Democratic policy trade-off) on perceived intentionality of primary effects ($b = -0.49$, $SE = 0.13$, $p < .001$, 95% CI [0.23, 0.75], $F(3, 169) = 6.98$, $R^2 = 0.11$). As can be seen in Fig. 3b, the more participants identified with the Democratic Party, the more they attributed intentionality to the primary effect of the Democratic policy and the less they attributed intentionality to the primary effect of the Republican policy. Republicans showed the same pattern, such that the more participants identified with the Republican Party, the more they perceived the primary effect of the Republican policy as intentional and the less they perceived the primary effect of the Democratic policy as intentional.

Lastly, we also replicated the correlations between negative attitudes and the biased attributions of intentionality to the side effects and the primary effects of Republican and Democratic policy trade-offs. Higher scores on *Democratic Negative Attitudes* were positively correlated with perceived intentionality of the Democratic side effect ($r\,(80) = 0.51$, $p < .001$) and negatively correlated with perceived intentionality of the Democratic primary effect ($r\,(80) = -0.48$, $p < .001$). Higher scores on *Republican Negative Attitudes* were positively correlated with perceived intentionality of the Republican side effect ($r\,(89) = 0.43$, $p < .001$) and negatively correlated with the perceived intentionality of the Republican primary effect ($r\,(89) = -0.46$, $p < .001$).

5.5. Discussion

In the second study we developed and put to test a content-free policy trade-off paradigm. Our goal was to eliminate the content of policy trade-offs in our experimental design, so as to be able to focus our investigation on the effects of the trade-off structure (a primary and positive intended effect coupled with a negative and unintended side effect) on the attribution of intentionality by partisans across the political spectrum. This paradigm has enabled us to run a stringent test on the partisan trade-off bias. As predicted, we replicated the effects from the first study, providing further evidence for an attributional bias in the interpretation of policy trade-offs. Together, the results of the first and

Table 4
Bivariate correlations between main variables; *p = .05, **p < .01, ***p < .001. Note: correlations were calculated for each condition: one condition in which participants evaluated primary and side effects of a Democratic policy trade-off ($N = 82$) and one in which participants evaluated primary and side effects of a Republican policy trade-off ($N = 91$). The correlation between Democratic and Republican negative attitudes is the only correlation based on the entire sample ($N = 173$).

	Political Partisanship	Rep. Side Effect Intention	Rep. Primary Effect Intention	Rep. Negative Attitudes	Dem. Side Effect Intention	Dem. Primary Effect Intention
Rep. Side Effect Intention	−0.28**	1				
Rep. Primary Effect Intention	0.20*	−0.36***	1			
Republican Negative Attitudes	−0.77***	0.42***	−0.46***	1		
Dem. Side Effect Intention	0.59***	–	–	−0.35**	1	
Dem. Primary Effect Intention	−0.34**	–	–	0.29**	−0.35**	1
Democratic Negative Attitudes	0.72***	−0.13	0.10	−0.50***	0.51***	−0.48***

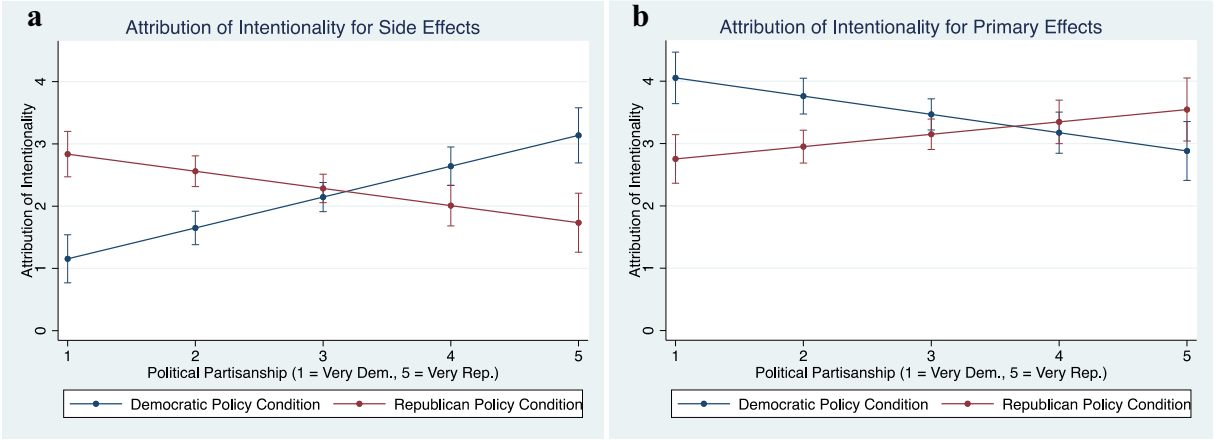

Fig. 3. a (left). Two-way interaction of political partisanship by condition (Republican Policy or Democratic Policy) on perceived intentionality of side effects. Error bars represent 95% confidence intervals. (Study 2). b (right). Two-way interaction of political partisanship by condition (Republican Policy or Democratic Policy) on perceived intentionality of primary effects. Error bars represent 95% confidence intervals. (Study 2).

second studies complete the test of the first two hypotheses (*The Partisan Side Effect Bias* and *The Partisan Primary Effect Bias*).

6. Study 3

In study 3, we used a negotiations design to test whether the partisan trade-off bias uniquely contributes to decreased willingness to accept policy deals from contrapartisans. This study enabled us to test *The Policy Deals* hypothesis (H3).

6.1. Participants

We recruited 200 participants via Prolific and excluded 14 for failing attention checks. Our final sample was 186 participants (51% male; $M_{age} = 30.26$, $SD_{age} = 10.67$). Given our repeated measures design, our long format data had an $N = 744$. Fifty-nine percent of participants reported being Democratic and 10.75% reported being Republican. We conducted sensitivity power analyses for logistic regression using standard criteria, *alpha* = 0.05 and *beta* = 0.80. They revealed a minimum effect size that could be detected by our sample for predicting the likelihood of accepting Republican policy deals (in our sample, proportion of Republican deals accepted was equal to 93 out of 375) equal to an odds ratio of 0.71 (critical $z = -1.95$) and for predicting the likelihood of accepting Democratic policy deals (in our sample, proportion of Democratic deals accepted was equal to 276 out of 369) equal to an odds ratio of 0.72 (critical $z = -1.95$).

6.2. Procedures

After providing informed consent, participants were informed that they were going to be asked to indicate whether they would like their party representatives to accept or reject a number of policy deals from Democratic and Republican Senators. Participants were then randomly assigned to one of eight conditions. Each condition contained a total of four policy deals across four policy domains: taxes, gun control, environmental regulation, and voting identification. The policies were counterbalanced between participants and the order randomized within participant. Here is an example of a policy deal proposal:

> "A Republican Senator wants to reform the tax code. After several rounds of negotiation between members of the Senate, the Republican Senator proposed the following deal: Implementing a tax break with the goal of creating more jobs, but knowing this policy also has the potential of cutting benefits for the poor."

After each policy deal, participants responded to our measures of willingness to accept or reject the deal and attribution of intentionality to the primary and side effects of each policy deal. Finally, participants answered demographic questions.

6.3. Measures

Political Partisanship. Participants rated their political partisanship on a 7-point scale from 1-Very Democratic to 7-Very Republican. Higher scores on this variable indicate greater identification with the Republican Party.

Willingness to Accept or Reject Policy Deal. Each participant was presented with four policy deals and indicated whether they would like their party representatives to accept or reject the deal on a binary scale (0 = reject, 1 = accept).

Perceived Intentionality of Side Effect (SE). Participants rated how intentional they perceived the side effect of each policy deal to be on a 7-point scale (from 1-Not at all to 7-Extremely).

Perceived Intentionality of Primary Effect (PE). Participants rated how intentional they perceived the primary effect of each policy deal to be on a 7-point scale (from 1-Not at all to 7-Extremely).

6.4. Results[7]

Given the repeated measures design, we analyzed the data in long format, which gave us a total of 369 Democratic policy deals and 375 Republican policy deals (total $N = 744$). The partisan trade-off bias was replicated in this sample and with this new design (policy trade-offs framed in terms of policy deals). Considering only the Democratic policy deals, the more individuals identified with the Republican Party, the more intentional they perceived the side effects of Democratic policy deals to be (r (367) = 0.27, $p < .001$). In contrast, considering Republican policy deals only, the more individuals identified with the Democratic Party, the more intentional they perceived the side effects of Republican policy deals to be (r (373) = −0.28, $p < .001$). In a similar fashion, higher identification with the Republican Party was associated with higher attribution of intentionality to the primary effects of Republican policy deals (r (373) = 0.17, $p < .01$), and higher identification with the Democratic Party was associated with higher attribution of intentionality to the primary effects of Democratic policy deals (r (367) = −0.36, $p < .001$).

[7] See SM for supplemental analyses using linear mixed effect models with participant as a random factor. All results replicated using this alternative method.

We used logistic regressions to investigate the likelihood of accepting policy deals from the opposing party. As predicted, the results showed that political partisanship significantly predicted the likelihood of accepting deals from contrapartisans. Considering Republican policy deals only, higher identification with the Republican party predicted higher likelihood of accepting deals from Republican senators (*Odds Ratio* = 2.53; *SE* = 0.30; $p < .001$; 95% CI [2.00, 3.19]; $\chi^2(1) = 83.09$; *Pseudo* $R^2 = 0.21$). More specifically, higher identification with the Republican party meant that the odds of accepting deals from Republican senators was 2.53 times the odds of rejecting these deals. Considering Democratic policy deals only, higher identification with the Republican party predicted higher likelihood of rejecting deals from Democratic Senators (*Odds Ratio* = 0.55; *SE* = 0.06; $p < .001$; 95% CI [0.45, 0.68]; $\chi^2(1) = 36.28$; *Pseudo* $R^2 = 0.09$). More specifically, higher identification with the Republican party meant that the odds of accepting deals from Democratic senators was 0.55 times the odds of rejecting these deals. Importantly, the partisan trade-off bias (biased attribution of intentionality to the side and primary effects of contrapartisans' policy trade-offs) predicted willingness to accept policy deals both for Democrats and Republicans—see Table 5 below.

These results shed light on an important practical implication of the partisan trade-off bias. Supporting policies that involve trade-offs lie at the heart of politics. We show that biased attributions of intentionality uniquely contribute to the public's unwillingness to support policy trade-offs proposed by contrapartisans.

7. Study 4

In the first three studies we have documented the partisan trade-off bias and showed evidence for the strength of this bias: it occurs even using content-free policy trade-offs. Yet the measure of negative partisanship that was used in the previous studies (including a series of items from the Pew Research Center) is not well-suited to capture differences between the two main psychological constructs that are potentially driving this bias: dislike versus distrust. Much research shows that trust is a necessary condition for negotiations, while "liking" can lead to positive outcomes but is not a requirement for successful agreements (Dirks & Ferrin, 2001). To examine the mechanism that underlies the partisan trade-off bias, in the fourth study we developed measures that more precisely allow for a test of the relative contribution of these two potential mediators (dislike versus distrust). We tested for mediation via both dislike and distrust. We predicted that distrust would be the main driver of the effect of political partisanship on the partisan trade-off bias. In this study, we provide evidence in support of the first part of *The Trust* hypothesis (H4): that the relative distrust in contrapartisans relative to copartisans will mediate the effect of political partisanship on attribution of intentionality.

Table 5
Logistic regressions predicting willingness to accept Democratic and Republican policy deals. Note: standard errors in parentheses; 744 nested data observations; *** p < .001, ** p < .01, * p < .05.

	Willingness to Accept Policy Deals			
	Democratic Policy Deals		*Republican Policy Deals*	
Independent Variables	Odds Ratio	95% CI	Odds Ratio	95% CI
Perceived Intentionality Side Effect	0.82*	[0.69, 0.96]	0.67***	[0.55, 0.81]
	(0.07)		(0.06)	
Perceived Intentionality Primary Effect	1.54***	[1.24, 1.92]	1.18	[0.95, 1.47]
	(0.17)		(0.13)	
Political Partisanship	0.67**	[0.53, 0.84]	2.21***	[1.74, 2.80]
	(0.07)		(0.26)	
Pseudo R-squared	0.18		0.27	
* < 0.05, ** < 0.01, *** < 0.001.				

7.1. Participants

We recruited 200 participants via Turk Prime and excluded 16 for failing open ended questions that functioned as attention checks. Our final sample was 184 participants (53.2% male; $M_{age} = 40.7$, $SD_{age} = 12.6$). Forty-eight percent of participants reported being Democratic and 34% reported being Republican. Sensitivity power analyses with linear multiple regression using three predictors and standard criteria, *alpha* = 0.05 and *beta* = 0.80, revealed a minimum effect size that could be detected by our sample equal to $R^2 = 0.004$ (Critical $F(3, 180) = 2.65$). We also conducted a sensitivity power analyses for a parallel mediation model using online tools for power calculations in Structural Equation Modeling (Want & Rhemtulla, in press) and mediation analyses (Schoemann, Boulton, & Short, 2017) using standard criteria, *alpha* = 0.05 and *beta* = 0.80, and found that the minimum standardized effect we could detect for the *a* and *b* path given our sample size was $b = 0.35$ and the minimum standardized indirect effect (*a*b)* was $b = 0.42$.

7.2. Procedures

Procedures were analogous to Study 2. We conducted a between-subjects design with two conditions: Republican Policy and Democratic Policy. The only difference is that, in this study, we included specific measures to capture the difference between dislike of contrapartisans versus distrust in contrapartisans.

7.3. Measures

Political Partisanship. This is the same measure used in the previous studies. Higher scores on this variable indicate greater identification with the Republican Party.

Perceived Intentionality of Side Effect (SE). Each participant read about only one policy (either a Republican Policy or a Democratic Policy) and rated how intentional they perceived the side effect to be on a 5-point scale (from 1-Not at all to 5-Extremely).

Perceived Intentionality of Primary Effect (PE). Each participant also rated how intentional they perceived the positive primary effect of the policy to be on a 5-point scale (from 1-Not at all to 5-Extremely).

Like/Dislike of Partisans and Contrapartisans. We measured how much people like/dislike members of both the Republican and the Democratic parties using a total of six items, measured on a 5-point scale (from 1-Strongly disagree to 5-Strongly agree). The items were: "I like Republicans (Democrats)," "I have positive feelings towards Republicans (Democrats)," and "I dislike Republicans (Democrats)." We reverse-scored the two dislike items and averaged the three Republican items to create a measure of "Like Republicans" ($\alpha = 0.94$) and the three Democratic items to create a measure of "Like Democrats" ($\alpha = 0.93$). Then, we subtracted "Like Democrats" from "Like Republicans" to create one measure that captured how much people like Republicans more than Democrats (Like Republicans *minus* Like Democrats; $\alpha = 0.88$).

Trust in Partisans and Contrapartisans. We measured how much people trust/distrust members of both the Republican and the Democratic parties using a total of six items, measured on a 5-point scale (from 1-Strongly disagree to 5-Strongly agree). The items were: "I believe that Republicans (Democrats) can be trusted," "I can trust Republicans (Democrats) to be sincere," and "Republicans (Democrats) are untrustworthy." We reverse-scored the two untrustworthy items and averaged the three Republican items to create a measure of "Trust Republicans" ($\alpha = 0.91$) and the three Democratic items to create a measure of "Trust Democrats" ($\alpha = 0.94$). Then, we subtracted "Trust Democrats" from "Trust Republicans" to create one measure that captures how much people trust Republicans more than Democrats (we called it *Trust Rep_Dem*; $\alpha = 0.82$).

7.4. Results

Firstly, we replicated the partisan trade-off bias in this study. The perceived intentionality of the side effects and of the primary effects of policy trade-offs across the political spectrum are depicted in Fig. 4a and 4b. The results of regression analysis showed that political partisanship interacted with condition (Republican versus Democratic policy) to significantly predict attribution of intentionality to side effects (b = 0.61, SE = 0.12, $p < .001$, 95% CI [0.37, 0.85], $F(3, 181) = 8.59$, R^2 = 0.12) and to primary effects ($b = -0.63$, SE = 0.13, $p < .001$, 95% CI [−0.90, −0.37], $F(3, 181) = 8.21$, $R^2 = 0.12$) of the policy trade-offs.

Secondly, we used mediation analyses to model trust in Republicans and Democrats as a mechanism by which political partisanship relates to attributions of intentionality to side effects, while controlling for liking of Republicans and Democrats. Although we cannot infer causality from the relations specified in the model or rule out other models (e.g., trust in each party or liking of each party could theoretically influence political partisanship), we use the model to contrast the effect of the partisan trust gap versus that of the partisan liking gap, and to set up a theoretical model that will be tested experimentally in Study 5.

Our results showed that while political partisanship predicted trust in Republicans and Democrats and liking of Republicans and Democrats, it was the gap in trust in contrapartisans relative to trust in copartisans, and not the partisan liking gap, that mediated the effect of political partisanship on perceived intentionality of side effects both in the Democratic and the Republican Policy conditions. In the Republican Policy condition, the trust gap in Republicans/Democrats mediated the effect of political partisanship on intentionality attribution to side effect (indirect standardized effect via *trust*: $b = -0.33$, $p < .001$; indirect standardized effect via *like*: $b = 0.04$, $p = .752$; see Fig. 5a). In the Democratic Policy condition, the trust gap in Republicans/Democrats mediated the effect of political partisanship on intentionality attribution to side effect (indirect standardized effect via *trust*: $b = 0.35$, $p < .01$; indirect standardized effect via *like*: $b = 0.01$, $p = .932$; see Fig. 5b). As a robustness check, we ran an identical mediation model but added two paths representing a feedback loop in which the trust gap and the liking gap in Republicans/Democrats predicted political partisanship (we fixed those parameters to an effect of 0.4), and found that, for the Republican policy condition, the trust gap in Republicans/Democrats still mediated the effect of political partisanship on intentionality attribution to side effect (indirect standardized effect via *trust*: $b = -0.19$, $p < .01$; indirect standardized effect via *like*: $b = 0.00$, $p = .829$). Similarly, for the Democratic policy condition, the trust gap in Republicans/Democrats still mediated the effect of political partisanship on intentionality attribution to side effect (indirect standardized effect via *trust*: $b = 0.13$, $p < .05$; indirect standardized effect via *like*: $b = 0.00$, $p = .927$).

7.5. Discussion

In the fourth study we were able to, once again, replicate the partisan trade-off bias. Additionally, we provided initial evidence for the mechanism driving this bias. While one could have predicted that this bias is a product of both like/dislike of and trust/distrust in members of the opposite party, our results suggest that the partisan trade-off bias is more closely related with trust. In the next study we manipulate perceived trust in order to provide a stricter test for this mechanism and also to investigate the possibility of an intervention to reduce this bias.

8. Study 5

In study 5, we designed an intervention with the goal of manipulating trust in Republican and Democratic policy makers. This afforded the opportunity to both further test our hypothesized mechanism and to provide a potential solution to the partisan trade-off bias. This study provides evidence for the second part of *The Trust* hypothesis (H4): that increasing trust in contrapartisan policy makers should attenuate the partisan trade-off bias. Building on the organizational literature on trust, we manipulated the extent to which the policymaker was seen as trustworthy by not only calling him trustworthy but explaining that he offered many opportunities for voice, that is, for members of both parties to express their views about and influence the policy. In addition to this trust manipulation, we also included a new measure in this study: policy support. Our idea was to test whether the partisan trade-off bias had downstream effects in people's attitudes and behaviors regarding policy trade-offs. Policy support is a crucial attitudinal feature in the political domain, as it is indicative of voting preferences. Hence, in this study, we also tested whether the partisan trade-off bias influenced policy support.

8.1. Participants

We recruited 400 participants via Turk Prime and excluded 39 for failing open ended questions that functioned as attention checks. Our final sample was 361 participants (52% male; M_{age} = 40.08, SD_{age} = 13.46). Sensitivity power analyses with linear multiple regression using four predictors and standard criteria *alpha* = 0.05 and *beta* = 0.80 revealed a minimum effect size that could be detected by our sample equal to $f^2 = 0.033$ or $R^2 = 0.001$ (Critical $F(4, 356) = 2.40$). Forty-seven percent of participants reported being Democratic, 35% reported being Republican, and 18% of participants reported being neither.

8.2. Procedures

We implemented a 2 × 2 between-subjects design manipulating the party proposing the policy trade-off (Republican Policy versus Democratic Policy) and trust in the policy proponent (trust/voice manipulation versus control). The control conditions were an identical replication of the Republican and Democratic versions of the *Policy A* vignette designed for Studies 2 and 4. The manipulation is described in the measures subsection. All other procedures and measures were identical to the ones in the fourth study.

8.3. Measures

Manipulations. The manipulation included the following modified vignette followed by a writing prompt:

> "A Democratic [Republican] policy maker wants to implement Policy A. He supports Policy A claiming it will help some people, but knowing it also has the potential of hurting some other people. This an honest and trustworthy policy maker who is committed to improving the lives of all members of society, not just some special groups. As evidence of his commitment, he spent the last year holding meetings with representatives of civil groups from across the political spectrum in order to adjust and reform Policy A until it was aligned with everyone's goals. People have different views about the intentions of the policy. What is your opinion?"
>
> Next page:
>
> "Before you give your opinion, please write down a couple of reasons why this Democratic [Republican] policy maker might really be honest and trustworthy."

Importantly, our manipulation tackles an important aspect involved in building trust in leaders: the presence of voice. In this way, our goal with this manipulation is to specifically influence trust in contrapartisans, and not how much people like/dislike contrapartisans (we measured both to check for the manipulations success).

Political Partisanship. This is the same measure used in the previous studies. Higher scores on this variable indicate greater identification with the Republican Party.

Perceived Intentionality of Side Effect (SE). Participants read about only one policy (Policy A Republican or Policy A Democratic, with or

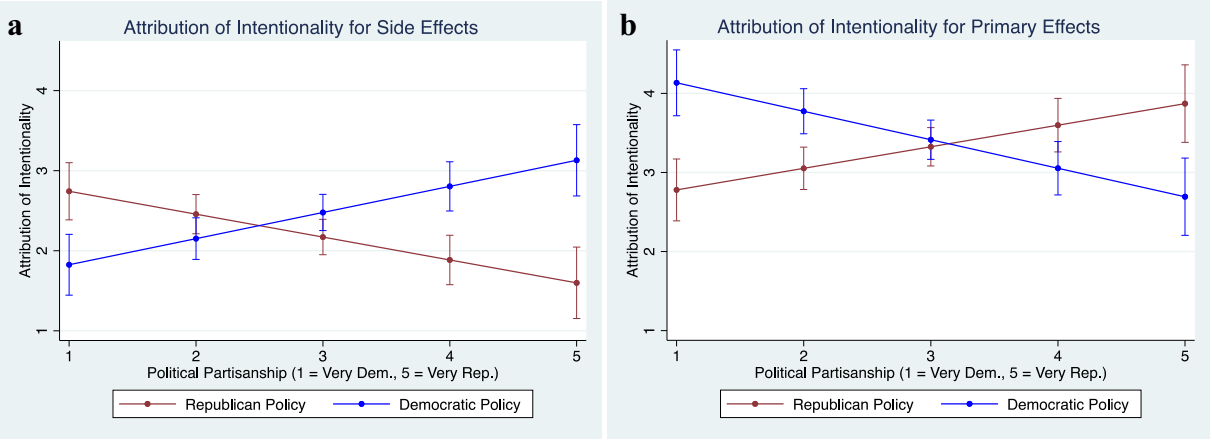

Fig. 4. a (left). Two-way interaction of political partisanship by policy trade-off proponent (Republican Policy Trade-off or Democratic Policy Trade-off) on perceived intentionality of side effects. Error bars represent 95% confidence intervals. (Study 4). b (right). Two-way interaction of political partisanship by policy trade-off proponent (Republican Policy Trade-off or Democratic Policy Trade-off) on perceived intentionality of primary effects. Error bars represent 95% confidence intervals. (Study 4).

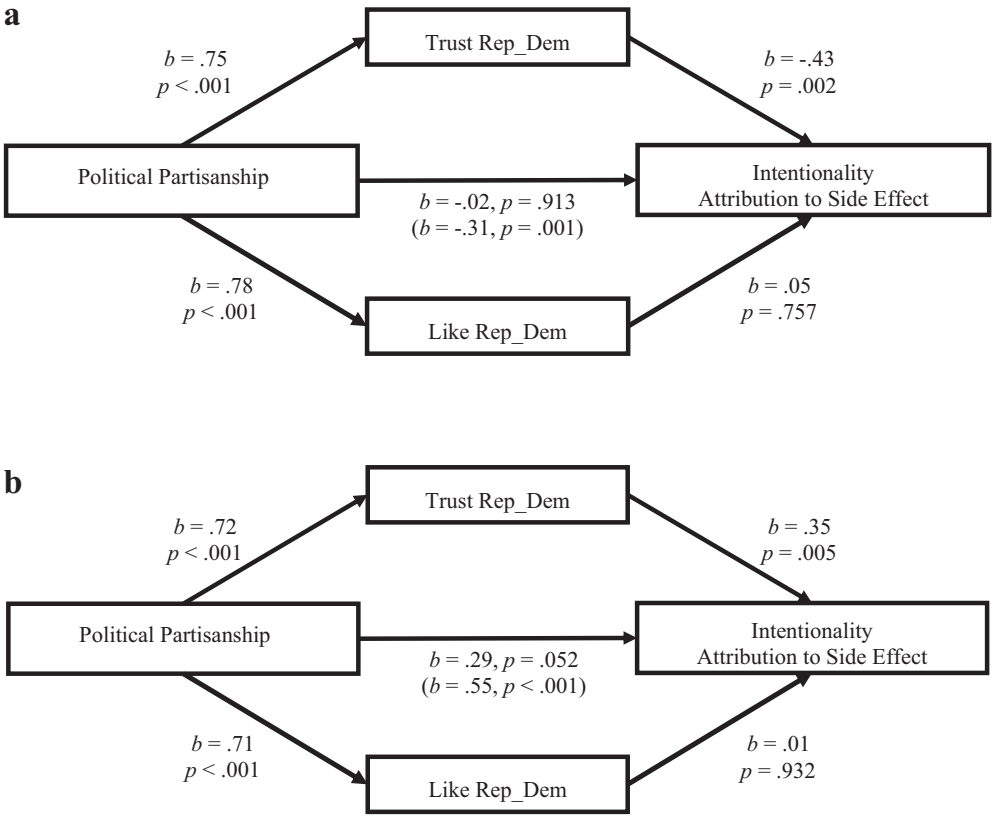

Fig. 5. a. In the Republican Policy condition, the trust gap in Republicans/Democrats mediated the effect of political partisanship on intentionality attribution to side effect (Study 4). b. In the Democratic Policy condition, the trust gap in Republicans/Democrats mediated the effect of political partisanship on intentionality attribution to side effect (Study 4).

without the manipulation) and rated how intentional they perceived the side effect to be on a 5-point scale (from 1-Not at all to 5-Extremely).

Perceived Intentionality of Primary Effect (PE). Participants also rated how intentional they perceived the primary effect of the policy to be on a 5-point scale (from 1-Not at all to 5-Extremely).

Policy Support. Participants rated the extent to which they would be willing to support the policy by answering the question: "Would you be willing to support Policy A?" (5-point scale from 1-Not at all to 5-Extremely).

Like/Dislike of Partisans and Contrapartisans. We measured how much people like/dislike members of both the Republican and the Democratic parties using the same six items from the fourth study. In this study: "Like Republicans" ($\alpha = 0.95$) and "Like Democrats" ($\alpha = 0.95$).

Trust in Partisans and Contrapartisans. We measured how much people trust/distrust the policy maker responsible for implementing Policy A using items analogous to the fourth study (in this study, $\alpha = 0.89$).

Table 6
Two-way interactions predicting main dependent variables (perceived intentionality of side effect, perceived intentionality of primary effect, and policy support) in the control conditions: Republican Policy and Democratic Policy with no manipulation. Note: standard errors in parentheses; 190 observations; *** p < .001, ** $p < .01$, * $p < .05$.

	Dependent Variables		
Independent Variables	Perc. Int. Side Effect	Perc. Int. Primary Effect	Policy Support
Political Partisanship	−0.30***	0.27**	0.39***
	(0.08)	(0.08)	(0.06)
Policy Party (Rep. vs Dem.)	−2.39***	1.39***	1.25***
	(0.35)	(0.38)	(0.28)
Political Partisanship *by* Policy Party	0.77***	−0.46***	−0.46***
	(0.11)	(0.12)	(0.09)
$F(3, 186)$	16.04	4.93	12.59
R-squared	0.20	0.07	0.17

8.4. Results

Our results revealed a replication of the partisan trade-off bias in the control conditions (Democratic versus Republican policy, no trust manipulation). The results of regression analysis showed that, in the control conditions, political partisanship interacted with policy party proposing the policy (Republican versus Democratic) to significantly predict attribution of intentionality to the side effects ($b = 0.77$, $SE = 0.11$, $p < .001$, 95% CI [0.54, 0.99], $F(3, 186) = 16.04$, $R^2 = 0.20$) and to the primary effects ($b = -0.46$, $SE = 0.12$, $p < .001$, 95% CI [−0.71, −0.21], $F(3, 186) = 4.93$, $R^2 = 0.07$) of the policy trade-offs. Results also showed a significant two-way interaction of political partisanship with policy party proposing the policy (Republican versus Democratic) predicting support for the proposed policy, our new measure introduced in this study ($b = -0.46$, $SE = 0.09$, $p < .001$, 95% CI [−0.64, −0.27], $F(3, 186) = 12.59$, $R^2 = 0.17$)—see Table 6 for results of two-way interactions in the two control conditions and Figs. 6, 7a, and 8a for graphical representations.

8.4.1. Trust manipulation: manipulation check

We tested whether our trust manipulation was successful in influencing how much people trust the contrapartisan policy maker and not how much people like/dislike contrapartisans in general. This is important as our claim is that the partisan trade-off bias results from a lack of trust in, and not dislike of, contrapartisans. The results of regression analyses showed that our trust manipulation successfully increased trust in the contrapartisan policy maker ($b = 0.80$, $SE = 0.09$, $p < .001$, 95% CI [0.62, 0.97], $F(1, 359) = 78.04$, $R^2 = 0.18$) while having no effect on how much people liked contrapartisans in general (effect on "Like Republicans": $b = 0.01$, $SE = 0.17$, $p = .933$, 95% CI [−0.33, 0.36], $F(3, 357) = 0.74$, $R^2 = 0.00$; effect on "Like Democrats": $b = -0.01$, $SE = 0.16$, $p = .926$, 95% CI [−0.34, 0.31], $F(3, 357) = 1.03$, $R^2 = 0.00$).

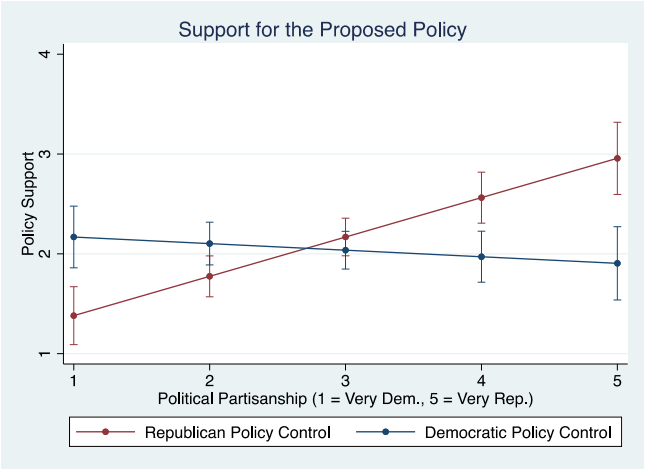

Fig. 6. Two-way interaction of political partisanship by policy proponent (Republican versus Democrat) on intentionality attribution to policy support in the control conditions. Error bars represent 95% confidence intervals. (Study 5).

8.4.2. Effect of trust manipulation on dependent variables

Results showed that the trust manipulation translated into a significant decrease in the partisan trade-off bias, both in terms of the attribution of intentionality to (1) the side effects (three-way interaction of political partisanship, Republican vs Democratic policy proponent, and trust manipulation on intentionality attribution to side effect: $b = 0.67$, $SE = 0.17$, $p < .001$, 95% CI [0.33, 1.00], $F(7, 353) = 12.59$, $R^2 = 0.17$) and to (2) the primary effects of policy trade-offs (three-way interaction of political partisanship, Republican vs Democratic policy proponent, and trust manipulation on intentionality attribution to the primary effect: $b = -0.31$, $SE = 0.17$, $p < .001$, 95% CI [−0.65, 0.03], $F(7, 353) = 5.94$, $R^2 = 0.10$). When it came to the new policy support measure, there was no three-way interaction, only a main effect of the trust manipulation on willingness to support the policy trade-off ($b = 0.74$, $SE = 0.10$, $p < .001$, 95% CI [0.54, 0.94], $F(1, 359) = 51.28$, $R^2 = 0.12$)—see Table 7 for two-way interactions in the trust manipulation conditions,

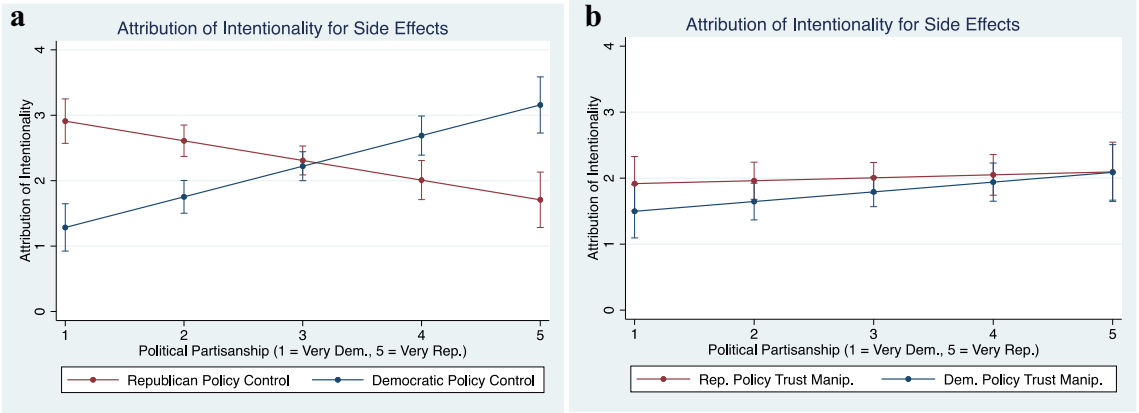

Fig. 7. a (left): Two-way interaction of political partisanship and policy proponent (Republican vs Democrat) on intentionality attribution to the side effect in the control conditions. Error bars represent 95% confidence intervals. (Study 5). b (right): Two-way interaction of political partisanship and policy proponent (Republican vs Democrat) on intentionality attribution to the side effect in the trust conditions. Error bars represent 95% confidence intervals. (Study 5).

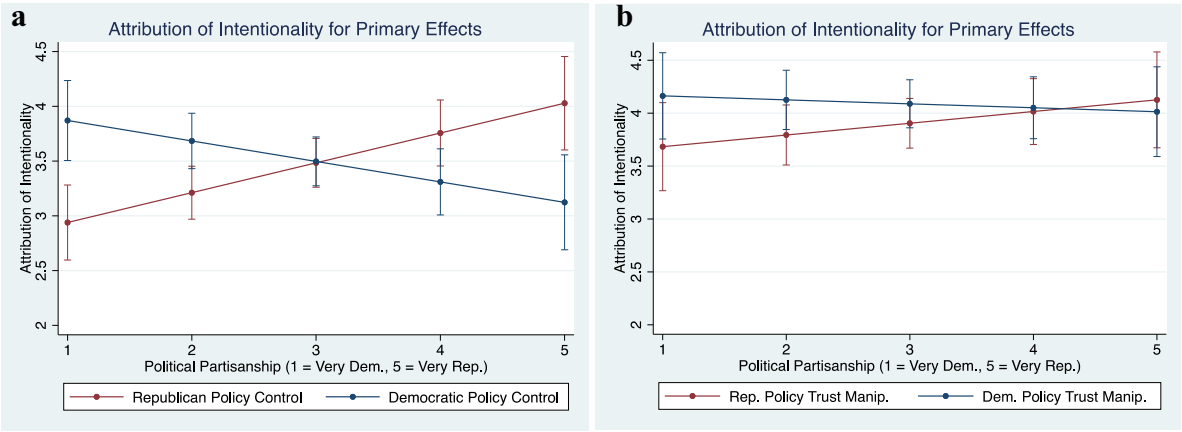

Fig. 8. a (left): Two-way interaction of political partisanship and policy proponent (Republican vs Democrat) on intentionality attribution to the primary effect in the control conditions. Error bars represent 95% confidence intervals. (Study 5). b (right): Two-way interaction of political partisanship and policy proponent (Republican vs Democrat) on intentionality attribution to the primary effect in the trust manipulation conditions. Error bars represent 95% confidence intervals. (Study 5).

Table 7
Two-way interactions predicting main dependent variables (perceived intentionality of side effect, perceived intentionality of primary effect, and policy support) in the trust manipulation conditions. Note: standard errors in parentheses; 171 observations; *** p < .001, ** *p* < .01, * *p* < .05.

	Dependent Variables		
Independent Variables	Perc. Int. Side Effect	Perc. Int. Primary Effect	Policy Support
Political Partisanship	0.04	0.11	0.27**
	(0.09)	(0.08)	(0.08)
Policy Party (Rep. vs Dem.)	−0.52	0.63	1.57***
	(0.40)	(0.36)	(0.36)
Political Partisanship *by* Policy Party	0.10	−0.15	−0.56***
	(0.12)	(0.11)	(0.11)
F(3, 167)	1.63	1.24	8.10
R-squared	0.03	0.02	0.13

Table 8
Three-way interactions predicting main dependent variables (perceived intentionality of side effect, perceived intentionality of primary effect, and policy support). Note: standard errors in parentheses; 361 observations; *** p < .001, ** p < .01, * *p* < .07.

	Dependent Variables		
Independent Variables	Perc. Int. Side Effect	Perc. Int. Primary Effect	Policy Support
Political Partisanship	0.47***	−0.19*	−0.07
	(0.08)	(0.08)	(0.07)
Policy Party (Rep. vs Dem.)	2.39***	−1.39***	−1.25***
	(0.35)	(0.35)	(0.30)
Trust Manipulation	0.53	0.14	1.38***
	(0.38)	(0.38)	(0.33)
Political Partisanship *by* Policy Party	−0.77***	0.46***	0.46***
	(0.12)	(0.12)	(0.10)
Political Partisanship *by* Trust Manip.	−0.32**	0.15	−0.22*
	(0.12)	(0.12)	(0.10)
Policy Party *by* Trust Manip.	−1.87***	0.76	−0.32
	(0.54)	(0.54)	(0.46)
Policy Party *by* Pol. Partis. *by* Trust Manip.	0.67***	−0.31*	0.10
	(0.17)	(0.17)	(0.15)
F(7, 353)	9.22	5.94	17.18
R-squared	0.15	0.10	0.25

Table 8 for overall three-way interactions, and Figs. 7b and 8b for graphical representations of the effect of the trust manipulation on the partisan trade-off bias.

8.5. Discussion

In this final study, we successfully manipulated trust in contrapartisan policy makers and significantly decreased the partisan trade-off bias. This study adds support to our proposed mechanism, that is, that trust in and not like/dislike of contrapartisans mediates the effect of political partisanship on attributions of intentionality to the primary effects and the side effects of policy trade-offs. Additionally, it brings to our attention the importance of emphasizing policy makers' commitment to providing voice to all groups in society as a path to building trust across the political spectrum.

9. General discussion

The increasing ideological divide between liberals and conservatives is corrupting public discourse in myriad ways (Achen & Bartels, 2016; Bennett & Iyengar, 2008; Iyengar et al., 2012; Sunstein, 2002). The polarization of political trust, in particular, has significantly contributed to decreased willingness to reach agreement on policies proposed by contrapartisans. This lack of willingness to reach agreements can be extremely costly. In 2013, for instance, Congress was unable to reach a budget compromise. This impasse activated, among other consequences, a series of mandatory cuts in Medicare and reductions in the defense budget. In that same year, Americans also experienced a government shutdown that cost the economy approximately $24 billion and losses of about 120,000 jobs (Binder, 2014).

As these political stalemates become increasingly common, scientists across social disciplines are working to better understand the psychology behind partisans' unwillingness to reach political compromise. Policy trade-offs—in which one has to pay an unavoidable cost in order to achieve a positive desired outcome—are rife territory for these political stalemates, but the psychology of policy trade-offs in the context of polarized trust remains to date underexplored. In our studies, we have documented a bias at the intersection of policy trade-offs and political polarization that poses yet another threat to the political debate. The partisan trade-off bias implies that both Democrats and Republicans view the side effects of policies proposed by members of the *opposite* party (contrapartisans) as intentional, but they do *not* attribute

intentionality to the very same side effects of policies proposed by members of their *own* party (copartisans). Worrisomely, both Democrats and Republicans also view the primary effects of policy trade-offs proposed by contrapartisans as less intentionally brought about relative to the primary effects of trade-offs proposed by copartisans. What is more, this bias is driven not by dislike of contrapartisans, but by the gap in trust in contrapartisans relative to trust in copartisans.

Our primary contribution is to the psychology of political polarization and rests on the uncovering of a novel bias that has the potential to covertly undermine political compromise. If our goal is to engage in fruitful political debate, people have to be capable of recognizing a trade-off for what it is: a policy that produces a *desired* and *intended* benefit along with a *foreseen* but *unintentional* side effect. The very possibility of compromise in the political arena depends on the ability of all parties to apply what is known as the principle of charity: to be charitable in a debate is to interpret the views of others under the most persuasive and reasonable light. In the context of political discourse as elsewhere, being able to accurately ascribe intentionality is a necessary condition for applying the principle of charity (Berg, 1987; Davidson, 1976; Grice, 1957; Grice, 1975). Attributing bad intentions to others fundamentally violates the principle of charity.[8] On a charitable interpretation of the motives and desires underlying policy trade-offs, one ought to conclude that policy makers neither intended nor intentionally brought about the side effect of adopting a certain policy. However, what we find is that people malign the motives of members of the opposite party in a way that they don't malign the motives of members of their own party, even in contexts where the policy trade-off is identical. Interestingly, the partisan trade-off bias is symmetrical across the political spectrum, lending support to the view that some partisan biases are symmetrical across liberals and conservatives (meta-analysis by Ditto et al., 2019).

A particularly worrisome aspect of the partisan trade-off bias is that partisans are displaying a tendency to malign not only the intentions underlying the side effects of policy trade-offs, but also the positive primary effects of these policies—of course, when the policies are proposed by contrapartisans. This tendency entails a reluctance to perceive contrapartisans as sincerely desiring to bring about the positive outcomes of their proposed policies. And this reluctance, in turn, is indicative of a political environment that is so distrusting across party lines that people appear to be seeing contrapartisans' actions as containing hidden motives under the guise of good policy outcomes. After all, it is one thing to believe that contrapartisans intended an unforeseen side effect, but it is yet another to not even believe that they desired the intended outcome. All in all, the partisan trade-off bias appears to map on to the new wave of conspiracy theories that are currently sweeping the United States, where people distrust what politicians say and are suspicious about the existence of secret reasons for their policies.

We have shown that simply knowing that the policy proposal comes from a copartisan or contrapartisan is sufficient to cause people to incur in the partisan trade-off bias. Yet we acknowledge that it may also be the case that people like or dislike the policy and, as a result, they manufacture arguments about contrapartisans' intentions to rationalize their favored policies—this is more likely for elites than for ordinary people. While still possible, we believe that this explanation is less likely in the case of the partisan trade-off bias in light of the studies with the content-free policy trade-offs (Studies 2, 4, and 5). In these studies, the only information participants have is about whether the proponent of the policy is Democrat or Republican. So, it is unlikely that participants are making inferences from the types of policies and how much they like or dislike them to how intentional they believe policy makers to be. In the complexity of life outside the laboratory, such effects could flow either from partisanship to perceived intentions to policy support or from partisanship to policy support to perceived intentions. While we cannot eliminate the possibility of causality in a different direction, we have relied on controlled laboratory experiments to provide evidence for the former causal chain.

We also contribute to the literature on the SE effect, showing that the attribution of intentionality to side effects can also occur when people evaluate the actions of groups, instead of individual agents (this effect has only been documented one time in the recent literature, see Michael & Szigeti, 2019). Moreover, in the context of public policies and political polarization, the SE effect is moderated by political partisanship. That is, the side effects of the policies proposed by one's own party are not evaluated as being intentionally brought about (to our knowledge, this has also not been documented in the literature).

Lastly, we contribute to the literature on trade-offs reasoning by uncovering a new obstacle in thinking through the logic of trade-offs; one that appears to be inherent to the context of public policies and fueled by political polarization. Tetlock (2000) has alluded to the fact that centrist ideologies should display superior performance in trade-off reasoning when compared with extremist ideologies. In our studies, we find evidence that is indicative of this view. People at both ends of the political spectrum (very Democratic or very Republican) are the most prone to incurring in the partisan trade-off bias. People in the middle, the ones that report being neither Democratic nor Republican, appear to be immune to this bias. Future research could investigate if independents are really unbiased or if they fail to display biased attributions of intentionality only because they are not partisans, that is, they do not have party in-groups and out-groups.

Our studies also have limitations that point us towards future directions. While we have used samples from two different participant pools (Amazon Mechanical Turk and Prolific), our samples are still exclusively from online panels and we would benefit from a replication with different samples. Another limitation of relying on online panels is that samples tend to be skewed towards liberals.[9] To this point, a limitation of our studies is that the absolute number of participants that have identified as "Republican" or "Very Republican" is particularly low in one of our studies (20 out of 186 participants have identified themselves as such in Study 3). In addition, future studies could explore how other types of interventions can potentially decrease the influence of the partisan trade-off bias on support for policies, thus contributing to a more balanced political debate.

All in all, our findings leave us with a warning: it may be fundamental for politicians to be careful about promoting voice in the design of their policies and to be clear about having included inputs from all stakeholders when expressing their proposals.

Open practices

We report all measures and manipulations in our studies and all materials, data, and code for all studies are made available in the project's OSF page: https://osf.io/qy23x/?view_only=1661466f2a024a609e81c3105bdc93e9

Appendix A. Policy proposals used in Study 1

Taxes

[8] There is a debate in the literature regarding how stringent the conditions of the principle of charity should be. For instance, Thagard and Nisbett (1983) deny that we should assume people are rational (when it comes to reasons and motivations). Yet, to our knowledge, no one suggests that we should not be as charitable as possible in interpreting intentions.

[9] Summary of Political Partisanship for all studies: Study 1: $N = 332$; 45% Dem. ($n = 149$); 31% Rep. ($n = 102$); Study 2: $N = 173$; 50% Dem. ($n = 85$); 34% Rep. ($n = 59$); Study 3: $N = 186$; 59% Dem. ($n = 110$); 10.75% Rep. ($n = 20$); Study 4: $N = 184$; 48% Dem. ($n = 88$); 34% Rep. ($n = 63$); Study 5: $N = 361$; 47% Dem. ($n = 169$); 35% Rep. ($n = 126$).

- Republicans say they want to reform the tax code. They support a tax break claiming it will help the rich have more money to create jobs, but knowing it also has the potential of cutting benefits for the poor. People have different views about the intentions of the law. What is your opinion?
- Democrats say they want to reform the tax code. They support a tax raise claiming it will increase benefits for the poor, but knowing it also has the potential of taking money that the rich could use to create jobs. People have different views about the intentions of the law. What is your opinion?

Environmental Regulation

- Republicans say they want to protect the future of American families. They support a plan to decrease environmental regulations claiming it would protect jobs in the fossil-fuels industry, but knowing it also has the potential to harm the environment. People have different views about the intentions of the law. What is your opinion?
- Democrats say they want to protect the future of American families. They support a plan to increase environmental regulations claiming it would protect the environment, but knowing it also has the potential to cut jobs in the fossil-fuels industry. People have different views about the intentions of the law. What is your opinion?

Gun Control

- Republicans say they want to increase public safety. They support a gun ownership law claiming it will increase individuals' ability to protect themselves, but knowing it also has the potential to increase the number of gun deaths. People have different views about the intentions of the law. What is your opinion?
- Democrats say they want to increase public safety. They support a gun ownership law claiming it will reduce the number of gun deaths, but knowing it also has the potential to restrict individuals' ability to defend themselves. People have different views about the intentions of the law. What is your opinion?

Voting ID Requirements

- Republicans say they want to protect the democratic process. They support a plan for stricter voter identification requirements claiming it would reduce the possibility of voter fraud, but knowing it also has the potential to make voting harder for disadvantaged groups. People have different views about the intentions of the law. What is your opinion?
- Democrats say they want to protect the democratic process. They support a plan for less strict voter identification requirements claiming it would make voting easier for disadvantaged groups, but knowing it also has the potential to increase voter fraud. People have different views about the intentions of the law. What is your opinion?

Appendix B. Measures of negative partisanship used in Studies 1 and 2

Participants rated Republicans and Democrats on a 5-point scale (from 1 = Not at all to 5 = Extremely) on the following traits:

1. Close-minded
2. Immoral
3. Lazy
4. Dishonest
5. Unintelligent
6. Open-minded
7. Moral
8. Hard-working
9. Honest
10. Intelligent

*Items 6–10 were reverse coded.

Feelings

Participants rated the extent to which Republicans and Democrats made them feel (5-point scale from 1 = Not at all to 5 = Extremely):

1. Frustrated
2. Afraid
3. Angry
4. Disgusted

Appendix C. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesp.2021.104231.

References

Achen, C., & Bartels, L. (2016). Democracy for realists: Holding up a mirror to the electorate. *Juncture, 22*(4), 269–275.

Bartels, M. (2002). Beyond the running tally: Partisan bias in political perceptions. *Political Behavior, 24*, 117–150.

Becker, G. (1981). *A treatise on the family*. Cambridge, MA: Harvard University Press.

Bennett, W. L., & Iyengar, S. (2008). A new era of minimal effects? The changing foundations of political communication. *Journal of Communication, 58*(4), 707–731.

Berg, J. (1987). Interpreting arguments. *Informal Logic, IX*(1), 13–21.

Binder, S. (2014). Polarized we govern?. Brooks Institution Report, Retrieved at https://www.brookings.edu/research/polarized-we-govern/.

Bolsen, T., Druckman, J. N., & Cook, F. L. (2014). The influence of partisan motivated reasoning on public opinion. *Political Behavior, 36*, 235–262.

Bullock, J. G. (2011). Elite influence on public opinion in an informed electorate. *The American Political Science Review, 105*, 496–515.

Carlin, R. E., & Love, G. J. (2018). Political competition, partisanship and interpersonal trust in electoral democracies. *British Journal of Political Science, 48*(1), 115.

Crawford, J. T., Kay, S. A., & Duke, K. E. (2015). Speaking out of both sides of their mouths: Biased political judgments within (and between) individuals. *Social Psychological and Personality Science, 6*, 422–430.

Davidson, D. (1976). Hume's cognitive theory of pride. *Journal of Philosophy, 73*, 744–757.

Dirks, K. T., & Ferrin, D. L. (2001). The role of trust in organizational settings. *Organization Science, 12*(4), 450–467.

Dirks, K. T., & Ferrin, D. L. (2002). Trust in leadership: Meta-analytic findings and implications for research and practice. *Journal of Applied Psychology, 87*(4), 611–628.

Ditto, P. H., Liu, B. S., Clark, C. J., Wojcik, S. P., Chen, E. E., Grady, R. H., … Zinger, J. F. (2019). *At least bias is bipartisan: A meta-analytic comparison of partisan bias in liberals and conservatives* (p. 19).

Druckman, J. N., Peterson, E., & Slothuus, R. (2013). How elite partisan polarization affects public opinion formation. *The American Political Science Review, 107*, 57–79.

Finkel, E. J., Bail, C. A., Cikara, M., Ditto, P. H., Iyengar, S., Klar, S., … Druckman, J. N. (2020). Political sectarianism in America. *Science, 370*(6516), 533–536.

Grice, H. P. (1957). Meaning. *Philosophical Review, 66*, 377–388.

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

Guglielmo, S., & Malle, B. F. (2010). Can unintended side effects be intentional? Resolving a controversy over intentionality and morality. *Personality and Social Psychology Bulletin, 36*(12), 1635–1647. https://doi.org/10.1177/0146167210386733

Hawkins, C. B., & Nosek, B. A. (2012). Motivated Independence? Implicit party identity predicts political judgments among self-proclaimed independents. *Personality and Social Psychology Bulletin, 38*(11), 1437–1452.

Hetherington, M. J., & Rudolph, T. J. (2018). Political trust and polarization. In E. M. Uslaner (Ed.), *The Oxford handbook of social and political trust* (pp. 579–597). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190274801.013.15.

Hetherington, Marc and Rudolph, Thomas J. (2014). The emergence of polarized trust. Available at SSRN: https://ssrn.com/abstract=2484755 or doi:https://doi.org/10.2139/ssrn.2484755.

Holland, P., Cooper, B., & Sheehan, C. (2017). Employee voice, supervisor support, and engagement: The mediating role of trust. *Human Resource Management, 56*(6), 915–929.

Itzchakov, G., & Kluger, A. (2017). Listening with understanding in negotiation and conflict resolution. *The Negotiator's Desk Reference*, 409–420.

Itzchakov, G., Weinstein, N., Legate, N., & Amar, M. (2020). Can high quality listening predict lower speakers' prejudiced attitudes? *Journal of Experimental Social Psychology, 91*, 104022.

Iyengar, S., & Krupenkin, M. (2018). The strengthening of partisan affect. *Political Psychology, 39*, 201–218.

Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideology. *Public Opinion Quarterly, 76*(3), 405–431. https://doi.org/10.1093/poq/nfs038

Iyengar, S., & Westwood, S. J. (2015). Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science, 59*(3), 690–707.
Kahan, D. M., Jenkins-Smith, H., & Braman, D. (2011). Cultural cognition of scientific consensus. *J Risk Res, 14*, 147–174.
Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.
Knobe, J. (2003). *Intentional action and side effects in ordinary language* (p. 6).
Knobe, J. (2006). The concept of intentional action: A case study in the uses of folk psychology. *Philosophical Studies, 130*(2), 203–231. https://doi.org/10.1007/s11098-004-4510-0
Lord, C., Ross, L., & Lepper, M. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37*, 2098–2109.
Lyu, S. C., & Ferrin, D. L. (2018). Determinants, consequences, and functions of interpersonal trust within organizations: What is the empirical evidence? In R. Searle, A. Nienaber, & S. Sitkin (Eds.), *The Routledge companion to trust* (pp. 65–104). London and New York: Routledge.
Machery, E. (2008). The folk concept of intentional action: Philosophical and experimental issues. *Mind & Language, 23*(2), 165–189.
Michael, J. A., & Szigeti, A. (2019). "The group knobe effect": Evidence that people intuitively attribute agency and responsibility to groups. *Philosophical Explorations, 22*(1), 44–61.
Nadelhoffer, T. (2010). Experimental philosophy of action. In J. Aguilar, A. Buckareff, & K. Frankish (Eds.), *New waves in the philosophy of action* (pp. 50–78). New York: Palgrave-MacMillan.
Pew Research Center. (2016). Partisanship and political animosity in 2016. Retrieved at: www.people-press.org/2016/06/22/partisanship-and-political-animosity-in-2016/.
Pew Research Center. (2019). Partisan antipathy: More intense, more personal. Retrieved at: https://www.pewresearch.org/politics/2019/10/10/partisan-antipathy-more-intense-more-personal/.
Schoemann, A. M., Boulton, A. J., & Short, S. D. (2017). Determining power and sample size for simple and complex mediation models. *Social Psychological and Personality Science, 8*(4), 379–386. https://doi.org/10.1177/1948550617715068
Sripada, C. S. (2010). The deep self model and asymmetries in folk judgments about intentional action. *Philosophical Studies, 151*(2), 159–176.
Sripada, C. S. (2012). Mental state attributions and the side-effect effect. *Journal of Experimental Social Psychology, 48*(1), 232–238.
Stine, M., Thompson, T., & Cusella, L. (1995). The impact of organizational structure and supervisory listening indicators on subordinate support, trust, intrinsic motivation, and performance. *International Journal of Listening, 9*(1), 84–105.
Sunstein, C. R. (2002). *Republic.com*. Princeton, New Jersey: Princeton University Press.
Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science, 50*, 755–769.
Tannenbaum, D., Fox, C. R., & Rogers, T. (2017). On the misplaced politics of behavioural policy interventions. *Nature Human Behaviour, 1*(7). https://doi.org/10.1038/s41562-017-0130
Tetlock, P. E. (1986). A value pluralism model of ideological reasoning. *Journal of Personality and Social Psychology, 50*(4), 819–827. https://doi.org/10.1037/0022-3514.50.4.819
Tetlock, P. E. (2000). Coping with trade-offs: Psychological constraints and political implications. In A. Lupia, M. D. McCubbins, & S. L. Popkin (Eds.), *Elements of reason: Cognition, choice, and the bounds of rationality* (pp. 239–263). New York: Cambridge University Press.
Tetlock, P. E., Hannum, K. A., & Micheletti, P. M. (1984). *Stability and change in the complexity of senatorial debate: Testing the cognitive versus rhetorical style hypotheses* (p. 12).
Thagard, P., & Nisbett, R. E. (1983). Rationality and charity. *Philosophy of Science, 50*(2), 250–267.
Van Boven, L., Ehret, P. J., & Sherman, D. K. (2018). Psychological barriers to bipartisan public support for climate policy. *Perspectives on Psychological Science, 13*(4), 492–507.
Vonasch, A. J., & Baumeister, R. F. (2017). Unjustified side effects were strongly intended: Taboo tradeoffs and the side-effect effect. *Journal of Experimental Social Psychology, 68*, 83–92.
Weisberg, H. F., & Rusk, J. G. (1970). Dimensions of candidate evaluation. *American Political Science Review, 64*(4), 1167–1185.