

Exposure to Arguments and Evidence Changes Partisan Attitudes Even in the Face of Countervailing Leader Cues

Ben M. Tappin ^{1,2,*}, Adam J. Berinsky ², David G. Rand ^{1,3}

¹Sloan School of Management, Massachusetts Institute of Technology

²Department of Political Science, Massachusetts Institute of Technology

³Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

*Corresponding author: benmtappin@googlemail.com

This manuscript is a preprint and has not yet been peer-reviewed.

Abstract

Patterns of public opinion recently observed in American politics tempt the conclusion that substantive arguments and evidence are less effective, or ineffective, at changing partisan minds when they overtly contradict cues from in-party leaders. This conclusion follows naturally from theories of partisan motivated reasoning. However, observations of public opinion do not provide the counterfactual outcomes required to draw this conclusion. Here we report a large-scale survey experiment in which we randomized exposure to the policy positions of Donald Trump and Joe Biden, as well as information that overtly contradicts their positions. Our design incorporates 24 policy issues and 48 information treatments. We find that the information does persuade partisans on average, and, critically, fully retains its persuasive force even when paired with countervailing cues from in-party leaders. This result holds across policy issues, demographic subgroups, and one- and two-sided cue environments, and is puzzling for partisan motivated reasoning theory.

Recent events in American politics constitute a stark reminder of the influence that party leaders can exert over their partisan supporters. Even months after the 2020 U.S. presidential election, large numbers of Republican voters continued to endorse former President Donald Trump's claim that the election was "stolen" from him by illegitimate means (Jackson and Duran 2021), despite widespread evidence and arguments to the contrary (Eggers, Garro, and Grimmer 2021; Kiely et al. 2020). Similarly, the relative skepticism observed among Republican voters over the health risks of Covid-19 during 2020 mirrored public communications from Donald Trump and other senior Republican and conservative sources (Brooks 2020; Deane, Parker, and Gramlich 2021; Summers 2020). Such skepticism appeared unwavering through 2020 and 2021, despite scientists and medical professionals repeatedly publicly attesting to the severity of the virus, and even as the number of U.S. infections, hospitalizations and deaths reached world-topping heights.

Together these examples are consistent with prevailing theories of political attitude formation, which suggest that substantive arguments and evidence are unlikely to change partisan minds when they overtly contradict in-party leaders. In particular, the influential theory of partisan motivated reasoning describes how people's motivation to form attitudes that are consistent with their party identification affects their receptiveness to new arguments and evidence (Bolsen, Druckman, and Cook 2014; Jerit and Barabas 2012; Kahan 2016; Taber and Lodge 2006). For example, the theory holds that this motivation reliably engages counterarguing that works to discredit and undermine politically uncongenial arguments and evidence (Kahan 2016; Taber, Cann, and Kucsova 2009; Taber and Lodge 2006). Notably, communications or "cues" from political parties are considered to be a key trigger of partisan motivated reasoning (Bolsen, Druckman, and Cook 2014; Druckman, Peterson, and Slothuus 2013; Petersen et al. 2013), especially when those cues come from party leaders (Barber and Pope 2019; Nicholson 2012).

Thus, the implication of partisan motivated reasoning theory is clear and straightforward: Substantive arguments and evidence will be less effective, or even entirely *ineffective*, at changing partisan minds when they overtly contradict the in-party leader.

This implication is deeply troubling from a normative perspective because it suggests that substantive arguments and evidence are undermined by partisan loyalties, thereby diminishing the accountability of party leaders. However, despite this normatively troubling implication, it is unknown whether substantive arguments and evidence do in fact lose their persuasive force among partisans when they contradict in-party leaders. Simply appealing to cases from world events — such as the public opinion dynamics over claims of election fraud and Covid-19, described above — is insufficient because of the lack of key counterfactual outcomes. For example, in these cases it is unknown to what extent the arguments and evidence would have influenced public opinion in the absence of countervailing cues from the in-party leader. But this counterfactual is required to test the hypothesis that substantive arguments and evidence become less effective at changing partisan minds when they contradict the party leader.

Experiments that randomize exposure to party leader cues and countervailing information offer a more rigorous method of testing the hypothesis. Unfortunately, such experiments are in short supply. While several previous studies do randomize substantive policy information alongside exposure to party cues (Bullock 2011; Ciuk and Yost 2016; Cohen 2003; Nicholson 2011; Peterson 2019), the design of these studies omits a control group in which people receive no information. Thus, the effect of exposure to countervailing information *per se* cannot be identified. Two studies that we know of do include the necessary control group, but their designs consider only a handful of policy issues and are beset by challenges to statistical power, thus limiting their ability to test the current hypothesis. Boudreau and MacKenzie (2014) randomized Californians to

receive generic party cues and/or countervailing policy information regarding several local ballot initiatives. However, the countervailing information treatment in their study had little persuasive power (no significant difference from control) absent the party cue, preventing a strong test of whether the information loses its persuasive force when the party cue is added.

Agadjanian (2020) also included the necessary control group, randomizing Democrats and Republicans to receive countervailing information and/or party leader cues on a single policy issue each. Among Republicans, the effect of the information absent the cue did not significantly differ from the control group — as above — again preventing a strong test of whether exposure to party cues undermines the effect of countervailing information. Among Democrats, the effect of the information *was* reliably detected absent the party cue, and there was no significant interaction between exposure to the information and the party cue. The null interaction implies that the information retained its persuasive force even in the face of the party cue. However, this null effect was imprecisely estimated, such that the data were consistent with the information losing more than half its power in the presence of the party cue. More importantly, the null result concerned just a single policy issue and corresponding information treatment, severely constraining its generalizability. A rapidly growing body of evidence suggests that the influence of party cues and substantive arguments are highly variable across policy issues (Blumenau and Lauderdale 2020; Clifford, Leeper, and Rainey 2020; Tappin 2020; Tappin and McKay 2021). Thus, despite observing a null interaction effect on one issue, party leader cues could still undermine the persuasive force of countervailing information across a majority of policy issues (Yarkoni 2020).

Our Experiment

Here we report the results of a large-scale randomized survey experiment designed to provide a high-powered and widely generalizable test of the hypothesis that substantive arguments and evidence lose their persuasive force among partisans when they contradict the in-party leader. In our experiment, we randomly expose Republicans and Democrats to cues from their party leaders — former President Donald Trump or current President Joe Biden, respectively — as well as arguments and evidence that overtly contradict the leader’s position across a large sample of 24 contemporary U.S. policy issues (discussed below). Altogether, our design incorporates 48 unique arguments and evidence (information) treatments, and we recruit a total of 5,071 American adults corresponding to 25,181 observations. The experiment design is as follows.

Each respondent is asked whether they agree or disagree with five policies, drawn randomly from a set of 24. Answers are given on a seven-point Likert scale running from strongly disagree (1) to strongly agree (7). The five policies are presented one per survey page and in a random order. There are two treatment factors: (1) the party leader cue {no cue, cue} and (2) the countervailing information that opposes the position taken by the in-party leader {no information, information}. On each policy question, respondents are randomly assigned with equal probability to receive one of the four conditions that results from crossing the two treatment factors. Random assignment is independent across policy questions.

For additional generalizability, we also randomize the specific nature of the party leader cue. Prior to seeing any policy questions, respondents are randomized with equal probability to one of two “cue type” conditions, determining whether they see a “one-sided” party leader cue or a “two-sided” party leader cue on policy questions where they are assigned to receive a party leader cue. Respondents assigned to the one-sided cue type condition see cues from their in-party leader

only, whereas respondents assigned to the two-sided cue type condition see cues from their in-party *and* out-party leader. Defining who is the in-party and out-party leader for each respondent is done with respect to their party identification obtained at the beginning of the survey, Democrat or Republican. For Democrats, the in-party leader is Joe Biden and the out-party leader is Donald Trump; for Republicans it is the reverse. This design feature allows us to examine whether our results are robust across one- and two-sided party leader cue environments. See Online Appendix S1 for further details regarding survey design.

The policy issues and corresponding cues (positions) of Donald Trump and Joe Biden were sourced from the website <https://www.isidewith.com>, an online encyclopedia that documents the positions of U.S. political figures on a range of contemporary policy issues. The 24 policy issues chosen for the experiment were diverse, corresponding to the broad policy areas of immigration, the economy, healthcare, the military, foreign policy, and the criminal justice system, among others. The positions of Trump and Biden were in opposition on all of the policy issues in the set, and the policies covered a mix of “hot button” issues, such as amnesty for illegal immigrants, as well as those perhaps less in the public eye, such as increasing capital gains tax. See Online Appendix S2 for further details regarding the policies and party leader cue treatments.

For each policy issue, we developed two information treatments that were each approximately 150 words in length; one that contradicted Trump’s position, and one that contradicted Biden’s position on the policy. The information treatments did not mention the positions of any political figures or party. Instead, they entailed substantive arguments for or against the policy, and often cited real-world evidence, such as statistics, in support of their argument. For policies where respondents were assigned an information treatment, they always received the treatment that contradicted the cue from their in-party leader. For example, if Trump

supported the policy, Republicans would receive the information treatment against the policy; if he was opposed, they would receive the information treatment in favor. (And the same for Democrats with respect to Biden’s position.) Online Appendix S2.1 reports the information treatments verbatim and provides further details. We contracted with Lucid to recruit 5000 U.S. adults quota matched to the national distributions of age, gender, education and region. The survey was fielded September 2 – 13, 2021. See Online Appendix S3.3 for sample demographics.

Results

Our pre-registered analysis¹ restricts to respondents who identified as Republicans or Democrats (including Independents who “lean” to one of the parties) and reported voting for Donald Trump or Joe Biden, respectively, in the 2020 U.S. presidential election ($n = 4,531$; 22,499 observations). For respondents who did not vote in 2020, we use their stated preference for Trump or Biden. This restriction theoretically maximizes the influence of the party leader cue by excluding pure Independents and a small minority of Republicans (Democrats) who preferred Biden (Trump) in 2020. Furthermore, we re-code the outcome variable — the extent to which respondents agreed or disagreed with the policy — such that higher numbers indicate greater agreement with the in-party leader cue. This allows us to meaningfully aggregate across policy issues and partisans.

To analyze the data, we fit a Bayesian multilevel linear regression model because the data are clustered by policy issue and respondent (McElreath 2020). The model specification includes a parameter for each of our treatment factors, as well as their interaction, and we allow these parameters to vary across both policy issues and respondents. The prior distributions on all model

¹ Pre-analysis plan: <https://osf.io/9gnaj>.

parameters are vague and weakly-informative, allowing the data to speak for themselves. See Online Appendix S4 for details regarding model specification and diagnostics.

Figure 1 shows the results from the model for the three key parameters: the average treatment effect (ATE) of the party leader cue (in the absence of countervailing information); the ATE of the countervailing information (in the absence of the party leader cue); and their interaction, which describes the extent to which the ATE of the countervailing information *changes* in the presence of the party leader cue. The point estimates and uncertainty intervals are the median and 95% highest posterior density intervals (HPDIs) of the posterior distribution for each parameter. The 95% HPDI contains the true value of the parameter with 95% probability, given the data and model. Figure 1 also shows the raw means in each condition (inset).

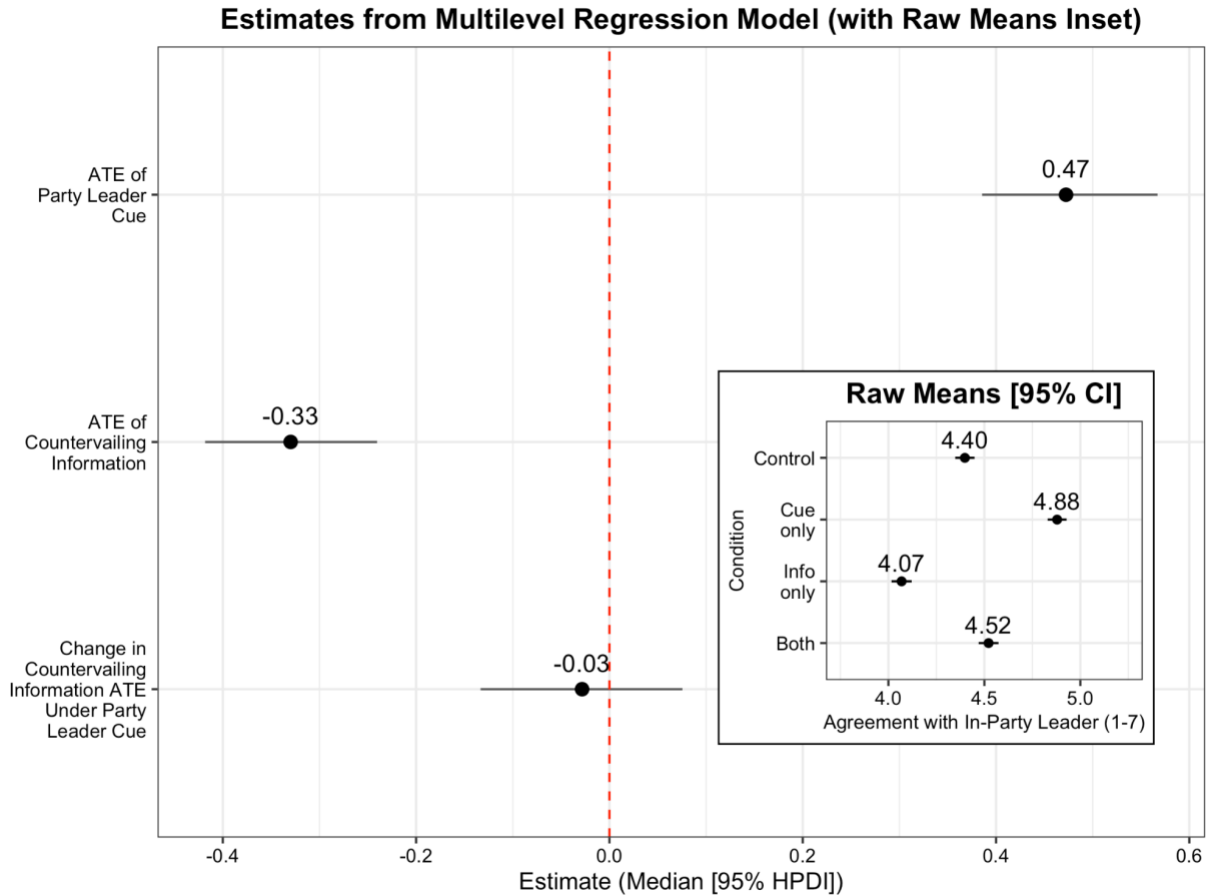


Figure 1. Key estimates from the multilevel model with raw means inset.

The ATE of the party leader cue is 0.47 [0.39, 0.57] of a Likert scale point; a modest but precisely estimated positive effect. As expected, partisans change their attitudes in the direction of cues from the in-party leader when these are learned. The ATE of the countervailing information is smaller in magnitude (and opposite in direction) at -0.33 [-0.42, -0.24]. On average, partisans update their attitudes toward the information when cues from party leaders are absent. But what about when cues from party leaders are present? The point estimate of the interaction effect is -0.03 [-0.13, 0.08], showing that the ATE of the information barely changes when party leader cues are present.

In other words, it is plausible the information retains all of its persuasive force despite the countervailing cue from the in-party leader. Furthermore, given that the upper bound of the 95% HPDI on the interaction effect is 0.08, we can be > 95% confident that the power of the information does not decrease by more than one-quarter under the party leader cue (because $0.08 / 0.33 \cong 0.24$). In summary, these results demonstrate that substantive arguments and evidence retain their persuasive force despite the presence of a countervailing cue from the in-party leader. (In Online Appendix S4.3 we show that this key result holds across the order of policy questions and worst-case imputation for post-treatment missing outcomes.)

There is a striking lack of heterogeneity in this key result across policy issues, demographic subgroups and cue environment. Figure 2A shows the ATE of the countervailing information for each policy issue in our sample when the party leader cue is absent versus present. The estimates come from two separate multilevel models fitted to observations where the party leader cue was absent versus present (see Online Appendix S5 for details). Summarizing the figure, the pairs of estimates across policy issues are similar, indicating little policy-level heterogeneity in the (lack of) impact of the party leader cue on the persuasive force of the countervailing information.

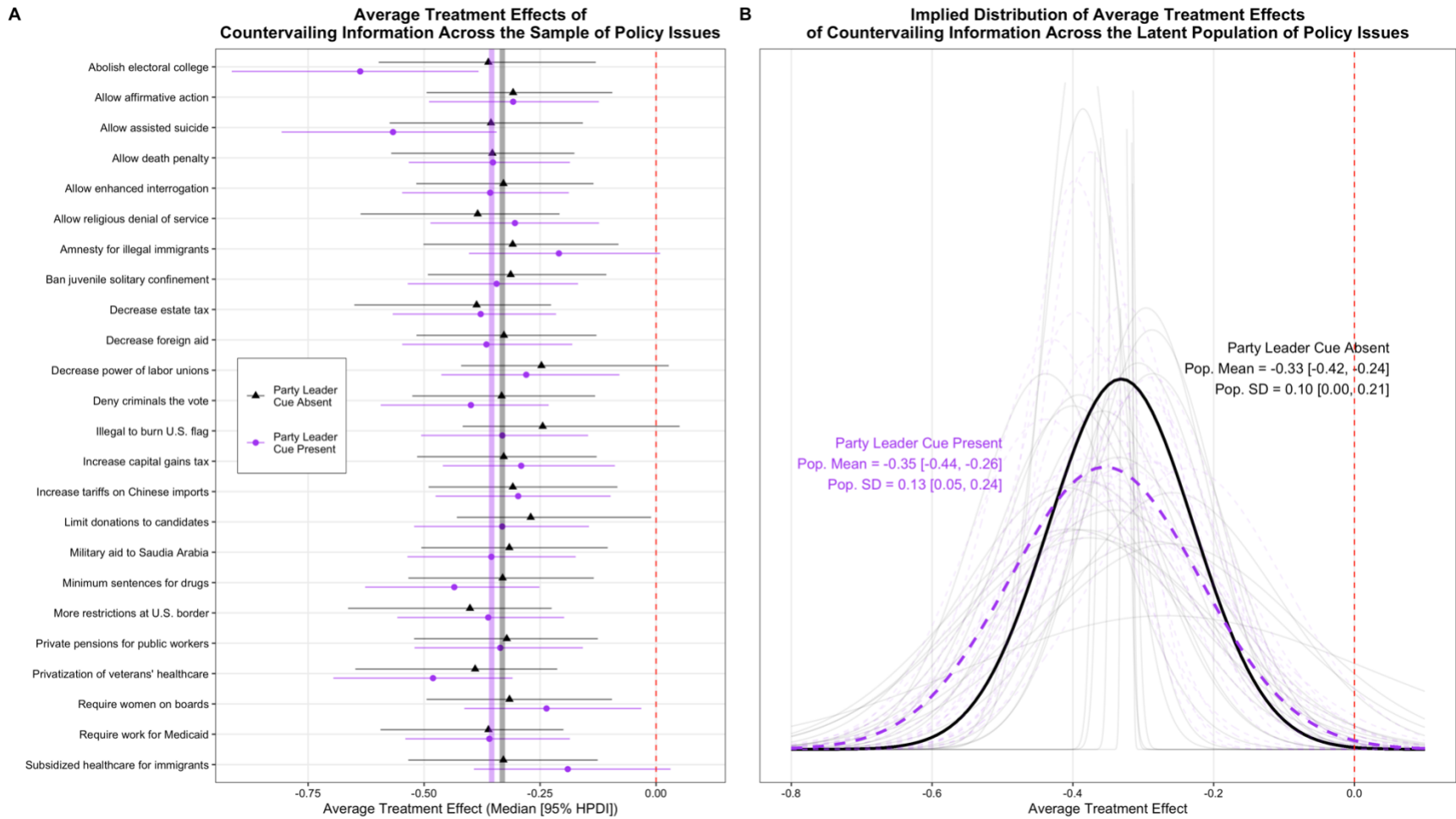


Figure 2. Average treatment effects of countervailing information across policy issues. (A) The estimates are mildly regularized, improving their out-of-sample accuracy on average (McElreath 2020). The faded vertical lines show the ATE averaged across policy issues. (B) The faded distributions in the background are other possibilities drawn from the posterior of the model, visually communicating the uncertainty in the Mean and SD.

However, while a key strength of our design is the large sample of policy issues, clearly we do not exhaust the space of all possible issues. Thus, Figure 2B shows the implied distributions of average treatment effects across the *unobserved population* of policy issues from which our sample is drawn. One distribution represents the “world” in which the party leader cue is absent (solid black), and the other in which it is present (dashed purple). The distributions are estimated by the models, given the data and assuming normality. The parameters of the distributions (Mean, SD) are annotated with 95% HPDIs appended. The key implication of the implied distributions is that, for most of the unobserved set of all possible policy issues, the model expects that partisans will update towards arguments and evidence on average even though they contradict the cue from the in-party leader. In other words, most of the probability mass of the dashed purple distribution lies below zero. Note, however, there is some probability mass in the tail of the distribution that is very close to zero. Thus, the model does acknowledge that, for a small minority of policy issues, partisans may not update much at all towards arguments and evidence on average when they contradict a cue from the in-party leader.

Figure 3 shows conditional average effects and their corresponding interaction estimates for subgroups defined by demographics and the cue environment (one-sided or two-sided cue). See Online Appendix S6 for the model specifications and diagnostics. There is limited evidence of subgroup heterogeneity, even where theory suggests we should observe such heterogeneity. For example, even among strong partisans, substantive arguments and evidence retain their persuasive force despite the countervailing cue from the in-party leader. Furthermore, the same is true for the nature of the cue environment: even in two-sided cue environments — where the arguments and evidence not only contradict the in-party leader but also *overtly support the out-party leader* — the persuasive force of the arguments and evidence remains undiminished.

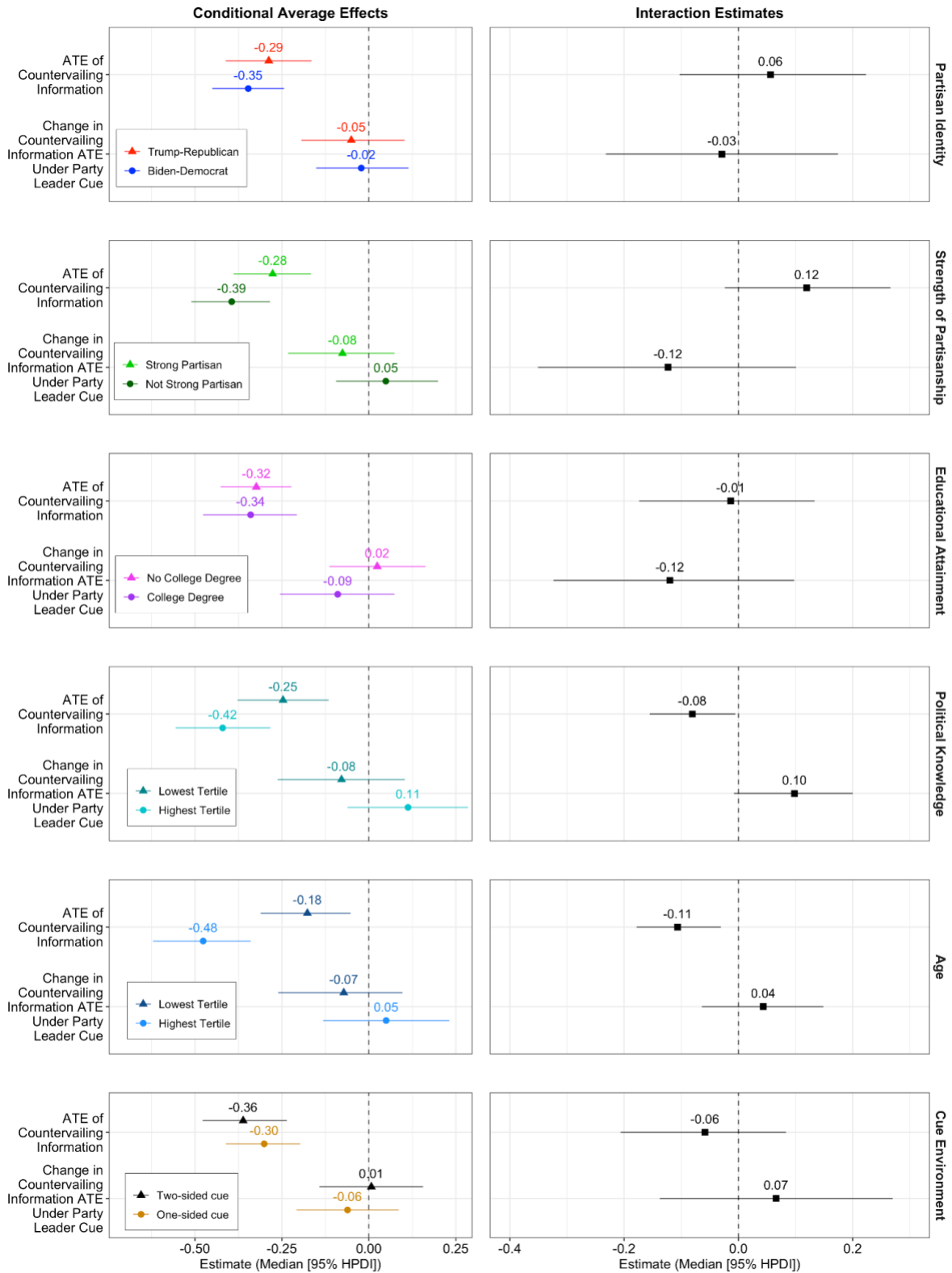


Figure 3. Subgroup conditional average effects and interaction estimates. Note *Political Knowledge* and *Age* are standardized in the interaction models, thus the estimates do not align perfectly with the conditional average effects.

Implications

Patterns of public opinion recently observed in American politics appear to suggest that substantive arguments and evidence are less effective, or entirely ineffective, at changing partisan minds when they contradict cues from in-party leaders. Despite the lack of relevant counterfactuals, observers could be forgiven for drawing this conclusion because it follows directly from the influential theory of partisan motivated reasoning. However, our results strongly dispute this conclusion. Across a large sample of 24 policy issues and 48 information treatments, we find that substantive arguments and evidence typically retain their persuasive force despite the presence of countervailing cues from in-party leaders. Moreover, this result generalizes across policy issues and demographic subgroups — notably, even among strong partisans — as well as in the absence or presence of antagonistic cues from out-party leaders.

Our results are puzzling for partisan motivated reasoning theory. According to that theory, exposure to cues from party leaders reliably triggers a motivation to form attitudes that are consistent with party identification (Barber and Pope 2019; Bolsen, Druckman, and Cook 2014; Druckman, Peterson, and Slothuus 2013; Petersen et al. 2013), in turn prompting counterarguing to undermine and discredit uncongenial information (Kahan 2016; Taber, Cann, and Kucsova 2009; Taber and Lodge 2006). Our results are not consistent with this theory. How might we reconcile this discrepancy? It could be the case that exposure to party leader cues does not trigger partisan motivated reasoning. Alternatively, perhaps the extent to which motivated partisans are able to undermine the persuasive effects of information simply by counterarguing is far more limited than currently assumed by the theory. In our view, the second explanation is most plausible. In fact, there are indirect clues as to this explanation in the real world. For example, typically parties (and their supporters in the media) do not simply announce their positions to the electorate,

but rather spend a great deal of time and energy providing justifications for those positions, as well as arguments against alternative positions. Alongside persuading people in their own right, these justifications and arguments may function to help motivated partisans better undermine and resist the persuasive attempts of their opponents. Future work should investigate this possibility.

Our results appear at odds with dynamics of public opinion observed in the real world, where persuasion across the aisle can seem slow or impossible. However, the design of our survey experiment differs from the dynamics of the real world in a variety of ways. One of the key differences is that partisans in our design are forcibly exposed to the information. Partisans in the real world may not be exposed to the information, or be more often exposed to information supportive of the in-party leader (rather than opposed), or be exposed repeatedly to cues from in-party leaders. Despite these differences, our results suggest that persuasion *is* nevertheless possible, given exposure. Thus, when faced with normatively troubling cues from party leaders, like unsubstantiated claims of election fraud, or health misinformation, our results imply that counter-communication strategies would be improved by making it harder for partisans to avoid countervailing information — thereby forcing exposure — alongside other strategies like sanctioning the media and party elites for disseminating and justifying such cues (Nyhan 2021).

In summary, the results reported here call for a revised understanding of key assumptions of partisan motivated reasoning theory, and offer an optimistic picture of persuasion in practice: arguments and evidence can persuade partisans, despite countervailing in-party leader cues.

Acknowledgements

We are grateful to Paul Irvine for research assistance, and for funding from the MIT Political Experiments Research Lab, the John Templeton Foundation (Grant #61061), and the Alfred P. Sloan Foundation.

Online Appendix and Materials

<https://osf.io/v3s72/>.

References

- Agadjanian, Alexander. 2020. "When Do Partisans Stop Following the Leader?" *Political Communication*: 1–19.
- Barber, Michael, and Jeremy C. Pope. 2019. "Does Party Trump Ideology? Disentangling Party and Ideology in America." *American Political Science Review* 113(1): 38–54.
- Blumenau, Jack, and Benjamin E Lauderdale. 2020. "The Variable Persuasiveness of Political Rhetoric." *American Journal of Political Science*.
<http://benjaminlauferdale.net/files/papers/BlumenauLauderdalePersuasion.pdf>.
- Bolsen, Toby, James N. Druckman, and Fay Lomax Cook. 2014. "The Influence of Partisan Motivated Reasoning on Public Opinion." *Political Behavior* 36(2): 235–62.
- Boudreau, Cheryl, and Scott A. MacKenzie. 2014. "Informing the Electorate? How Party Cues and Policy Information Affect Public Opinion about Initiatives." *American Journal of Political Science* 58(1): 48–62.
- Brooks, Brad. 2020. "Like the Flu? Trump's Coronavirus Messaging Confuses Public, Pandemic Researchers Say." *Reuters*. <https://www.reuters.com/article/us-health-coronavirus-mixed-messages-idUSKBN2102GY> (August 6, 2021).
- Bullock, John G. 2011. "Elite Influence on Public Opinion in an Informed Electorate." *American Political Science Review* 105(3): 496–515.
- Ciuk, David J., and Berwood A. Yost. 2016. "The Effects of Issue Salience, Elite Influence, and Policy Content on Public Opinion." *Political Communication* 33(2): 328–45.
- Clifford, Scott, Thomas J Leeper, and Carlisle Rainey. 2020. "Increasing the Generalizability of Survey Experiments Using Randomized Topics: An Application to Party Cues."
- Cohen, Geoffrey L. 2003. "Party Over Policy: The Dominating Impact of Group Influence on

- Political Beliefs.” *Journal of Personality and Social Psychology* 85(5): 808–22.
- Deane, Claudia, Kim Parker, and John Gramlich. 2021. “A Year of U.S. Public Opinion on the Coronavirus Pandemic.” *Pew Research Center*.
<https://www.pewresearch.org/2021/03/05/a-year-of-u-s-public-opinion-on-the-coronavirus-pandemic/> (August 6, 2021).
- Druckman, James N., Erik Peterson, and Rune Slothuus. 2013. “How Elite Partisan Polarization Affects Public Opinion Formation.” *American Political Science Review* 107(1): 57–79.
- Eggers, Andrew C, Haritz Garro, and Justin Grimmer. 2021. “No Evidence for Voter Fraud: A Guide To Statistical Claims About the 2020 Election.” *Working Paper*: 1–51.
- Jackson, Chris, and Jocelyn Duran. 2021. “Majority of Republicans Still Believe the 2020 Election Was Stolen from Donald Trump.” *Ipsos*. <https://www.ipsos.com/en-us/news-polls/majority-republicans-still-believe-2020-election-was-stolen-donald-trump> (July 6, 2021).
- Jerit, Jennifer, and Jason Barabas. 2012. “Partisan Perceptual Bias and the Information Environment.” *The Journal of Politics* 74(3): 672–84.
- Kahan, Dan M. 2016. “The Politically Motivated Reasoning Paradigm, Part 1: What Politically Motivated Reasoning Is and How to Measure It.” In *Emerging Trends in the Social and Behavioral Sciences*, , 1–16.
<https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118900772.etrds0417> (June 4, 2019).
- Kiely, Eugene, Lori Robertson, Rem Rieder, and D’Angelo Gore. 2020. “The President’s Trumped-Up Claims of Voter Fraud.” *FactCheck.org*.
<https://www.factcheck.org/2020/07/the-presidents-trumped-up-claims-of-voter-fraud/>

- (July 6, 2021).
- McElreath, Richard. 2020. *Statistical Rethinking: A Bayesian Course with Examples in R and STAN*. CRC Press.
- Nicholson, Stephen P. 2011. “Dominating Cues and the Limits of Elite Influence.” *The Journal of Politics* 73(4): 1165–77.
- . 2012. “Polarizing Cues.” *American Journal of Political Science* 56(1): 52–66.
- Nyhan, Brendan. 2021. “Why the Backfire Effect Does Not Explain the Durability of Political Misperceptions.” *Proceedings of the National Academy of Sciences* 118(15): e1912440117.
- Petersen, Michael Bang, Martin Skov, Søren Serritzlew, and Thomas Ramsøy. 2013. “Motivated Reasoning and Political Parties: Evidence for Increased Processing in the Face of Party Cues.” *Political Behavior* 35(4): 831–54.
- Peterson, Erik. 2019. “The Scope of Partisan Influence on Policy Opinion.” *Political Psychology* 40(2): 335–53.
- Summers, Juana. 2020. “Timeline: How Trump Has Downplayed The Coronavirus Pandemic.” *NPR*. <https://www.npr.org/sections/latest-updates-trump-covid-19-results/2020/10/02/919432383/how-trump-has-downplayed-the-coronavirus-pandemic> (August 6, 2021).
- Taber, Charles S., Damon Cann, and Simona Kucsova. 2009. “The Motivated Processing of Political Arguments.” *Political Behavior* 31(2): 137–55.
- Taber, Charles S., and Milton Lodge. 2006. “Motivated Skepticism in the Evaluation of Political Beliefs.” *American Journal of Political Science* 50(3): 755–69.
- Tappin, Ben M. 2020. *Estimating the Between-Issue Variation in Party Elite Cue Effects*.

PsyArXiv. preprint. <https://osf.io/p48zb> (October 8, 2020).

Tappin, Ben M., and Ryan McKay. 2021. "Estimating the Causal Effects of Cognitive Effort and Policy Information on Party Cue Influence." <https://psyarxiv.com/tdk3y/> (February 14, 2021).

Yarkoni, Tal. 2020. "The Generalizability Crisis." *Behavioral and Brain Sciences*.
<https://osf.io/jqw35> (December 10, 2019).