nature human behaviour

Article

https://doi.org/10.1038/s41562-023-01551-7

Partisans' receptivity to persuasive messaging is undiminished by countervailing party leader cues

Received: 13 July 2022

Accepted: 6 February 2023

Published online: 02 March 2023

Check for updates

Ben M. Tappin [1,2] ✉, Adam J. Berinsky [1] & David G. Rand [2,3]

It is widely assumed that party identification and loyalty can distort partisans' information processing, diminishing their receptivity to counter-partisan arguments and evidence. Here we empirically evaluate this assumption. We test whether American partisans' receptivity to arguments and evidence is diminished by countervailing cues from in-party leaders (Donald Trump or Joe Biden), using a survey experiment with 24 contemporary policy issues and 48 persuasive messages containing arguments and evidence (N = 4,531; 22,499 observations). We find that, while in-party leader cues influenced partisans' attitudes, often more strongly than the persuasive messages, there was no evidence that the cues meaningfully diminished partisans' receptivity to the messages—despite them directly contradicting the messages. Rather, persuasive messages and countervailing leader cues were integrated as independent pieces of information. These results generalized across policy issues, demographic subgroups and cue environments, and challenge existing assumptions about the extent to which party identification and loyalty distort partisans' information processing.

A central question in the study of political psychology is to what extent, and under what conditions, exposure to persuasive arguments and evidence ('persuasive messages') causes people to change their political attitudes[1]. In this paper we test whether American partisans' receptivity to such persuasive messaging is diminished by countervailing cues from favoured party leaders Donald Trump and Joe Biden. While cues from party leaders and other elites are ubiquitous in US politics, and their effects on Americans' opinions are well documented, there is limited evidence as to whether persuasive messages that explicitly cut against these cues retain (versus lose) their persuasive force. However, robustly answering this question is important for various reasons.

First, recent events in US politics call for an answer to this question. Even months after the 2020 US 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[2], despite widespread arguments and evidence to the contrary[3,4]. Similarly, the relative scepticism observed among Republican voters over the health risks of coronavirus disease 2019 during 2020 mirrored public communications from Donald Trump and other Republican-aligned elites[5–7]. Such scepticism appeared unwavering through 2020 and 2021, despite scientists and medical professionals attesting to the severity of the virus, and even as the number of US infections, hospitalizations and deaths reached world-topping heights. These events are of acute practical importance, and suggest that arguments and evidence fall on deaf partisan ears when pitted against countervailing cues from party leaders. However, they lack the required counterfactual outcomes to warrant this inference. For example, perhaps public opinion would have been further skewed in a party-consistent direction were it not for the arguments and evidence in public domain.

[1]Department of Political Science, Massachusetts Institute of Technology, Cambridge, MA, USA. [2]Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA. ✉e-mail: benmtappin@gmail.com

Second, consistent with these recent events in American politics and trends in public opinion, major theoretical accounts of how party leader cues influence partisans' psychology predict that partisans' receptivity to persuasive messaging can indeed be diminished by countervailing cues from party leaders. Specifically, exposure to such cues is theorized to activate people's party identification and loyalty, producing an emotional reaction and (partisan) motivation to adopt the party position[8]. This process can be expected to diminish partisans' receptivity to persuasive messaging, insofar as partisans either blindly conform to the leader's position (thus ignoring the messaging) or strive to actively defend the leader's position (thus refuting the messaging).

Third, whether or not partisans' receptivity to persuasive messaging is diminished by countervailing cues from party leaders speaks to longstanding debates over the extent to which party leaders lead (versus follow) public opinion[9,10]. Specifically, if countervailing cues from party leaders diminish partisans' receptivity to arguments and evidence, leaders plausibly possess even greater power to direct public opinion than currently thought—thus further limiting the extent to which they may be constrained by public opinion. On the other hand, if exposure to persuasive messaging largely retains its causal effect despite countervailing party leader cues, then it is possible (in principle) for arguments and evidence to counteract leaders' influence on public opinion and facilitate constraint.

Fourth, a large body of evidence from randomized survey experiments indicates that partisans on the left and right update their political attitudes and beliefs in broadly similar ways when exposed to the same arguments and evidence[11–14]. Yet, there remains substantial political polarization in the standing attitudes and beliefs of the American public. What explains this discrepancy? One explanation is that, in the real world, partisans are exposed to different arguments and evidence—resulting in different attitudes and beliefs. However, another explanation is that partisans are exposed to broadly similar arguments and evidence, but the persuasive causal effect of these messages is selectively diminished by exposure to countervailing cues from favoured party leaders. The aforementioned experiments cannot adjudicate between these possibilities.

In this Article, we bring empirical evidence to bear on each of these points. We test whether the causal effect of persuasive messaging on American partisans' attitudes is diminished by countervailing cues from party leaders Donald Trump and Joe Biden, using a large-scale pre-registered survey experiment with $N = 4{,}531$ American partisans ($N = 22{,}499$ observations), 24 contemporary US policy issues and 48 unique persuasive message treatments containing arguments and evidence.

Our results show the following. As in past work, cues from favoured party leaders reliably influenced partisans' attitudes and, in our case, typically to a greater extent than the persuasive messages. Critically, however, we found no evidence that the cues meaningfully diminished partisans' receptivity to the messages—despite standing in direct contradiction to the messages. Moreover, this result generalized broadly across policy issues, demographic subgroups and cue environments. When Trump-voting Republicans or Biden-voting Democrats were exposed to persuasive messaging about a policy issue, they responded by (1) updating their attitudes towards the message on average, and (2) updating their attitudes by a similar amount even when confronted with the fact that Trump or Biden's position, respectively, was opposed to the message. They responded this way largely irrespective of the policy issue in question, and largely irrespective of their age, gender, education, knowledge of politics or strength of partisanship. Finally, they responded this way even in polarized, or 'two-sided' cue environments; that is, even when they knew that not only was the position of their in-party leader opposed to the message, but in addition that the position of the out-party leader was aligned with the message.

We draw two main conclusions from these results. First, party loyalty- and partisan-motivated conformity exert a more limited effect on information processing than currently understood. It is widely held that party loyalty and the partisan motivation to conform are 'activated' by party cues and can thereby exert a powerful influence over people's information processing—potentially distorting their perception, reasoning and thus receptivity to other types of (especially counter-partisan) information.

For example, the authors of *The American Voter* famously wrote that party identity raises a partisan 'perceptual screen' over information processing[15], and many scholars corroborate this assessment, concluding that party loyalties 'have pervasive effects on perceptions'[16] (p. 138), altering 'information processing linked to reasoning, memory, implicit evaluation and even perception'[17] (p. 214). An authoritative synthesis of the literature describes the prevalent view that party cues activate party identity and thus 'guide reasoning'[8] (p. 136), and a recent empirical study concludes that the influence of party identity is 'so powerful' that when people receive cues from party leaders they override their ideological values[18] (p. 39). Further consequences of the activation of party identity and the partisan motivation to conform include that people are prone to 'interpret information through the lens of their party commitment'[19] (p. 235), 'rel[y] more on partisan endorsements and less on substantive arguments'[20] (p. 57) and 'often abandon their cherished values in favour of party loyalty'[17] (p. 214). As a result, 'even intellectually forceful messages [can get] distorted' by party cues[21] (p. 852), partisans can 'reject counter partisan messages, even when these messages align with their political values'[22] (p. 1181), party cues can 'interfere with partisans' ability to make decisions'[23], 'limit the effectiveness' of exposure to other types of information[24] (p. 5) and 'reduce to nil' the persuasive impact of other relevant content[25] (p. 811).

These consequences are not apparent in our data. We found no evidence that countervailing cues from favoured party leaders meaningfully diminished partisans' receptivity to persuasive arguments and evidence—in contrast to what one would expect if party loyalty distorted partisans' information processing. Importantly, this does not imply that party cues (from leaders or otherwise) have no effect on people's attitudes—on the contrary, we find clear evidence that they do. Rather, the implication is that cues do not meaningfully interfere with or distort partisans' processing of other (even counter-partisan) information. Notably, however, our results do not imply that directional motivated reasoning is limited in a more general sense. People have various identities and motivations beyond those derived from their party[8,26], and it is possible that these were 'activated' by our persuasive messaging treatments, explaining our results. We consider this question in greater detail in the discussion section.

Our second main conclusion regards the influence of party leaders on public opinion. While previous research demonstrates the power of cues from party leaders to influence partisans' attitudes[9,18,27], our results indicate that leader influence stops short of providing immunity from counter-partisan persuasive messages, or even appreciably diminishing their causal effect. Yet the implication of this finding for real-world public opinion formation should not be overstated; when party leader cues are ubiquitous and reliably propped up by partisan media talking points, even if people are exposed to counter-partisan messages (which is not guaranteed), these messages may represent a tiny portion of the otherwise-partisan causal effects acting on their opinions. Furthermore, in our experiment, exposure to the party leader cues typically influenced partisans' attitudes more strongly than exposure to the persuasive messages.

Nevertheless, our results indicate that counter-partisan persuasion is possible—given exposure to persuasive messages. This suggests that at least some cases of political polarization in public attitudes and beliefs are maintained by patterns of asymmetric exposure: greater exposure to cues from favoured party leaders versus counter-partisan messages; greater exposure to pro-partisan messages versus counter-partisan messages; or both. Thus, when faced with normatively troubling cues from party leaders, such as unsubstantiated

https://doi.org/10.1038/s41562-023-01551-7

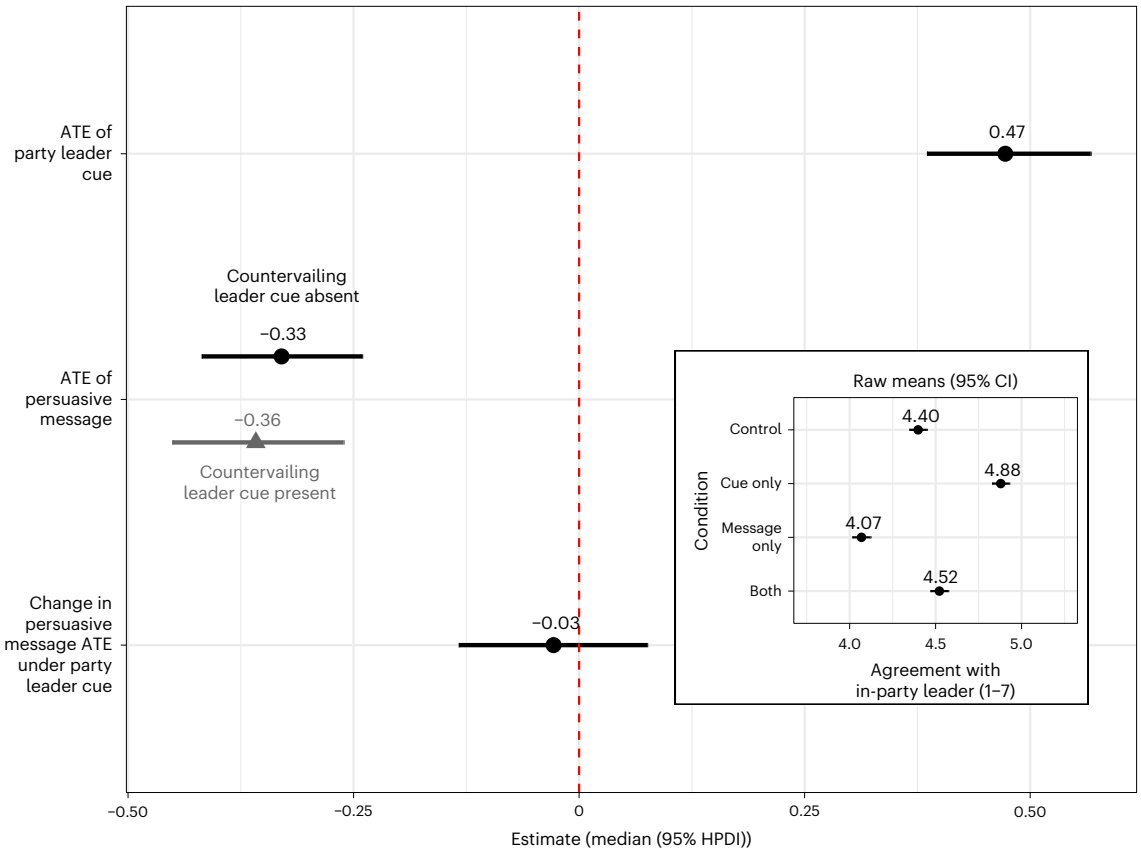

Fig. 1 | Key estimates from the primary multilevel model with raw means inset. Main panel: The top row is the ATE estimate of the party leader cue treatment (absent the persuasive message treatment). The first estimate in the second row is the ATE of the persuasive message treatment when the party leader cue is absent. The second estimate in the second row is the ATE of the persuasive message treatment when the party leader cue is present. The third row is the estimated interaction effect, which describes the change in the persuasive message ATE when the party leader cue is present (versus absent). Error bars are 95% HPDI. In-set panel: Mean value of the outcome variable (agreement with in-party leader) in each condition of our experiment design. Error bars are 95% confidence intervals (CI). Estimates are based on $n = 4{,}531$ respondents; 22,499 observations.

claims of election fraud, or health misinformation, our findings suggest that counter-communication strategies are not futile, and could be improved by making it harder for people to avoid counter-partisan messages—thereby forcing exposure—combined with other strategies such as sanctioning the partisan media and other elites for disseminating and justifying such cues[28].

In a pre-registered survey experiment conducted in September 2021, we recruited US adults online who identified as either Republican or Democrat and who reported voting for Donald Trump or Joe Biden, respectively, in the 2020 presidential election ($n = 4{,}531$; 22,499 observations). Each respondent was asked whether they agreed or disagreed with five policies, drawn randomly from a larger set of 24 contemporary American policy issues. The set of policies covered a broad array of issue areas, including immigration, the economy, healthcare, the military, foreign policy and the criminal justice system, among others (for more details, see Methods).

On each policy question, respondents were randomized to one of four conditions in a 2 × 2 design. The first treatment factor was whether they received a message intended to persuade them to either support or oppose the policy (message, no message), while the second treatment factor was whether they learned the position of their in-party leader on the policy—that is, a party leader cue (cue, no cue). For Trump-voting Republicans, the in-party leader was Donald Trump; for Biden-voting Democrats it was Joe Biden. Importantly, the in-party leader's position was always opposed to the position argued for in the persuasive message. Thus, it was a countervailing leader cue. We selected issues for which Trump and Biden had opposing positions, based on their public statements and voting records (for more details, see Methods). The persuasive message treatments were each approximately 150 words of text, and did not mention the policy positions of any political figures or party. Instead, they entailed substantive arguments for or against the policy, appealing to the values of the intended audience, and often cited evidence, such as statistics, in support of their argument (for more details, see Methods). Secondarily, we also randomized whether respondents in the cue condition would receive cues from their in-party leader only (one-sided cues), or cues from their in-party and out-party leader (two-sided cues), thus allowing us to probe generalizability across cue environments (for more details, see Methods).

Results

Following our pre-registered analysis protocol, we fit a multilevel linear regression model because the data are clustered by policy issue and respondent[29,30]. We re-code the outcome variable such that higher numbers indicate greater agreement with the in-party leader cue, allowing us to meaningfully aggregate across policy issues and partisans. Thus, the sign of the treatment effect of persuasive messaging is expected to be negative, while the sign of the treatment effect of the leader cue is expected to be positive.

Our model specification includes a parameter for each of our two treatment factors, as well as their interaction term. The parameter on the interaction term is our key quantity of interest: a positive interaction effect indicates that the average causal effect of the persuasive messaging is diminished by the presence of the countervailing leader cue. We fit the model in a Bayesian framework, and specify weakly

https://doi.org/10.1038/s41562-023-01551-7

Table 1 | Results and diagnostics of primary Bayesian multilevel regression model

Parameter group	Parameter	Estimate	Standard error	Lower 95% HPDI	Upper 95% HPDI	Effective samples	$\hat{R}$
Fixed effects	Intercept	4.40	0.09	4.22	4.56	1,337	1.00
	Cue	0.47	0.05	0.39	0.57	4,447	1.00
	Message	−0.33	0.05	−0.42	−0.24	5,686	1.00
	Message × Cue	−0.03	0.05	−0.13	0.08	8,000	1.00
Random effects (policy issues)	s.d.(Intercept)	0.41	0.07	0.30	0.55	2,418	1.00
	s.d.(Cue)	0.13	0.04	0.06	0.22	3,480	1.00
	s.d.(Message × Cue)	0.05	0.04	0.00	0.15	3,540	1.00
	s.d.(Message)	0.13	0.05	0.04	0.23	2,532	1.00
	Corr(Intercept, Cue)	−0.64	0.19	−0.94	−0.24	8,000	1.00
	Corr(Intercept, Message × Cue)	0.21	0.36	−0.54	0.81	8,000	1.00
	Corr(Intercept, Message)	0.23	0.25	−0.28	0.71	8,000	1.00
	Corr(Cue, Message × Cue)	−0.18	0.37	−0.81	0.57	8,000	1.00
	Corr(Message, Cue)	0.06	0.30	−0.55	0.60	4,834	1.00
	Corr(Message, Message × Cue)	−0.02	0.37	−0.69	0.69	8,000	1.00
Residual	Sigma	1.79	0.01	1.77	1.81	647	1.00

Note. The estimate is the median value of the posterior distribution for each parameter. Units are Likert scale points. For brevity, and because we do not analyse them, the respondent-level random effects parameters are not displayed here (for the full model output, see Supplementary Information 2.2.2).

informative prior distributions on all model parameters (for more details, see Methods). When reporting the parameters estimated by the model, we report the median of the posterior distribution and the 95% highest posterior density interval (HPDI). The HPDI is the narrowest region that covers the value of the parameter with 95% probability, given the data and model.

The results show that, on average, we find no evidence that partisans' receptivity to the persuasive messaging is meaningfully diminished by the countervailing party leader cues (Fig. 1; see also Table 1). First, we estimate the average treatment effect (ATE) of the party leader cue (in the absence of the persuasive messaging) to be 0.47 (0.39 to 0.57), approximately half a scale point on a seven-point Likert scale, a sizable and precisely estimated positive effect. As expected, partisans tend to change their attitudes in the direction of cues from their in-party leader when these are learned. The estimated ATE of the persuasive messaging (in the absence of the party leader cues) is smaller in magnitude—and opposite in direction, as expected—at −0.33 (−0.42 to −0.24). On average, partisans update their attitudes towards the message when countervailing cues from in-party leaders are absent.

How, then, does the effect of the persuasive messaging change when countervailing cues from in-party leaders are present? The point estimate of the interaction effect is −0.03 (−0.13 to 0.08), showing that the ATE of the persuasive messaging barely changes when countervailing cues from party leaders are present. Indeed, the ATE of the persuasive messaging in the presence of countervailing party leader cues is estimated to be −0.36 (−0.45 to −0.26) (Fig. 1). How precise is the null interaction effect? The upper bound of the 95% HPDI is 0.08; thus, we can conclude with greater than 95% confidence that the average causal effect of persuasive messaging does not decrease by more than 0.08 Likert scale points in the presence of a countervailing party leader cue. This is smaller than one-quarter of the estimated magnitude of the messaging effect here (−0.33). Moreover, dividing the upper bound of 0.08 by the standard deviation (s.d.) in our outcome variable (1.98) shows that we can confidently rule out a decrease in the effect of persuasive messaging that is greater than 0.04 s.d., typically considered a very small effect[11]. These results are robust across a series of alternative analyses (Methods).

Analysing the distribution of attitudes

We found no evidence that the presence of countervailing cues from their party leader meaningfully diminished the average causal effect of persuasive messaging on partisans' attitudes. However, perhaps the countervailing cue changed other features of this causal effect that are not revealed by the average attitude change. For example, perhaps the persuasive messages caused a minority of partisans to form attitudes that explicitly disagreed with their party leader when the cue was absent, but, when the cue was present, the messaging caused a larger number of partisans to agree slightly less—but nevertheless all still agree—with their party leader. This would provide an important qualification to our results thus far.

To illuminate this question requires looking beyond the effect of the treatments on average attitudes and looking instead at their effects on the distribution of attitudes[31]. Thus, in Fig. 2 we visualize the distribution of attitudes in each of the four conditions of our design, and we compute the difference between the distributions for those that did not receive a persuasive message versus those that did. This difference-in-distributions shows how the mass of the attitude distribution shifts in response to the persuasive messaging treatment. We compute this difference-in-distributions both for the condition in which the countervailing party leader cue is (1) absent and (2) present. We are interested in whether the difference-in-distributions differs between (1) and (2). Such a difference would indicate that the countervailing leader cue has an impact on the causal effect of the persuasive messaging that is not revealed simply by looking at average attitude change.

Figure 2 shows that this is not the case: the persuasive messages had a qualitatively similar causal effect on the distribution of partisans' attitudes, whether or not there was a countervailing cue from the party leader. Specifically, exposure to persuasive messaging caused the distribution of attitudes to shift such that fewer partisans agreed or strongly agreed with the position of the in-party leader—a score of 7 or 6 on the outcome scale, respectively—and more partisans explicitly disagreed with the position of the in-party leader—a score less than 4. As Fig. 2 shows, this distributional shift appears visually similar whether or not there was a countervailing cue from the in-party leader.

https://doi.org/10.1038/s41562-023-01551-7

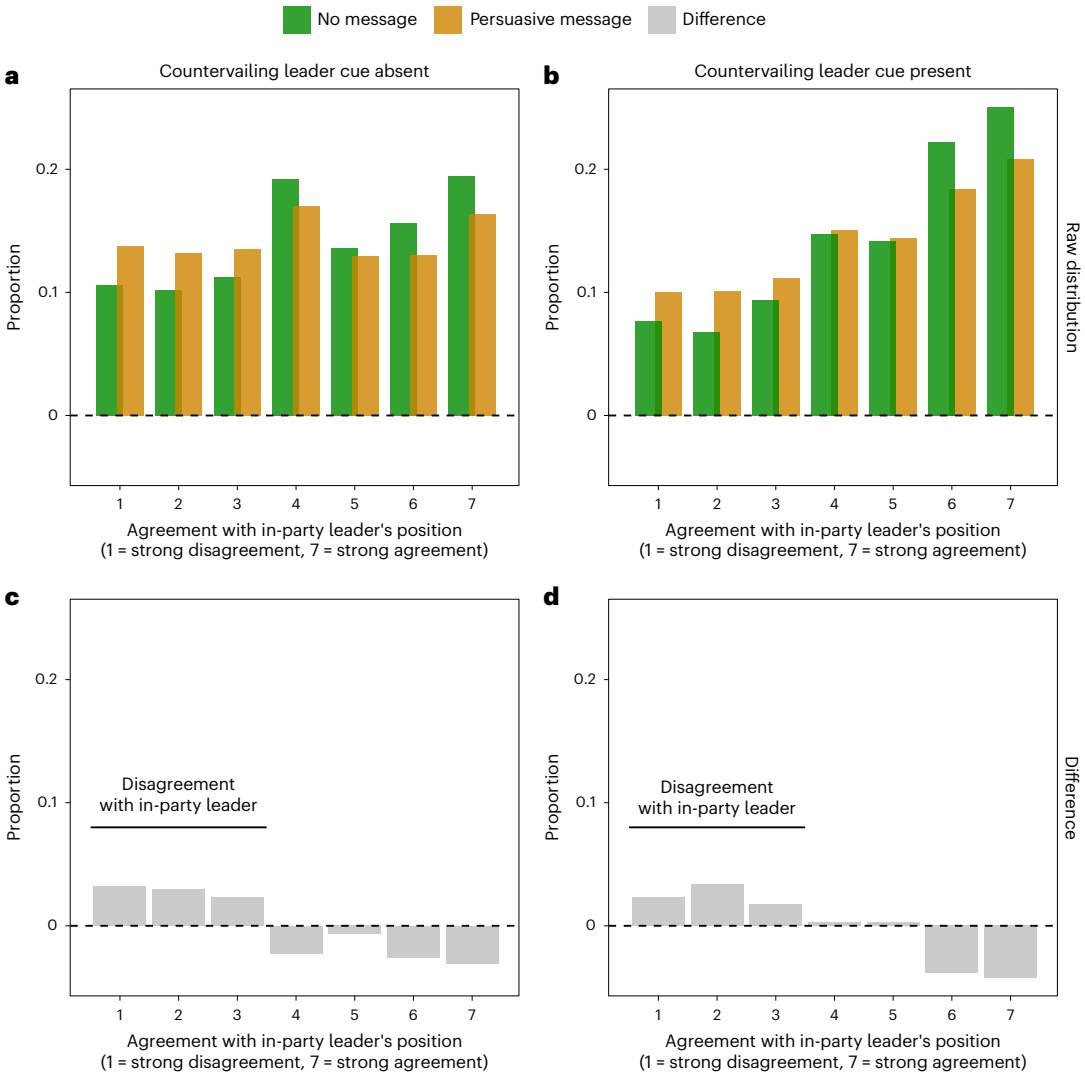

Fig. 2 | Analysing the effect of persuasive messaging on the distribution of attitudes. a,**b**, Histograms showing the distribution of the outcome variable disaggregated by the four main conditions of our design (the values on the y-axis are displayed as a within-condition proportion). Panel (**a**) shows the distribution of the outcome variable in the control (green) and message-only (orange) conditions. Panel (**b**) shows the distribution of the outcome variable in the cue-only condition (green) and the condition in which there was both a cue and a message (orange). **c**,**d**, The difference between the distributions for those observations that were assigned to receive a persuasive message versus those that were not. Panel (**c**) shows the difference for the conditions where the party leader cue was absent (i.e. the difference between the distributions shown in panel (**a**) directly above). Panel (**d**) shows the difference for the conditions where the party leader cue was present (i.e. the difference between the distributions shown in panel (**b**) directly above). The distributions are based on $n = 4{,}531$ respondents; 22,499 observations.

Heterogeneity across policy issues?

Now we examine whether our results are heterogeneous across policy issues. Figure 3a and Table 2 display the model-estimated interaction effects for each of the 24 policy issues in our design (the average interaction effect is also overlaid). Recall that positively signed interaction effects indicate that the causal effect of the persuasive messaging is diminished by the countervailing leader cue. In short, we find no evidence of this for any of the policy issues in our design: the 95% HPDIs all comfortably overlap with zero, with upper bounds that range from 0.10 to 0.21 Likert scale points across policy issues. Thus, while the intervals are of course wider than for the average interaction effect, we can still confidently rule out a decrease in the effect of persuasive messaging that is greater than 0.05–0.11 s.d. at the policy issue level, typically considered small effects (numbers calculated via dividing the upper bounds by the s.d. of our outcome variable, 1.98).

To visually reinforce this, Fig. 3b shows the corresponding conditional average treatment effects (CATEs) of the persuasive messaging for each policy issue, as estimated by the model (these estimates are also reported numerically in Table 2). Notably, there is visible variability between policy issues in the overall causal effect of the persuasive messaging treatment; the treatment causes qualitatively greater attitude change on average for some policies versus others. In contrast, however, there are only negligible differences within any given policy issue between the effect of the persuasive messaging when countervailing leader cues were present versus absent. This is visually apparent by examining the pairs of estimates for any given policy issue in Fig. 3b. In sum, we conclude there is little evidence of heterogeneity in our main result across these 24 policy issues; we find no evidence that the causal effect of the persuasive messaging was meaningfully diminished by countervailing leader cues for any of the issues in our study.

The minimal variation in the interaction effect across policy issues renders it difficult to 'explain' this variation by reference to other potentially relevant issue-level variables—such as the baseline level of political polarization on the issues. Indeed, the only issue-level

https://doi.org/10.1038/s41562-023-01551-7

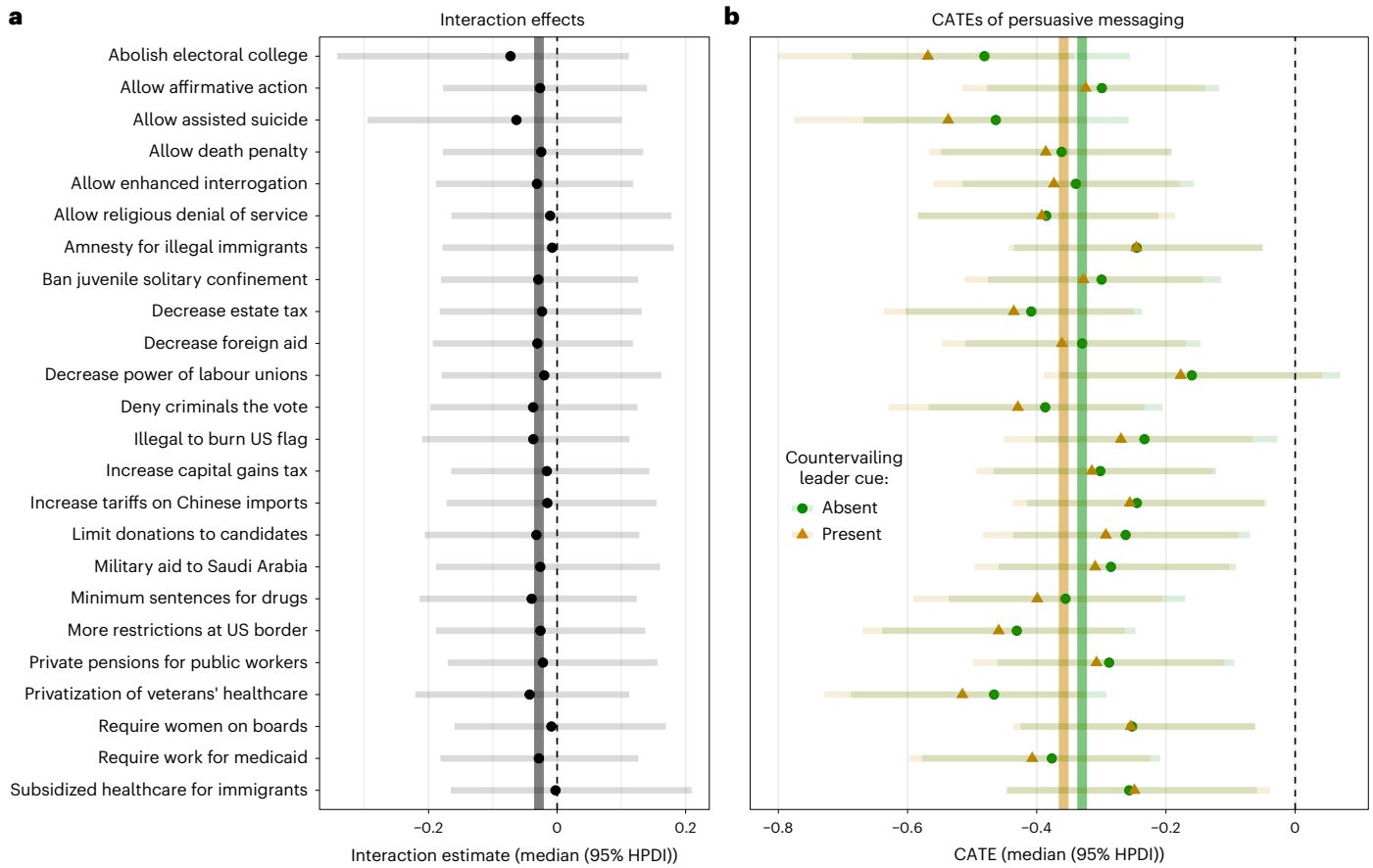

Fig. 3 | Estimated effects of persuasive messaging across the 24 policy issues in our design. a,**b**, Estimated interaction effects (**a**) and CATEs (**b**). The overall average interaction effect and CATEs are overlaid as vertical lines. Error bars are 95% HPDI. The estimates are based on $n = 4{,}531$ respondents; 22,499 observations.

pattern we reliably observe is a negative correlation between the baseline level of polarization on an issue and the ATE of the party leader cue on that issue: $r = -0.64$ (−0.94 to −0.24) (Table 1 and Methods). In other words, exposure to the party leader cue had a weaker effect on people's attitudes for issues that were more polarized at baseline, which is consistent with prior work[27,32,33]. Critically, however, we do not find any evidence that baseline polarization (or any other issue-level parameter) is associated with the extent to which party cues diminished partisans' receptivity to the persuasive messages (Table 1). That is, while the direct effect of the party leader cues on people's attitudes was larger for less polarized issues (where the cue is perhaps more surprising), we found no evidence that the interaction between the party cues and the persuasive messaging depended upon issue polarization.

While a key strength of our design is the large sample of policy issues—affording generalizability—clearly, we do not exhaust the space of all possible policy issues. Moreover, the estimates for the policy issues that we do observe could be more precise. This prompts the question: How much heterogeneity across policy issues could we expect to see if we had a much larger sample of policy issues and could estimate their effects with perfect precision? In other words, what is the plausible upper bound on heterogeneity across policy issues suggested by our model and data?

To answer this question, we turn to interpreting the model's formal estimate of the variation in our effects across policy issues (that is, the estimated s.d. of our effects across policy issues; Table 1). The multilevel model assumes that the true effects for each policy issue are sampled from an unobserved population of policy issues, represented as a multivariate Gaussian distribution. The model learns the mean and covariance of that population from the data, including the variation in our effects of interest. To interpret these variance parameters, we use simulation. In particular, we sample 1,000 hypothetical 'policy issues' from the population learned by the model, and we plot the distribution of the interaction effects (Fig. 4a) and CATEs (Fig. 4b) corresponding to these 1,000 policy issues. The distribution is arranged by the size of the interaction effect (for further details, see Methods).

The interpretation of the distribution is simple. Assume we were to conduct another experiment where we examined a new sample of policy issues that are similar, but not identical, to our current sample. Furthermore, assume we had infinite data and were thus able to estimate effects with perfect precision. In this new study, we would expect to observe a positive interaction effect for approximately one-third of policy issues, because approximately one-third of the distribution lies above zero in Fig. 4a (solid vertical black line). In other words, for one-third of policies we would expect the causal effect of persuasive messaging to be diminished by countervailing leader cues. Importantly, however, in many such cases the magnitude by which countervailing leader cues are expected to diminish the causal effect of persuasive messaging is minimal. This is shown by the distribution of CATEs in Fig. 4b. Only at the very extremes of the distribution—for example, the largest 2.5% of interaction effects, corresponding to 1 in 40 policy issues—is the countervailing leader cue expected to substantively diminish the causal effect of persuasive messaging.

We conclude from Fig. 4 that, for a majority of policy issues, it is most likely that the causal effect of persuasive messaging is not substantively diminished by countervailing cues from party leaders. However, for a minority of policy issues, such diminishing of the causal effect may occur. We also note that the model has much uncertainty over the distribution of effects across policy issues. Future work could

https://doi.org/10.1038/s41562-023-01551-7

Table 2 | Estimates of key parameters for the 24 policy issues in our design

Policy issue	Message ATE (party cue absent)	Message ATE (party cue present)	Interaction effect
Abolish electoral college	−0.48 (−0.69 to −0.26)	−0.57 (−0.80 to −0.34)	−0.07 (−0.34 to 0.11)
Allow affirmative action	−0.30 (−0.48 to −0.12)	−0.32 (−0.52 to −0.14)	−0.03 (−0.18 to 0.14)
Allow assisted suicide	−0.46 (−0.67 to −0.26)	−0.54 (−0.78 to −0.33)	−0.06 (−0.29 to 0.10)
Allow death penalty	−0.36 (−0.55 to −0.19)	−0.39 (−0.57 to −0.19)	−0.02 (−0.18 to 0.13)
Allow enhanced interrogation	−0.34 (−0.52 to −0.16)	−0.37 (−0.56 to −0.18)	−0.03 (−0.19 to 0.12)
Allow religious denial of service	−0.39 (−0.58 to −0.21)	−0.39 (−0.58 to −0.19)	−0.01 (−0.16 to 0.18)
Amnesty for illegal immigrants	−0.25 (−0.44 to −0.05)	−0.25 (−0.44 to −0.05)	−0.01 (−0.18 to 0.18)
Ban juvenile solitary confinement	−0.30 (−0.48 to −0.11)	−0.33 (−0.51 to −0.14)	−0.03 (−0.18 to 0.13)
Decrease estate tax	−0.41 (−0.60 to −0.24)	−0.44 (−0.64 to −0.25)	−0.02 (−0.18 to 0.13)
Decrease foreign aid	−0.33 (−0.51 to −0.15)	−0.36 (−0.55 to −0.17)	−0.03 (−0.19 to 0.12)
Decrease power of labour unions	−0.16 (−0.36 to 0.07)	−0.18 (−0.39 to 0.04)	−0.02 (−0.18 to 0.16)
Deny criminals the vote	−0.39 (−0.57 to −0.21)	−0.43 (−0.63 to −0.23)	−0.04 (−0.20 to 0.13)
Illegal to burn US flag	−0.23 (−0.40 to −0.03)	−0.27 (−0.45 to −0.07)	−0.04 (−0.21 to 0.11)
Increase capital gains tax	−0.30 (−0.47 to −0.12)	−0.31 (−0.49 to −0.13)	−0.02 (−0.16 to 0.14)
Increase tariffs on Chinese imports	−0.24 (−0.41 to −0.05)	−0.26 (−0.44 to −0.04)	−0.02 (−0.17 to 0.15)
Limit donations to candidates	−0.26 (−0.44 to −0.07)	−0.29 (−0.48 to −0.09)	−0.03 (−0.21 to 0.13)
Military aid to Saudi Arabia	−0.29 (−0.46 to −0.09)	−0.31 (−0.50 to −0.10)	−0.03 (−0.19 to 0.16)
Minimum sentences for drugs	−0.36 (−0.54 to −0.17)	−0.40 (−0.59 to −0.20)	−0.04 (−0.21 to 0.12)
More restrictions at US border	−0.43 (−0.64 to −0.25)	−0.46 (−0.67 to −0.26)	−0.03 (−0.19 to 0.14)
Private pensions for public workers	−0.29 (−0.46 to −0.09)	−0.31 (−0.50 to −0.11)	−0.02 (−0.17 to 0.16)
Privatization of veterans' healthcare	−0.47 (−0.69 to −0.29)	−0.52 (−0.73 to −0.33)	−0.04 (−0.22 to 0.11)
Require women on boards	−0.25 (−0.43 to −0.06)	−0.25 (−0.44 to −0.06)	−0.01 (−0.16 to 0.17)
Require work for Medicaid	−0.38 (−0.58 to −0.21)	−0.41 (−0.60 to −0.22)	−0.03 (−0.18 to 0.13)
Subsidized healthcare for immigrants	−0.26 (−0.45 to −0.06)	−0.25 (−0.45 to −0.04)	0.00 (−0.17 to 0.21)

Note. Estimates are medians of the posterior distribution for each policy–parameter combination, and the brackets report the lower 95% and upper 95% HPDI. The estimates are shrunk towards the average value of each parameter by the model, which mitigates against overfitting the data for any individual policy issue; this serves a similar function as performing a correction for multiple comparisons (for further detail, see ref. [51]). The estimates reported in this table correspond to those displayed visually in Fig. 3.

reduce this uncertainty by studying an even larger sample of policy issues or collecting more observations per policy issue, increasing the precision of the effects.

Relatedly, it is unlikely that the 24 policy issues in our set represent a random sample of all possible policy issues—we may be systematically missing particular types of issue, which adds some extra uncertainty to our conclusion here. For example, while our design incorporated a wide range of issues—from the politicized (for example, undocumented immigration) to the not-so-politicized (for example, capital gains tax)—it did not include hyper-salient and politicized issues such as abortion. Future work should test whether such issues exhibit a systematically different pattern of results, although we believe this to be unlikely given that party positions on those issues are already very well known (and thus their addition seems unlikely to change persuasion effects). Furthermore, testing the current hypothesis on such issues may be statistically challenging, given that attitudes are likely to be more crystallized and thus persuasion effect sizes are likely to be much smaller across the board. Another consideration is that, given the source from which we sampled our policy issues (Methods), it is plausible that our set of issues also omits those for which there is minimal (or zero) public communication. Future work could examine the interaction between party cues and persuasive messaging in such contexts.

Heterogeneity across respondents or cue environment?

Now we examine whether our results are heterogeneous across characteristics of our respondents (for example, demographics) or the nature of the cue environment (one- or two-sided party leader cues). This is pertinent, because previous work suggests that the strongest effects of partisan motivated reasoning occur among the most committed and politically engaged partisans[34], and that exposure to out-party (versus in-party) leader cues can have a stronger impact on information processing[35].

Figure 5 shows conditional average effect estimates for subgroups defined by demographics and the cue environment (one-sided or two-sided cue) for our main parameters of interest; it also shows estimates of the difference (interaction) between subgroups for each parameter (Table 3 presents the results numerically). These estimates come from separate multilevel models, in which we examine whether the relevant demographic covariate or cue environment condition (1) the ATE of the persuasive messaging (in the absence of the countervailing leader cue) and (2) the extent to which this ATE is diminished by the presence of the countervailing leader cue (for further details about the specification of these models, see Methods).

To summarize Fig. 5 and Table 3, we find limited evidence of heterogeneity, even where theory suggests we should find it. For example, even among strong partisans, we find no evidence that the causal effect of persuasive messaging was meaningfully diminished by the countervailing party leader cue: the interaction term indicating the change in the message ATE under the party cue is null, −0.08 (−0.23 to 0.07) (Table 3), and the upper bound of the 95% HPDI is equivalent to 0.07/1.98 = 0.04 s.d. Furthermore, even in two-sided cue environments—where partisans knew that not only was the position of

https://doi.org/10.1038/s41562-023-01551-7

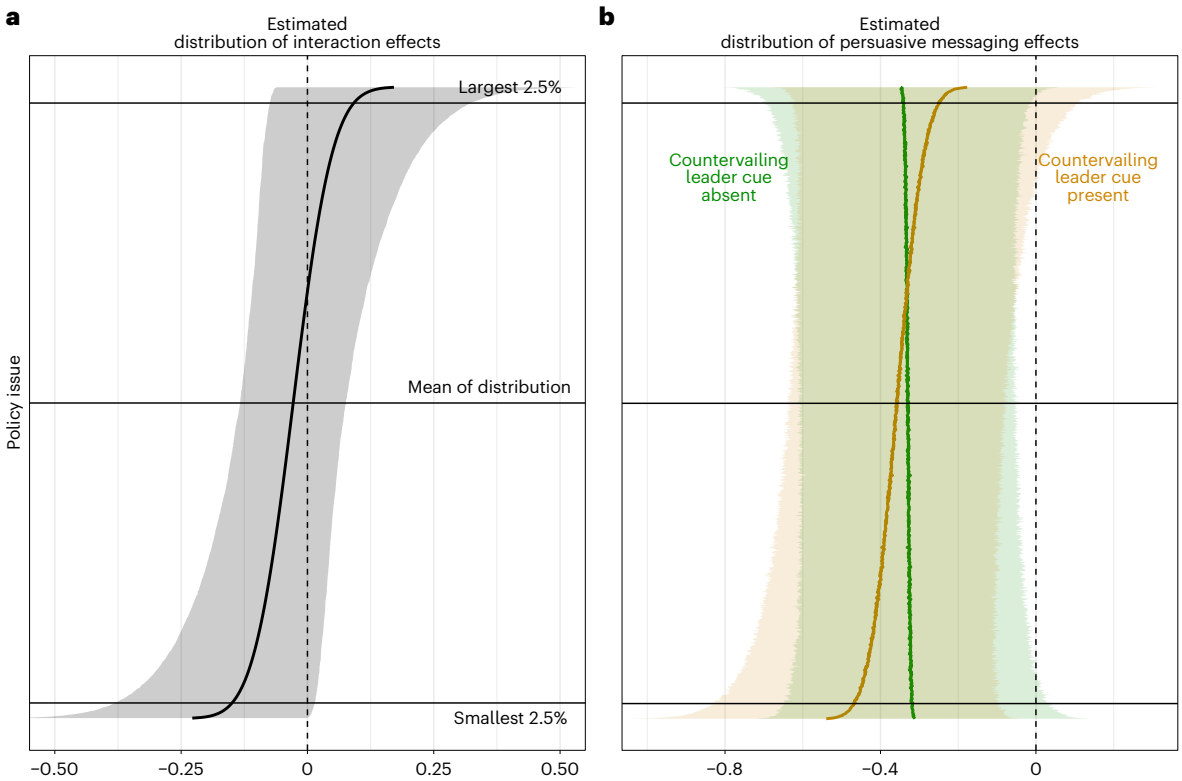

Fig. 4 | Posterior predictive distribution of interaction effects and CATEs of persuasive messaging across hypothetical policy issues. a, Interaction effects. **b**, CATEs. The solid vertical lines are the expected value of the distribution. The shaded regions are 95% posterior uncertainty (quantile) intervals over the distribution. The estimates are based on $n = 4{,}531$ respondents; 22,499 observations.

their in-party leader opposed to the message, but the position of the out-party leader was consistent with the message—we found no evidence that the causal effect of the message was reliably diminished: 0.01 (−0.14 to 0.16) (Table 3; see also Fig. 5), with an upper bound 95% HPDI equivalent to 0.16/1.98 = 0.08 s.d. This last result is especially notable, given that two-sided party leader cues had a direct effect on partisans' attitudes that was twice as strong as that of one-sided cues (Supplementary Fig. 14).

Discussion

In this paper we tested whether American partisans' receptivity to persuasive messaging was diminished by countervailing cues from favoured party leaders Donald Trump and Joe Biden. Our results showed that this was not the case: we found no evidence that the average causal effect of the persuasive messages was meaningfully diminished by the countervailing party leader cues. Moreover, this result held broadly across policy issues, demographic subgroups and cue environments.

These findings contrast with the notion that party loyalty and a partisan motivation to conform override people's values and interfere with, distort or otherwise limit their processing of counter-partisan messages. If such interference and distortion does occur, our findings suggest that it is relatively minor or uncommon or may be avoided with ease.

Importantly, this does not imply that party cues (from leaders or otherwise) have minimal impact on partisans' attitude formation per se. On the contrary, in line with much previous work[18,25,36–38], we found that exposure to such cues had a clear effect on partisans' attitudes—and in our particular case this effect was larger than the average effect of the persuasive messages. In this way, our results draw a clear distinction between two key research questions in political psychology: to what extent do party cues influence people's attitudes versus by what mechanism do they exert their influence? While there is relative consensus on the first question—party cues reliably influence people's attitudes, sometimes by a great deal—the second question remains unsettled[8,38]. Our results advance understanding of this second question because they indicate that party leader cues do not in general affect how people process counter-partisan persuasive messages. This result is inconsistent with an influential view of party cues' mechanism that contends that they trigger powerful party loyalties that can override people's values, and interfere with, distort or otherwise limit their processing of other types of (especially counter-partisan) information[8,17,18,21,22,24,25].

However, while our results place tighter constraints on the power of party loyalty and partisan motivation specifically to interfere with and distort information processing, they do not suggest that directional motivations in general are limited. People have various identities and motivations that are not reducible to their party[8,26]. One interpretation of our results is that the persuasive messages influenced people's attitudes by activating directional motivations in the opposite direction to the party leader cues; and party loyalty did not (or could not) override the influence of these other directional motivations. For example, our message treatments often attempted to appeal to people's values. Insofar as values are reinforced by one's community[39,40], appealing to people's values may change attitudes by triggering a directional motivation to conform. But this is just one example. Regardless of the various mechanisms one could posit for how persuasive messaging affects people's attitudes, our contribution remains unchanged: the causal effect of the mechanisms does not appear distorted by party loyalty.

The scale and design of our study contributes to existing evidence regarding the question of whether countervailing leader cues diminish the causal effect of persuasive information. Several previous studies

https://doi.org/10.1038/s41562-023-01551-7

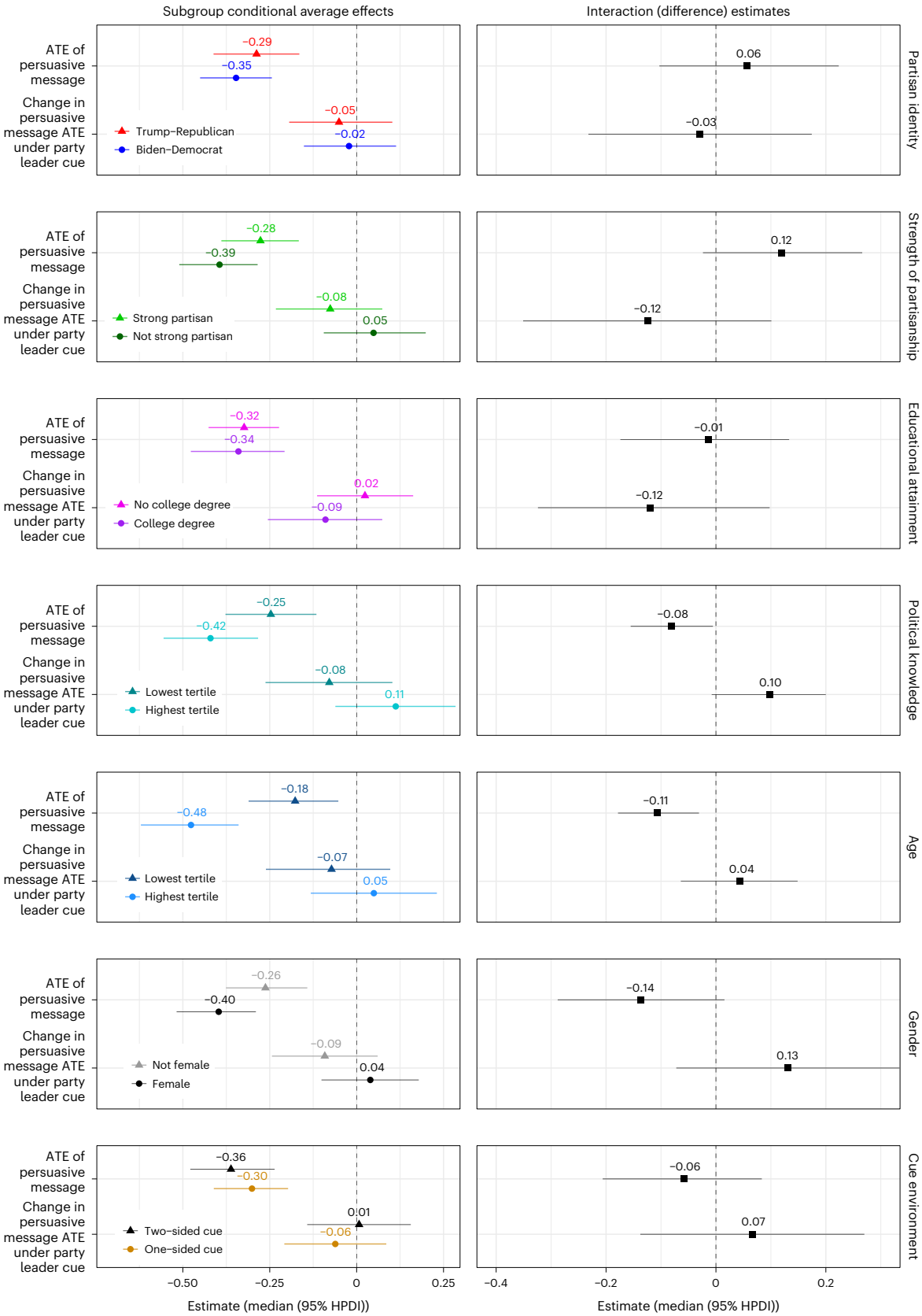

Fig. 5 | Subgroup conditional average effects and their corresponding interaction (difference) estimates. The estimates in the right-hand panels (the interaction estimates) model the difference between the corresponding estimates in the left-hand panels. Note that the full distribution of political knowledge and age covariates are analysed in the difference-test models, which is why the difference estimates from those models do not perfectly equal the difference between subgroup conditional average effects (which are based on splitting the covariates into upper and lower tertiles). Note also that the CATEs of the persuasive messaging (top rows in each of the left-hand panels) are those estimated in the absence of the party leader cues. Error bars are 95% HPDI.

https://doi.org/10.1038/s41562-023-01551-7

Table 3 | Results of subgroup conditional average effects models and corresponding difference-test (interaction) model for each covariate

Covariate	Parameter	Model	Subgroup value	Estimate	Lower 95% HPDI	Upper 95% HPDI
Age	Message ATE (absent party cue)	Subgroup 1	Lowest tertile	−0.18	−0.31	−0.05
		Subgroup 2	Highest tertile	−0.48	−0.62	−0.34
		Difference	—	−0.11	−0.18	−0.03
	Change in message ATE under party cue	Subgroup 1	Lowest tertile	−0.07	−0.26	0.10
		Subgroup 2	Highest tertile	0.05	−0.13	0.23
		Difference	—	0.04	−0.06	0.15
Cue environment	Message ATE (absent party cue)	Subgroup 1	One-sided cue	−0.30	−0.41	−0.20
		Subgroup 2	Two-sided cue	−0.36	−0.48	−0.24
		Difference	—	−0.06	−0.21	0.08
	Change in message ATE under party cue	Subgroup 1	One-sided cue	−0.06	−0.21	0.09
		Subgroup 2	Two-sided cue	0.01	−0.14	0.16
		Difference	—	0.07	−0.14	0.27
Educational attainment	Message ATE (absent party cue)	Subgroup 1	No college degree	−0.32	−0.43	−0.22
		Subgroup 2	College degree	−0.34	−0.48	−0.21
		Difference	—	−0.01	−0.17	0.13
	Change in message ATE under party cue	Subgroup 1	No college degree	0.02	−0.11	0.16
		Subgroup 2	College degree	−0.09	−0.26	0.07
		Difference	—	−0.12	−0.32	0.10
Gender	Message ATE (absent party cue)	Subgroup 1	Not female	−0.26	−0.38	−0.14
		Subgroup 2	Female	−0.40	−0.52	−0.29
		Difference	—	−0.14	−0.29	0.02
	Change in message ATE under party cue	Subgroup 1	Not female	−0.09	−0.24	0.06
		Subgroup 2	Female	0.04	−0.10	0.18
		Difference	—	0.13	−0.07	0.33
Partisan identity	Message ATE (absent party cue)	Subgroup 1	Biden-Democrat	−0.35	−0.45	−0.24
		Subgroup 2	Trump-Republican	−0.29	−0.41	−0.16
		Difference	—	0.06	−0.10	0.22
	Change in message ATE under party cue	Subgroup 1	Biden-Democrat	−0.02	−0.15	0.11
		Subgroup 2	Trump-Republican	−0.05	−0.19	0.10
		Difference	—	−0.03	−0.23	0.17
Political knowledge	Message ATE (absent party cue)	Subgroup 1	Lowest tertile	−0.25	−0.38	−0.12
		Subgroup 2	Highest tertile	−0.42	−0.56	−0.28
		Difference	—	−0.08	−0.16	−0.01
	Change in message ATE under party cue	Subgroup 1	Lowest tertile	−0.08	−0.26	0.10
		Subgroup 2	Highest tertile	0.11	−0.06	0.28
		Difference	—	0.10	−0.01	0.20
Strength of partisanship	Message ATE (absent party cue)	Subgroup 1	Not strong partisan	−0.39	−0.51	−0.28
		Subgroup 2	Strong partisan	−0.28	−0.39	−0.17
		Difference	—	0.12	−0.02	0.27
	Change in message ATE under party cue	Subgroup 1	Not strong partisan	0.05	−0.09	0.20
		Subgroup 2	Strong partisan	−0.08	−0.23	0.07
		Difference	—	−0.12	−0.35	0.10

Note. For each covariate, the table reports the results of three Bayesian multilevel regression models: one model for each of the two subgroup values, and one 'difference' model that tests whether the listed parameters are different across the two subgroups by interacting the covariate with the parameters in question. The full distribution of political knowledge and age covariates are analysed in the difference-test models, which is why the difference estimates from those models do not perfectly equal the difference between subgroup conditional average effects (which are based on splitting the covariates into upper and lower tertiles). The results reported in this table correspond to those displayed visually in Fig. 5. The estimate is the median value of the posterior distribution for each parameter. The message ATE parameter is the estimated effect of the persuasive messaging in the absence of the countervailing party leader cue. Units are Likert scale points. For full model outputs and diagnostics, see Supplementary Information 2.4.

have randomized substantive policy information alongside exposure to party cues[25,37,41–43], but their designs omitted a control group in which people received no information. Thus, the effect of exposure to counter-partisan information cannot be identified using these designs. One study[44] included the necessary control group, but the information treatment exerted little persuasive effect (no significant difference from the control group) when the countervailing cue was absent. Thus, the data cannot provide a clear answer to the question of whether such information loses its persuasive force when countervailing leader cues are present.

Another study[36] also included the necessary control group, randomizing policy information and countervailing party leader cues on two policy issues. For one issue, the information exerted a significant persuasive effect absent the countervailing cue that did not diminish in size when the countervailing cue was present. However, the key difference-in-difference (interaction) test was imprecisely estimated: the null effect could not confidently rule out a decrease of -0.25 points (on a 1–7 Likert scale) in the information effect when the cue was present, equal to more than half the magnitude of the information effect observed in that study and almost the full magnitude of the message effect observed in the current study. Thus, those data are unable to rule out a substantial decrease in the causal effect of persuasive messaging due to countervailing cues from party leaders. Moreover, the result concerned just a single policy issue and corresponding information treatment. This constrains its generalizability, given the wide variation in the effects of party cues and political messages across policy issues in prior work[27,33,45,46].

In sum, relevant previous work has used a design that either cannot answer the current research question or is beset by the twin challenges of low statistical power and small samples of policy issues—severely limiting their ability to comprehensively answer the current research question. By contrast, our design incorporated a larger sample of policy issues, persuasive message treatments and respondents.

Another implication of our results that warrants further discussion regards the constraint on party leaders to influence public opinion. While our results suggest that leader influence stops short of diminishing the causal effect of counter-partisan messages, a less rosy perspective on our results is that counter-partisan messages likewise fail to diminish the causal effect of party leader cues on public opinion—since the null interaction effect cuts both ways. This is the perspective adopted by a recent study[36] whose results are conceptually similar to ours. In considering these different perspectives, it appears that our results (and those) occupy a middle ground between the most and least normatively optimistic outcome. The most optimistic outcome is that exposure to arguments and evidence diminishes the causal effect of party leader cues, while the least optimistic is the reverse; that the latter diminishes the effect of the former. That we observe neither such outcome leaves room for the different perspectives.

Nevertheless, we contend that the least optimistic outcome is more plausible *ex ante*, owing to the relative dominance of party loyalty and partisan motivation for explaining people's political psychology and behaviour[8], as well as influential research that points towards the power of party loyalty and partisan motivation to override people's values, and to interfere with, distort or otherwise limit people's processing of other types of (especially counter-partisan) information. That we find little evidence of this outcome is therefore theoretically important, even if not the most normatively optimistic.

Our results provide evidence for the 'persuasion in parallel' hypothesis[11,12], which holds that most people respond to persuasive information by updating their attitudes towards the information, and by about the same amount. Our results suggest that this hypothesis holds even in contexts where people are explicitly confronted with the fact that the position of their in-party leader is opposed to that of the information. That we found this result to be largely homogeneous across policy issues, demographic subgroups and cue environments offers further support for the hypothesis. While clear evidence of heterogeneity may yet be found elsewhere, we do not find it here.

Now we consider some limitations of our study. The main limitation concerns the generalizability of our results to other contexts. A growing body of evidence shows that persuasion phenomena can be highly variable across contexts[1,45,47,48]. Notably, while we included an unusually large and diverse sample of policy issues and persuasive message treatments in our design, each issue had only two corresponding message treatments (one that contradicted the Biden cue and one that contradicted the Trump cue). Meanwhile the 'space' of potential messages that we could have included is extremely large given the numerous dimensions along which persuasive messages can vary[45]. It is possible that different types of messages would produce different results than ours.

Another potentially relevant dimension for generalizability concerns the party leader cue treatment. In line with previous work, our treatment consisted in simply communicating the position of the party leaders on the policy issue in question. However, in the real world of political communication, typically party leaders (and their supporters in the partisan 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 and evidence against alternative positions. The presence of such justifications may enable partisans to more easily ignore counter-partisan messages and simply fall into line with their party leader—as they are better able to rationalize this action[49]. To systematically test this proposition would involve adding a further treatment factor to our design: randomizing whether partisans receive a message supportive of their party leader. We consider this extension of our design a priority for future research.

To conclude, we reiterate our primary result: we found no evidence that American partisans' receptivity to persuasive messaging was meaningfully diminished by countervailing cues from party leaders. This result generalized broadly across policy issues, demographic subgroups and cue environments. Future work should further test the boundaries of this phenomenon.

Methods

The hypothesis, sample size, experiment design and analysis plan were pre-registered on 1 September 2021, before data collection, at https://osf.io/9gnaj. Respondents provided informed consent, and the survey was deemed exempt from requiring ethics approval by the Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects (ID: E-2285). Qualtrics survey software was used to collect the survey responses.

Experiment design

Respondents began the survey by providing informed consent and answering an attention check question that they were required to pass in order to continue with the survey (they were given one opportunity to pass this question; Supplementary Information 1.1). Following that, respondents answered a series of pre-treatment questions to measure their demographic and other characteristics, starting with their US party identification. Those classified as true Independents were not eligible to continue with the survey (for the classification scheme and other covariates, see Supplementary Information 1.1).

Respondents then arrived at the policy questions. Each policy question appeared on its own survey page, and respondents' attitudes were measured on a seven-point Likert scale running from strongly disagree (1) to strongly agree (7). On each policy question, respondents were randomized to one of the four treatment conditions with equal probability; that is, randomization occurred at the policy-question level, and was independent across policy questions. The party leader cue treatment (where assigned) always appeared before the persuasive message treatment (where assigned) on the survey page. As noted in the main text, for additional generalizability, we also randomized the

specific nature of the party leader cue treatment. Before seeing any policy questions, respondents were randomized with equal probability to one of two 'cue type' conditions, determining whether they saw a 'one-sided' leader cue or a 'two-sided' leader cue on policy questions where they were assigned to receive a party leader cue. Respondents assigned to the one-sided cue type condition saw cues from their in-party leader only, whereas respondents assigned to the two-sided cue type condition saw cues from their in-party and out-party leader. The two leaders disagreed on all of the policies in our set (if Trump supports, Biden opposes; and vice versa).

Policy issues and treatments

The policy issues and corresponding positions (cues) of Donald Trump and Joe Biden were sourced from the website https://www.isidewith.com, an online encyclopaedia that documents the real positions of US political figures on a range of contemporary American policy issues. The party leader cue treatment consisted in informing respondents whether Trump and/or Biden agreed or disagreed with the policy in question, alongside a thumbnail picture of the leader's face. Supplementary Information 1.2 provides further details regarding the set of policy issues and the party leader cue treatments.

For each policy issue, we developed two persuasive message treatments that were each approximately 150 words in length: one that contradicted Trump's position on the policy, and one that contradicted Biden's position. Supplementary Information 1.2 reports the persuasive message treatments in full and provides additional details; however, an example treatment is shown below for the policy, 'Allow the military to use enhanced interrogation techniques, such as waterboarding, to gain information from suspected terrorists.' The message treatment is that which would have been shown to Republican respondents, since the Trump cue supported the policy:

> Evidence shows that enhanced interrogation techniques are simply not effective. Therefore, they should not be allowed. In 2014, the US Senate published a 525-page report into the CIAs enhanced interrogation program. It found that enhanced interrogation techniques did not produce reliable intelligence, nor gain cooperation from suspects. It also found the CIAs justification for using the techniques relied on bad evidence. For example, information that led to Osama bin Laden was reportedly obtained through standard techniques, and suspects who were subjected to enhanced interrogation techniques in fact tried to provide false and misleading information about bin Laden's whereabouts. Ultimately, while such techniques may occasionally provide useful information, evidence suggests this is rarely the case. The question then becomes: is it worth violating international law by torturing people—who are effectively innocent until proven guilty by a jury—for mainly useless information? The answer is No. America is better than that.

Sample

We contracted with the survey provider Lucid to recruit 5,000 US adults quota matched to the national distributions of age, gender, education and region. The survey was fielded 2–13 September 2021. A total of 7,483 respondents began the survey, and a total of 5,071 respondents completed it—corresponding to 25,181 observations. Supplementary Information 2.1.1 provides information about sample demographics, and Supplementary Fig. 4 (in Supplementary Information 2.1.4) shows the points of attrition during the survey. No statistical methods were used to pre-determine our sample size, but our sample size is substantially larger than that reported in previous relevant publications[36,44].

Analytic strategy

Following our pre-registered protocol, our analysis restricts to respondents who (1) identified as Republicans or Democrats (including Independents who 'lean' to one of the parties) and (2) reported voting for Donald Trump or Joe Biden, respectively, in the 2020 US presidential election (n = 4,531; 22,499 observations). For those who did not vote in 2020, we use their stated preference for Trump or Biden. This restriction is designed to maximize the influence of the party leader cue by excluding pure Independents and a small minority of Republicans (Democrats) who preferred Biden (Trump) in 2020. Data collection and analysis were not performed blind to the conditions of the experiment.

As described in the main text, we analysed the data using Bayesian multilevel linear regression models. All multilevel models in this paper are fitted using the R package brms[50], and all achieved satisfactory convergence criteria, including $\hat{R}$ values less than 1.05 for all parameters, effective samples often in the thousands, and no divergent transitions during MCMC sampling (tables of results and diagnostics for all models are reported in Supplementary Information 2). The multilevel model offers us two advantages over, for example, ordinary least squares (OLS) with clustered standard errors, because it provides a principled framework for examining heterogeneity in our results across policy issues.

First, as well as estimating average effects aggregated across all 24 policy issues in our design, we also sought to estimate the effects at the level of each individual policy issue. The multilevel model allows us to do so while avoiding overfitting the data. The intuition here is simple: because our design contains many different policy issues, even though we have a large number of observations overall there is still a relatively small number of observations with which to estimate effects at the level of each individual policy issue. Thus, estimating these effects using just the raw data for each policy issue would produce some estimates that are large (or small) simply due to sampling variability. Such estimates would not generalize well to a new dataset; they are overfit. The multilevel model addresses this problem by adaptively 'shrinking' the individual estimates towards the mean estimate, thereby reducing overfitting and improving the out-of-sample accuracy of the individual estimates on average (for example, see Chapter 13 in ref. [30]). Another way of stating this benefit of multilevel modelling is that we are less likely to fall victim to increased type I error rates that result from the multiple 'comparisons' involved in examining our effects across each of 24 policy issues[51].

Second, the multilevel model allows us to formally estimate the heterogeneity in our effects across policy issues. The intuition here is again simple: our design contains a sample of policy issues, but really we would like to understand the heterogeneity in effects across the wider population of policy issues from which our sample is drawn. The multilevel model explicitly estimates the parameters of this population, given our data and some assumptions, thereby providing insight into the expected heterogeneity in effects across the wider population of policy issues (for example, see Chapter 13 in ref. [30]).

Primary multilevel model. Our primary multilevel model specification includes a parameter for each of our two treatment factors, as well as their interaction term, and the model allows these parameters (and the intercept) to vary across policy issues as well as across respondents. We specify weakly informative prior distributions on the model parameters, allowing the data to 'speak for itself'. For example, for our fixed effects of interest (two treatment factors, interaction term), we specify a normal distribution for the prior with mean = 0 and s.d. = 2. Given the scale of the outcome variable (1–7) and the dummy variables for the treatment factors, this prior distribution is consistent with a wide range of values for these parameters, such as an ATE of 3 (very large) or 0.01 (very small) Likert scale points—thereby allowing the data itself to determine the value of the parameter. The formal specification and diagnostics of our primary model is reported in full in Supplementary Information 2.2.1 and 2.2.2, respectively, and we reiterate that the model specification was pre-registered. The primary model forms the basis of the results that are reported in the results sections before

the section in which we examine heterogeneity across respondents and cue environment. Our primary results (shown in Fig. 1) passed a series of robustness checks.

First, there were small amounts of post-treatment differential missingness across conditions in the outcome variable (Supplementary Information 2.1.3 and 2.1.4 report balance checks and analyses of missing data, respectively). Thus, we conducted a 'worst-case' imputation analysis as a robustness check: we imputed values for the post-treatment missing outcomes that would work maximally in favour of us finding a positive interaction effect, and we refitted the multilevel model including the observations with these imputed values. This analysis provides an estimate of the interaction effect that assumes the worst case of bias caused by the differential missingness. However, the pattern of results from this analysis was substantively identical to those of our primary model (that is, a precise null interaction effect; Supplementary Information 2.2.3).

Second, recall that respondents answered five policy questions in our design, and randomization was independent across questions. A potential concern could be that, after being exposed to the treatments on the first or earlier questions, respondents answered systematically differently on the remaining questions. To confirm this was not the case, we subsetted the data by policy question order {1, 2, 3, 4, 5} and refitted the multilevel model to each subset. The results were substantively identical across the order of policy questions (Supplementary Information 2.2.3).

Third, in Supplementary Information 2.1.2 we show that we get the same key results from simple OLS regression.

Finally, while concerns have recently been raised about the inattentiveness of survey respondents recruited via Lucid[52], we consider it highly unlikely that inattentiveness can explain our pattern of results here. First, we note that, of the 7,483 respondents who started the survey, 1,145 (~15%) failed an initial attention check and were not eligible to continue with the survey. This suggests that our attention check was working to filter out a substantial portion of inattentive responders. Second, we observed clear and precisely estimated main effects of both the party leader cues and persuasive messaging treatments—yet no interaction effect (as per our key results). It is unclear how inattention could cause both the main effects to be clearly and precisely estimated, yet their interaction to be a precise null effect. Third, our results hold across levels of educational attainment and performance on a political knowledge quiz (Fig. 5), covariates that have been found to be correlated with attentiveness in other research. For these reasons, we consider it highly unlikely that inattention explains our pattern of results.

Heterogeneity across policy issues. In addition to estimating the variation across policy issues in our parameters of interest (that is, the interaction effect), our primary multilevel model also estimates the correlations between the parameters across policy issues. We examined these correlation estimates to determine whether any of our parameters were reliably associated across policy issues. The only reliable correlation we observed was a negative correlation between the intercept parameter and the parameter on the party leader cue treatment indicator: r = −.64 [−.94, −.24] (see Table 1). Given that the intercept indicates the degree of alignment with the in-party cue in the control group (that is, at baseline) collapsed across partisans, this correlation implies that policy issues with weaker levels of baseline political polarization tended to exhibit larger effects of the party leader cue. This is consistent with prior work[27,32,33], but tangential to our main results here (we found no evidence that the ATE of the messages nor the interaction term was associated with any parameters across policy issues; see Table 1).

To generate the distribution in Fig. 4, we sampled the estimates of one-thousand hypothetical 'new' policy issues from the posterior distribution of our fitted primary multilevel model (that is, a posterior predictive distribution). Thus, for each sampled 'policy issue' we obtained four parameters: an intercept, two treatment effects (one for the persuasive message effect, one for the party leader cue effect) and the interaction effect. For each policy issue, the CATE of the persuasive message in the absence of the countervailing leader cue is simply equal to the sampled value of the treatment effect of the persuasive message; by contrast, the CATE of the persuasive message in the presence of the countervailing leader cue is equal to the sum of (1) the sampled value of the treatment effect of the persuasive message and (2) the sampled value of the interaction effect. We ranked the distribution of interaction effects and CATEs across policy issues by order of size, running from the largest to smallest interaction effect (recall that positively signed interaction effects indicate that the causal effect of the persuasive messaging is diminished by the countervailing party leader cue). Finally, because our model is Bayesian, we performed the above process for each MCMC draw from the posterior distribution of the model and, for each rank-ordered policy issue (1 through 1,000), computed the mean and 95% quantiles across the draws. In Fig. 4, the mean corresponds to the vertical solid dark lines and the 95% quantiles are the shaded regions. A more detailed description of this simulation is provided in Supplementary Information 2.3.

Heterogeneity across respondents and cue environment. For the results reported in this section, we fitted additional multilevel models corresponding to each respondent-level covariate examined, as well as the cue-environment factor. The formal specifications and diagnostics of these models are reported in full in Supplementary Information 2.4. In Supplementary Information 2.4.3, we reproduce Fig. 5 in the main text but additionally show the estimates of the party leader cue effect from the different subgroup models.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The dataset generated and analysed during the current study is available in the Open Science Framework repository, https://osf.io/v3s72/.

Code availability

The code used to analyse the data during the current study is available in the Open Science Framework repository, https://osf.io/v3s72/.

References

1. Druckman, J. N. A framework for the study of persuasion. *Annu. Rev. Political Sci.* https://doi.org/10.2139/ssrn.3849077 (2021).
2. Jackson, C. & Duran, J. 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 (2021).
3. Eggers, A. C., Garro, H. & Grimmer, J. No evidence for systematic voter fraud: a guide to statistical claims about the 2020 election. *Proc. Natl Acad. Sci. USA* **118**, e2103619118 (2021).
4. Kiely, E., Robertson, L., Rieder, R. & Gore, D. 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/ (2020).
5. Brooks, B. Like the flu? Trump's coronavirus messaging confuses public, pandemic researchers say. *Reuters* https://www.reuters.com/article/us-health-coronavirus-mixed-messages/like-the-flu-trumps-coronavirus-messaging-confuses-public-pandemic-researchers-say-idUSKBN2102GY (2020).
6. Deane, C., Parker, K. & Gramlich, J. 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/ (2021).

7. Summers, J. 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 (2020).
8. Leeper, T. J. & Slothuus, R. Political parties, motivated reasoning, and public opinion formation. *Polit. Psychol.* **35**, 129–156 (2014).
9. Lenz, G. S. *Follow the Leader? How Voters Respond to Politicians' Policies and Performance* (Univ. Chicago Press, 2012).
10. Zaller, J. R. *The Nature and Origins of Mass Opinion* (Cambridge Univ. Press, 1992).
11. Coppock, A. *Persuasion in Parallel* (Univ. Chicago Press, 2022).
12. Coppock, A. *Positive, Small, Homogeneous, and Durable: Political Persuasion in Response to Information* (Columbia Univ., 2016).
13. Guess, A. & Coppock, A. Does counter-attitudinal information cause backlash? Results from three large survey experiments. *Br. J. Polit. Sci.* https://doi.org/10.1017/S0007123418000327 (2020).
14. Wood, T. & Porter, E. The elusive backfire effect: mass attitudes' steadfast factual adherence. *Polit. Behav.* **41**, 135–163 (2019).
15. Campbell, A., Converse, P. E., Miller, W. E. & Stokes, D. E. *The American Voter* (Univ. Chicago Press, 1960).
16. Bartels, L. M. Beyond the running tally: partisan bias in political perceptions. *Polit. Behav.* **24**, 117–150 (2002).
17. Van Bavel, J. J. & Pereira, A. The partisan brain: an identity-based model of political belief. *Trends Cogn. Sci.* **22**, 213–224 (2018).
18. Barber, M. & Pope, J. C. Does party trump ideology? Disentangling party and ideology in America. *Am. Polit. Sci. Rev.* **113**, 38–54 (2019).
19. Bolsen, T., Druckman, J. N. & Cook, F. L. The influence of partisan motivated reasoning on public opinion. *Polit. Behav.* **36**, 235–262 (2014).
20. Druckman, J. N., Peterson, E. & Slothuus, R. How elite partisan polarization affects public opinion formation. *Am. Polit. Sci. Rev.* **107**, 57–79 (2013).
21. Petersen, M. B., Skov, M., Serritzlew, S. & Ramsøy, T. Motivated reasoning and political parties: evidence for increased processing in the face of party cues. *Polit. Behav.* **35**, 831–854 (2013).
22. Nickerson, D. W. & Rogers, T. Campaigns influence election outcomes less than you think. *Science* **369**, 1181–1182 (2020).
23. Swigger, N., Buelow, M., Wirth, J. & Okdie, B. Partisans hear, but they don't listen: testing the limits of partisanship in risky decision making. *Am. Polit. Res.* https://doi.org/10.1177/1532673X221081252 (2022).
24. Guilbeault, D., Becker, J. & Centola, D. Social learning and partisan bias in the interpretation of climate trends. *Proc. Natl Acad. Sci. USA* **115**, 9714–9719 (2018).
25. Cohen, G. L. Party over policy: the dominating impact of group influence on political beliefs. *J. Pers. Soc. Psychol.* **85**, 808–822 (2003).
26. Bayes, R., Druckman, J. N., Goods, A. & Molden, D. C. When and how different motives can drive motivated political reasoning. *Polit. Psychol.* **41**, 1031–1052 (2020).
27. Tappin, B. M. Estimating the between-issue variation in party elite cue effects. *Public Opin. Q.* https://doi.org/10.31234/osf.io/p48zb (2022).
28. Nyhan, B. Why the backfire effect does not explain the durability of political misperceptions. *Proc. Natl Acad. Sci. USA* **118**, e1912440117 (2021).
29. Gelman, A. & Hill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge Univ. Press, 2006).
30. McElreath, R. *Statistical Rethinking: A Bayesian Course with Examples in R and STAN* (CRC Press, 2020).
31. Slothuus, R. & Bisgaard, M. Party over pocketbook? How party cues influence opinion when citizens have a stake in policy. *Am. Polit. Sci. Rev.* **115**, 1090–1096 (2021).
32. Slothuus, R. Assessing the influence of political parties on public opinion: the challenge from pretreatment effects. *Polit. Commun.* **33**, 302–327 (2016).
33. Clifford, S., Leeper, T. J. & Rainey, C. Increasing the generalizability of survey experiments using randomized topics: an application to party cues. *Polit. Behav.* (2023).
34. Taber, C. S. & Lodge, M. Motivated skepticism in the evaluation of political beliefs. *Am. J. Polit. Sci.* **50**, 755–769 (2006).
35. Nicholson, S. P. Polarizing Cues. *Am. J. Polit. Sci.* **56**, 52–66 (2012).
36. Agadjanian, A. When do partisans stop following the leader? *Polit. Commun.* https://doi.org/10.1080/10584609.2020.1772418 (2020).
37. Bullock, J. G. Elite influence on public opinion in an informed electorate. *Am. Polit. Sci. Rev.* **105**, 496–515 (2011).
38. Bullock, J. G. Party cues. In *The Oxford Handbook of Electoral Persuasion* (eds Suhay, E., Grofman, B. & Trechsel, A. H.) 129–150 (Oxford University Press, 2020).
39. Connors, E. C. The social dimension of political values. *Polit. Behav.* https://psycnet.apa.org/doi/10.1007/s11109-019-09530-3 (2019).
40. Tetlock, P. E. Thinking the unthinkable: sacred values and taboo cognitions. *Trends Cogn. Sci.* **7**, 320–324 (2003).
41. Ciuk, D. J. & Yost, B. A. The effects of issue salience, elite influence, and policy content on public opinion. *Polit. Commun.* **33**, 328–345 (2016).
42. Nicholson, S. P. Dominating cues and the limits of elite influence. *J. Polit.* **73**, 1165–1177 (2011).
43. Peterson, E. The scope of partisan influence on policy opinion. *Polit. Psychol.* **40**, 335–353 (2019).
44. Boudreau, C. & MacKenzie, S. A. Informing the electorate? How party cues and policy information affect public opinion about initiatives. *Am. J. Polit. Sci.* **58**, 48–62 (2014).
45. Blumenau, J. & Lauderdale, B. E. The variable persuasiveness of political rhetoric. *Am. J. Polit. Sci.* https://doi.org/10.1111/ajps.12703 (2022).
46. Yarkoni, T. The generalizability crisis. *Behav. Brain Sci.* https://doi.org/10.31234/osf.io/jqw35 (2020).
47. O'Keefe, D. J. & Hoeken, H. Message design choices don't make much difference to persuasiveness and can't be counted on—not even when moderating conditions are specified. *Front. Psychol.* **12**, 664160 (2021).
48. Hewitt, L. & Tappin, B. M. Rank-heterogeneous effects of political messages: evidence from randomized survey experiments testing 59 video treatments. Preprint at *psyArXiv* https://doi.org/10.31234/osf.io/xk6t3 (2022).
49. Williams, D. The marketplace of rationalizations. *Econ. Philos.* **39**, 99-123 (2023); https://doi.org/10.1017/S0266267121000389
50. Bürkner, P.-C. brms: an R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**, 1–28 (2017).
51. Gelman, A., Hill, J. & Yajima, M. Why we (usually) don't have to worry about multiple comparisons. *J. Res. Educ. Eff.* **5**, 189–211 (2012).
52. Ternovski, J., & Orr, L. A note on increases in inattentive online survey-takers since 2020. *J. Quant. Descr. Digit. Media* https://doi.org/10.51685/jqd.2022.002 (2022).

Acknowledgements

For helpful feedback on an earlier version of the manuscript we are grateful to D. Williams, A. Agadjanian, B. Guay and members of the Human Cooperation Lab at MIT and the Social Psychology Seminar Series at the University of Kent. We also thank P. Irvine for research assistance and three anonymous reviewers for their helpful comments. No specific funding was received for this project.

Author contributions

Conceptualization: B.M.T., A.J.B. and D.G.R. Data collection and analysis: B.M.T. Writing: B.M.T., with critical revisions from A.J.B. and D.G.R.

Competing interests

B.M.T is co-founder of a research organization that conducts public opinion research. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-023-01551-7.

Correspondence and requests for materials should be addressed to Ben M. Tappin.

Peer review information *Nature Human Behaviour* thanks David Ciuk and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Ben M. Tappin

Last updated by author(s): 2023/02/02

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Qualtrics survey software (2021) was used to collect the survey data in this study.
Data analysis	The statistical computing software R (version 4.1.0) was used to analyze the data in this study. The Bayesian models were fitted using the R package "brms" (version 2.17). The code used to analyze the data is available in the Open Science Framework repository at https://osf.io/v3s72/.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The data set generated and analyzed during the current study is available in the Open Science Framework repository at https://osf.io/v3s72/. The policy issues were sourced from https://www.isidewith.com.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	We report one analysis disaggregated by gender in the manuscript.
Population characteristics	Nothing to add. See above.
Recruitment	Participants were recruited online using the survey provider Lucid, which quota-matches its samples of U.S. adults to the national distributions of age, gender, education and region. Nevertheless, our sample is still a convenience sample and is not representative of the U.S. population; therefore, descriptive statistics such as the average value of the outcome variable should not be taken as an estimate of the population average value. However, because we are focused primarily on estimating treatment effects (not describing U.S. population opinions or other characteristics), we consider it unlikely that the use of a convenience sample by itself seriously impacts or biases the generalizability of our results (for more discussion/ detail, we refer to https://doi.org/10.1017/psrm.2018.10).
Ethics oversight	The survey was deemed exempt from requiring ethics approval by the Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects (ID: E-2285).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Quantitative experimental.
Research sample	Participants were recruited online using the survey provider Lucid, which quota-matches its samples of U.S. adults to the national distributions of age, gender, education and region. Our sample for primary analysis was restricted to Democrats/Republicans who reported voting for Biden/Trump (respectively) in the 2020 U.S. election. The sample is therefore unlikely to be representative of the general U.S. population. The mean age of our sample for analysis was 48 years; 52% were women; 80% were white; and 42% had a college degree.
Sampling strategy	Participants were an online convenience sample but were quota-matched to the U.S. national distributions of age, gender, education and region (see above). Our target sample size was pre-registered (n = 5000), and was based on our resource constraints.
Data collection	The data were collected online on participants' computers or phones. The researchers were not with participants during data collection. The researchers were not blind to experimental condition or study hypotheses during data collection.
Timing	The survey was fielded September 2–13, 2021.
Data exclusions	Exclusion criteria were preregistered and are described in the manuscript: "Our preregistered analysis restricts to respondents who (1) identified as Republicans or Democrats (including Independents who “lean” to one of the parties) and (2) reported voting for Donald Trump or Joe Biden, respectively, in the 2020 U.S. presidential election (n = 4,531; 22,499 observations)". A total of n = 1,627 respondents were excluded on the basis of these pre-registered criteria (see Supplementary Figure 4).
Non-participation	Some participants dropped out or did not answer particular questions during the survey. In total, there were 84 missing responses on the outcome variable. We analyze rates and implications of dropout/missingness at length in the Supplementary Information.
Randomization	Participants were randomized with equal probability to our experimental conditions, as described in the manuscript.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging