



Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpubePreempting polarization: An experiment on opinion formation[☆]Daniel Kashner^a, Mateusz Stalinski^{b,*}^a Kenneth C. Griffin Department of Economics, The University of Chicago, Chicago 60637, IL, United States^b Department of Economics, The University of Warwick, Coventry CV4 7AL, United Kingdom

ARTICLE INFO

JEL classification:

D83
D91
P16
Z18

Keywords:

Political polarization
Opinion formation
Partisan identity

ABSTRACT

Blind adoption of opinions put forward by political parties and influential figures can sometimes be harmful. Focusing on cases where the partisan gap on policy support has not yet arisen, we investigate whether its formation can be prevented by encouraging prior active engagement with non-partisan information. To address this question, we recruited N=851 Republicans for a study about net neutrality, an issue largely unfamiliar to the electorate, which refers to equal treatment of all internet traffic. In a pre-registered experiment, we randomly changed the order in which the following two types of information were provided: (i) partisan, underscoring Republicans' opposition and Democrats' support, and (ii) non-partisan, where the participants evaluated factual arguments about the pros and cons of the policy. Despite holding total information constant, we find that those who saw the non-partisan block first donated 46% more to a charity advocating for net neutrality ($p=0.001$). The treatment effect persisted in an obfuscated follow-up study, conducted several weeks after the intervention. However, we do not find an effect on donations when repeating the main study with a sample of Democrats.

1. Introduction

Reliance on partisan cues when considering high-stake policy issues may lead to prominent welfare consequences.¹ For example, in the healthcare context, despite the abundance of information supplied by non-partisan experts, political affiliation and opinion-based conservative media coverage were linked to a lower use of COVID-19 preventive measures and higher fatalities (Gollwitzer et al., 2020; Bursztyn et al., 2023). With party position shown as a dominating factor in evaluation of policies, overwhelming their objective content and one's ideological beliefs (Cohen, 2003; Druckman et al., 2013), the goal to understand the set of minimum conditions which can either prevent a partisan gap in policy support from arising or mitigate its impact on people's decisions should have a high priority.

Addressing polarization on a policy issue through information interventions is demanding. In fact, previous research offers suggestive

evidence that the provision of non-partisan information is often ineffective (e.g., Kahan et al., 2012),² a point corroborated especially in settings where people already have well-defined views (Long et al., 2023; Baysan, 2021). The challenge of tackling the partisan gap appears even more difficult given the findings that individuals choose to oppose a position they previously backed after the topic becomes politicized (Satherley et al., 2018). This indicates that party cues are not only powerful in persuading individuals to support or oppose a policy on an unfamiliar issue, but they can also easily undo previously formed opinions on a familiar one. Despite the multitude of discouraging evidence, there are still relevant factors which have not been explored. Crucially, researchers typically consider interventions where factual information about an issue is provided simultaneously with the party stances, or where the political context is well-known. This neglects the impact of *early* non-partisan campaigning, particularly using methods

[☆] This research is funded by the Becker Friedman Institute at the University of Chicago, United States. We would like to thank Leonardo Bursztyn, Joshua Dean, Angela Doku, Justin Holz, Alex Imas, Rafael Jiménez-Durán, Emir Kamenica, Andrew Kao, John List, Lester Lusher, Ricardo Perez-Truglia, Michael Price, Aakaash Rao, Andrew Simon, Egon Tripodi, and numerous seminar participants for helpful comments and suggestions. We are grateful to Jasmine Han for excellent research assistantship. This study was approved by the University of Chicago Institutional Review Board. The main experiment was registered in the American Economic Association Registry for randomized control trials under trial number AEARCTR-0009704. The additional experiment conducted with a sample of Democrats was pre-registered under trial number AEARCTR-0011471.

* Corresponding author.

E-mail addresses: kashner@uchicago.edu (D. Kashner), mateusz.stalinski@warwick.ac.uk (M. Stalinski).

¹ Schuldt et al. (2011) provides an example of the extent to which people follow partisan cues without a good understanding of the issue at hand, even on serious topics such as environmental policy. By slightly changing the policy name from "climate change" (used more often by liberal websites) to "global warming" (used more often by conservative websites), the authors generate a reduction in recognition of the problem as real among Republicans by 16 pp.

² Kahan et al. (2012) pointed out that science literacy enhances cultural polarization as a predictor of opinions on climate change.

that foster active engagement with information – a policy that an NGO or an issue advocacy group could pursue. We fill in the missing piece by exploring this opportunity.

Focusing on cases where the partisan gap on policy support has not yet arisen, we investigate whether its formation can be prevented by encouraging prior active engagement with non-partisan information. In this context, we are interested in whether the *order* in which individuals face partisan and non-partisan content matters for opinion formation. We explore the decisions made by individuals who first interact with factual information on an issue, and do not learn about the party stances until after they have formed a well-informed personal opinion. We ask whether, upon exposure to partisan cues, they exhibit a desire to conform to their party's position, or maintain their personal opinion. We compare this scenario to the counterfactual in which individuals first experience partisan information, followed by non-partisan information. Lastly, as a supplementary question, we investigate whether partisan information alters which arguments individuals find convincing when engaging with non-partisan content.

To address the research questions, we recruited a sample of 851 Republicans on a platform called Prolific, which matches researchers with participants for online surveys and experiments.³ The recruited individuals took a survey introducing the concept of net neutrality. Net neutrality is the principle that internet service providers (ISPs) – companies that connect users to the internet – cannot block or favor particular content. It means that they cannot create fast lanes for certain websites while slowing down others. This topic remains largely outside the public discourse and, thus, it is likely unfamiliar to the US electorate.

All participants in the study were provided with both (i) partisan and (ii) non-partisan information about net neutrality. The former consisted of texts and images revealing the history of policy changes related to net neutrality rules in the United States. The materials highlighted Democrats' support for net neutrality along with Republicans' opposition. When choosing this content, we attempted to eliminate any facts affecting the evaluation of the policy merits. Unlike the partisan block, the non-partisan information involved factual pros and cons of net neutrality, discussed in videos.⁴ These materials did not have any references to the party positions. At the end of the non-partisan block, we elicited the side of the argument that the participant preferred — we asked them to provide a short written statement explaining which argument they found convincing. As a central element of our design, we randomized the order in which (i) the partisan block and (ii) the non-partisan block were presented. In the Partisan-2nd condition, the partisan information followed the non-partisan information. In the Partisan-1st condition, the opposite was true.

As our main outcome, we measured donations to the Electronic Frontier Foundation (EFF), a charity advocating for net neutrality. Specifically, participants had to decide how to split a bonus payment of \$0.5 between themselves and the EFF. At the point of elicitation, the overall informational content was the same for both conditions — only the order differed. Several weeks after the main survey, as a robustness check, we conducted an obfuscated follow-up study to check if the effect of the intervention on the support for net neutrality persisted (in comparison to placebo issues).

³ Prolific ranks highly in data quality in comparison to its competitors, such as CloudResearch, Qualtrics, and Dynata (Eyal et al., 2021). You can find more information about Prolific by visiting their website <https://www.prolific.com/>, accessed 2023-01-05.

⁴ An important feature of our design is that both the partisan and the non-partisan blocks are “two-sided”. The former outlines the stance of both Republicans and Democrats, while the latter reports both pros and cons of net neutrality. This structure is intended to minimize experimenter demand effects, as it is difficult to guess what is the researchers' preferred political party or their stance on net neutrality rules.

We proceed to report the main results of the experiment. We find that preceding partisan information with non-partisan videos (Partisan-2nd group) was successful in increasing donations by 46.9% relative to when the non-partisan information was shown only after the party view was clear (Partisan-1st group). In particular, in the Partisan-2nd group the mean proportion of the bonus payment donated to the EFF was 21.6%, whereas in the Partisan-1st group it was equal to 14.7%. This translates into the average treatment effect of 6.9 pp ($p = 0.001$), or 0.23 SD. We conclude that preempting the partisan gap on an issue, especially one that would otherwise receive bipartisan support, is possible by preceding partisan information with two-sided non-partisan content (covering pros and cons) and ensuring active engagement with it in a way that leads to formation of a well-informed personal opinion. This outlines the set of sufficient conditions. We hope that our results will inspire a discussion on whether they are also necessary, or if a similar outcome can be achieved with less.

It is important to acknowledge limitations in interpreting the results. First, our strategy of acting preemptively, by providing non-partisan information before an issue becomes politicized, is only relevant for policy domains in which stakeholders (such as NGOs) can act at the time when individuals do not yet have a strong prior belief. Second, the experiment was conducted in a fairly artificial environment of an online experiment where the main incentivized outcome is low-stake (participants were splitting a bonus payment of \$0.5 between themselves and the EFF). The scope for preempting polarization on policy issues as well as the results' generalizability to high-stake environments are left to future scientific inquiry, which we hope our results will encourage. Lastly, we conducted an identical experiment with a sample of Democrats as a robustness check, and we do not find an effect on donations. This indicates that our protocol to preempt polarization by exposure to non-partisan information may not apply in all partisan contexts. We discuss the results of the additional experiment later on in the introduction and in Section 3.3, including possible ways of reconciling the difference in treatment effects for Republicans and Democrats.

Additionally, we report that receiving prior partisan information (Partisan-1st group), in comparison to not receiving it (Partisan-2nd group), alters which arguments Republicans consider convincing when viewing the non-partisan block. In particular, we find that the proportion of individuals who chose an argument in favor of net neutrality was higher in the Partisan-2nd group by 16 pp ($p < 0.001$). This outcome was measured during the non-partisan block, and thus, unlike for the donation outcome, total information was not held constant — here, one group knows their party's stance when watching the videos whereas the other does not. The result indicates that despite viewing the same substantive content (the partisan block did not contain any objective pros and cons), the participants demonstrated the capacity to fit their reception of the arguments to match their party stance. This finding is notable given that they not only had to pick a side, but also provide a written explanation of why in their own words. Overall, our result serves as a proof-of-concept that partisan signals can make people blind to other arguments.

We conducted a series of robustness checks to address potential concerns. First, we demonstrate that our results are probably not explained by attrition, as they are robust to applying (Lee, 2009) bounds. Furthermore, we strive to rule out alternative explanations. To that end, we provide evidence that our order intervention did not change the beliefs about Republican support for net neutrality. More importantly, we address the possibility that our results could have been driven by participants' desire to appear consistent in the eyes of the experimenter. It is natural to worry that once they selected a pro argument after viewing non-partisan information, they might choose to donate more just to satisfy experimenters' demand.⁵ Similarly, the subjects might

⁵ It is important to note that a desire to be consistent within self, rather than to appear consistent in the eyes of the experimenter, is one of the

be unwilling to “admit” that they were swayed by the partisan stance alone, which would be apparent through their donation choice. We dispel both criticisms by conducting a follow-up study, with an obfuscated purpose, several weeks after the initial experiment (which the median person took after 30 days). We report that the treatment effect on the support for net neutrality persisted while we recorded null results for two placebo issues — blockchain voting and carbon capture.

Lastly, we conducted an additional experiment with a sample of $N = 841$ Democrats as a further robustness check, and to uncover any potential heterogeneity of the treatment effect by party affiliation. We used exactly the same protocol as in the main study (including the same information screens and outcome wording) to ensure that the studies with Republicans and Democrats are directly comparable.

We proceed to report our findings. First, just like in the case of Republicans, we find that knowing the party position affects which non-partisan arguments about net neutrality Democrats consider persuasive. In the Partisan-1st group (already knowing that the Democratic party is in favor of net neutrality), the likelihood of making an argument in favor was higher by 7.6 pp ($p = 0.006$). Thus, the participants react to the partisan cue and shift their support in the direction consistent with the signal. As previously discussed, this outcome was measured at the end of the non-partisan block, so the total information was not held constant (one group knows the party position when watching the videos and the other does not). Second, unlike for Republicans, we do not find that the order of partisan vs. non-partisan information (here, the total information is held constant) affects the proportion of the bonus payment donated to the charity advocating for net neutrality. In the Partisan-2nd group, the proportion of the bonus payment donated was insignificantly higher than in the Partisan-1st group (1.5 pp, $p = 0.585$). There are multiple possible explanations for why the result for Republicans does not extend to Democrats in our study. First, the cost of ignoring or going against partisan signals may be significantly heterogeneous with respect to party affiliation, i.e., it could be higher for Democrats, rendering preemption efforts more difficult. Secondly, the finding might suggest that whether the party position is favorable or unfavorable is critical. It is plausible that despite higher personal support for net neutrality, learning that one’s party also supports it crowds out the need to offer financial support to the charity (as the party they trust will advocate for the policy). This extra channel is not present in the case of an unfavorable party signal, which could explain why we detect an effect on donations for Republicans but not for Democrats. Lastly, the base level of support for net neutrality among Democrats is already high, which makes the partisan signal less impactful in encouraging further support, thus leaving less room to detect effectiveness of preemptive efforts to hinder the increase in support following the signal in the Partisan-2nd group. Further discussion of the additional experiment with Democrats is provided in Section 3.3.

The paper is related to several strands of literature. First, our paper contributes to the literature on opinion formation and politically motivated reasoning. Opinion formation on policy issues is often affected by factors that are not directly related to the available objective information (e.g., Akesson et al., 2022; Lind et al., 2022). Chief among them, political signals play a major role, leading to partisan polarization. For example, a recent paper by Afrouzi et al. (2023) showcases that when individuals are presented with speeches on the topic of immigration by political leaders vs. placebo actors (holding the content constant and controlling for leader priming), the leaders have a significant impact on beliefs beyond the content of the speech. Satherley et al. (2018), who rely on a longitudinal survey on voters’ opinions about a flag

channels through which our intervention is intended to operate, and does not constitute an alternative explanation that we attempt to rule out. We discuss the possible mechanisms behind our results in the literature review section and in Section 3.1.3.

referendum before and after the issue became politicized, report similar conformity to the leader’s position.

One usual suspect for partisan polarization is motivated reasoning: even if Democrats and Republicans are exposed to the same facts, they may update their beliefs differently depending on how those facts align with the interests of their parties (e.g., Thaler, 2023; Di Tella et al., 2015; Schwardmann et al., 2022; Lord et al., 1979; Bolsen et al., 2014; Bisgaard, 2019; Schaffner and Roche, 2016). We provide causal evidence from a survey experiment consistent with this mechanism. Our evidence is based on a novel experimental design. While information-provision experiments typically randomize whether the information is shown or not, our experimental design randomizes the order in which the information is presented.

Second, our paper adds to the literature on methods of reducing issue polarization. Previous attempts to bridge gaps in political opinions yielded mixed results. These include efforts to provide unbiased information (e.g., Joslyn and Demnitz, 2021) and use message framing (Arpan et al., 2018; Bechtel et al., 2015; Singh and Swanson, 2017). Information provision is particularly ineffective when people already have defined beliefs (Long et al., 2023; Chen, 2022; Nyhan and Reifler, 2010; Baysan, 2021). This is not surprising in light of the literature showing that when polarization has already occurred, party endorsements have greater effects (Aaroe, 2012; Druckman et al., 2013).

In light of the small effectiveness of depolarization efforts, our paper focuses on preempting issue polarization with provision of two-sided information on the merits of the policy. Our novel result in an experiment with Republicans indicates that such a strategy can successfully inoculate individuals against an incoming strong partisan signal. Crucially, we achieved it by varying the order of partisan vs. non-partisan information, without changing the total information available to people. This highlights the importance of early campaigning by NGOs and issue advocacies before the debate becomes politicized.

When evaluating the non-partisan arguments on net neutrality, participants were required to write at least 20 words describing their reasoning for their preferred position, ensuring that some level of effortful thinking about the information was needed to move on in the study. This element of our protocol adds to a variety of previous research on the relationship between effort and attitude strength (e.g., Barden and Petty, 2008; Falk and Zimmermann, 2018; Strandberg et al., 2018). Another mechanism to consider is the possibility that participants who receive partisan information only after choosing their position on the basis of the non-partisan videos can be influenced by a preference for consistency. Previous research on cognitive dissonance avoidance has found that people tend towards behaving in ways that are consistent with their past behaviors, like, for instance, voters holding more favorable views of a candidate after having voted for them (Mullainathan and Washington, 2009; Shachar, 2003). Bénabou and Tirole (2011) explains this by proposing a theory wherein individuals have a sense of their identity and beliefs that they would like to act consistently with.

Lastly, due to the design of our protocol, our paper contributes to work focusing on the effects of the order or timing of information on a variety of related outcomes, including attitude strength (Haugtvedt and Wegener, 1994), political information effectiveness (Bositis et al., 1985), and willingness to donate (Bae, 2021). Regarding the question of whether “primacy” (participants privileging information received earlier) is more important than “recency” (information received later having a greater influence), previous results are mixed, with some evidence favoring the recency effect (Conlon et al., 2022; Brashier et al., 2021). Our results are inconsistent with the recency effect, as demonstrated by higher donations in the Partisan-2nd group, where the partisan block, expected to discourage donations from Republicans, was shown second. Methodologically, our work is most closely related to two other experimental papers on order and timing of information: Babcock et al. (1995) and Gneezy et al. (2020). The latter paper shows

that if participants are told about a bonus payment for recommending a lottery before (rather than after) they learn that it has a lower expected value than the alternative, a higher proportion of people recommend the less attractive option. Our paper offers a related result, with the monetary incentive replaced by a partisan signal.

The paper is organized as follows. Section 2 introduces the setting and outlines the experiment design. Section 3 provides a discussion of results and addresses potential concerns. Section 4 concludes.

2. Experiment design

2.1. Setting

In this section we introduce the issue of net neutrality, which provides the setting for the experiment, and explain its importance to our study. In a nutshell, net neutrality rules are designed to ensure that Internet Service Providers (ISPs) treat all internet traffic equally, which prevents them from favoring certain content e.g. by creating “fast lanes”. Proponents could argue that the ISPs should not have any impact on what content users access or its quality. At the same time, this lack of control implies that the ISPs cannot charge websites with high demand for bandwidth relatively more. This short summary of net neutrality rules masks the legal complexity of the issue, with the fine details remaining arcane to non-experts.⁶

Our decision to conduct the experiment centered on the issue of net neutrality stems from the unique opportunity associated with the way in which it is present in public discourse. First, the question of net neutrality rules is outside the scope of interest for most of the electorate, with infrequent media coverage. Even in 2015, a time with heightened media attention to net neutrality rules, polling data suggests that over 85% of U.S. adults heard little or nothing at all about the issue.⁷ This lack of familiarity increases the likelihood that the experimental intervention dominates any prior information that the participants might have.

Second, public policy polling of U.S. adults suggests that net neutrality received bipartisan support among survey participants exposed to information about the problem, both in 2017 (75% of Republicans and 89% of Democrats opposed dropping net neutrality rules)⁸ and in 2022 (65% of Republicans and 82% of Democrats favored reinstating net neutrality rules).⁹ A recent Morning Consult survey, with less background information, also does not reveal a significant partisan gap among the electorate, with 57% of Democrats, 60% of Independents, and 49% of Republicans backing net neutrality.¹⁰ At the same time, the issue is highly polarizing among political elites, with Republican and Democratic lawmakers clashing on whether to adopt net neutrality rules. This enabled us to create a set of materials inducing a strong sense of polarization around the topic for the purpose of the study. As a result, net neutrality offers a perfect framework for studying how to preempt a partisan gap from arising on a policy issue. This is the

⁶ Formally, the division between those who support and those who oppose net neutrality stems from opposing views on how internet service providers (ISPs) should be regulated — as “common carrier services”, to be governed by Title II of the Communications Act of 1934 and regulated by the Federal Communications Commission (FCC); or as providing “information services”, to be governed by the less stringent regulations under Title I of the Communications Act of 1934 and regulated by the Federal Trade Commission (FTC). Most supporters of net neutrality favor Title II regulation.

⁷ <https://www.cpc.udel.edu/content-sub-site/Documents/NatAgenda2015-PR-NetNeutrality-12-11-2015.pdf>, accessed: 2022-10-24.

⁸ https://publicconsultation.org/wp-content/uploads/2017/12/NetNeutrality_Quaire_121217.pdf, accessed: 2022-10-24.

⁹ https://publicconsultation.org/wp-content/uploads/2022/05/NetNeutrality_Quaire_0322.pdf, accessed: 2022-10-24.

¹⁰ <https://morningconsult.com/2022/04/27/net-neutrality-survey/>, accessed: 2022-10-24.

case because it combines three unique elements: unfamiliarity, the lack of ex-ante partisan gap, and abundance of publicly available materials suggesting strong polarization.

With Democratic lawmakers strongly supporting and Republican legislators strongly opposing net neutrality rules, we were constrained in the way in which we could use partisan materials to induce a sense of polarization. One option was to conduct the experiment with Republicans and use partisan information to demonstrate their party’s opposition to net neutrality. Another one was recruiting Democrats and inducing a sense of own party’s support. For the main experiment with an obfuscated follow-up, we chose to focus on the former alternative for two reasons. First, given the bipartisan support for net neutrality observed in surveys, this option allowed us to induce a stronger perception of issue polarization (by focusing on own party’s opposition) – otherwise, already high individual support could not move much in the counterfactual situation (that we experimentally create) where the issue gets politicized. Secondly, in the context of relevance of the study to various stakeholders, such as NGOs and issue advocacy groups, the question of how to overcome the impact of a party’s position on a policy issue (rather than enhance it) by providing non-partisan information is more pressing.¹¹ Taken together, these considerations led us to conduct the main experiment with Republicans. Nevertheless, we completed an additional experiment with a sample of Democrats and the same protocol (though without the follow-up study). The results of the extra experiment are discussed in Section 3.3.

2.2. Sample

2.2.1. Recruitment

We recruited a sample of participants through a platform called Prolific, which matches researchers with participants for online surveys and experiments (see Eyal et al., 2021 for the discussion of data quality). During recruitment, we targeted only individuals who reported affiliation with the Republican party. To do so, we relied on Prolific’s pre-screening data. In particular, we only allowed prospective participants who answered “Republican” to the following question: “In general, what is your political affiliation?”, with the other options being: “Democrat”, “Independent”, “None”, and “Other”. The recruitment window spanned three weeks — the first person enrolled on June, 30, 2022, and the last one on July 19, 2022. This allowed us to meet our sample size target of N = 800 despite the limited availability of Republicans on Prolific. In total, 802 individuals completed our main survey, including 602 participants in a study conducted after the pre-registration and 200 in the preceding pilot. In the pre-registration, we highlighted that we will pool the subjects from both studies when testing our hypotheses in order to overcome power limitations caused by Prolific’s scarcity of Republican respondents. As a robustness check, we verify that our main results hold even if we limit the sample to the observations recorded after the pre-registration.

2.2.2. Sample size and covariate balance

We report that 851 people passed the attention check and were assigned a treatment group – Partisan-2nd (423 individuals) or Partisan-1st (428 individuals). Overall, 802 individuals (94.2%) completed the survey i.e. answered all of the questions. This includes 393 (92.9%) in the Partisan-2nd group and 409 (95.6%) in the Partisan-1st group. For each specific outcome discussed in the paper, we report the results using the sample of all individuals for whom it is available. In particular, 807 participants chose the side of the argument after watching

¹¹ Depending on the framing, this could mean going against the party’s support or opposition to a particular proposition. The primary concern is that the party’s political elite might adopt a position that *disagrees* with the NGO’s position, influencing the party-affiliated electorate who might otherwise agree with the NGO’s position.

Table 1
Sample balance.

	Partisan-1st			Partisan-2nd			diff
	n	mean	sd	n	mean	sd	
Trump in 2020	406	0.72	0.45	382	0.72	0.45	0.001
Male	411	0.49	0.50	394	0.51	0.50	0.021
Age	411	42.68	14.30	394	41.21	13.77	-1.471
College	411	0.51	0.50	394	0.52	0.50	0.012
White	411	0.83	0.38	394	0.87	0.34	0.038
Income > 70k	411	0.48	0.50	394	0.53	0.50	0.041
West	411	0.17	0.38	394	0.16	0.36	-0.015
Midwest	411	0.23	0.42	394	0.23	0.42	-0.003
South	411	0.42	0.49	394	0.44	0.50	0.013
Northeast	411	0.17	0.38	394	0.18	0.38	0.005
Household size	411	3.00	1.39	394	3.01	1.49	0.003
Christian	411	0.49	0.50	394	0.52	0.50	0.034

Note: The table presents balance on covariates by treatment group (Partisan-1st and Partisan-2nd). In the Partisan-2nd group, participants (Republicans) see the non-partisan information block on net neutrality (non-political videos on its pros and cons) prior to receiving information about the position of political parties (Republicans against and Democrats in favor). In the Partisan-1st group, the order is reversed. For each covariate, we report the sample size, mean, standard deviation, and the difference in means, all by treatment group. We report the significance of the coefficient in the regression of each covariate on a dummy variable equal to one if the participant was assigned the Partisan-2nd group (* significant at 10%; ** significant at 5%; *** significant at 1%). We report the following covariates in order: (1) a dummy equal to one if the participant reported voting for Donald Trump in 2020 (*Trump in 2020*), (2) a dummy equal to one if a person is male (*Male*), (3) age, defined as the difference between 2022 and the reported year of birth (*Age*), (4) a dummy equal to one if they have at least a 4-year degree (*College*), (5) a dummy equal to one if they are white/Caucasian (*White*), (6) a dummy equal to one if they have household income exceeding \$70,000 (*Income > 70k*), (7) four regional dummies based on the state of residence (*West, Midwest, South, Northeast*), (8) the household size capped at 6 (*Household Size*), (9) a dummy equal to one if their religion can be classified as Christian (*Christian*). The table is based on a sample of N = 805 individuals (411 in the Partisan-1st group and 394 in the Partisan-2nd group) for whom we collected the main outcome — the proportion of the bonus payment donated to the Electronic Frontier Foundation, a charity supporting net neutrality. The only exception is *Trump in 2020* with N = 788.

the non-partisan information, and 805 individuals decided how much to donate to a charity supporting net neutrality. Since the latter is our primary outcome, we consider the 805 participants our main sample. We address the issue of attrition in Section 3.2.1, where we verify that our results are robust to applying (Lee, 2009) bounds.

Table 1 indicates that the main sample is well-balanced. None of the twelve reported covariates reveals a significant difference in means by treatment at the 10% significance level. Additionally, in the online appendix, we demonstrate that the distributions of three key categorical demographics: religion, income, and education, do not differ by treatment even at a high level of granularity.

Lastly, we corroborated our recruitment strategy targeting Republicans by eliciting our own electoral measures in the survey: support for Republicans for Congress and for Donald Trump in 2024, as well as measures of affective polarization. Details of the analysis are provided in the online appendix. There, we provide evidence supporting a conclusion that our sample consists of individuals exhibiting nearly universal congressional support for Republicans, significant approval for Donald Trump as a presidential candidate, and a large pro-Republican ingroup bias.

2.3. Study flow

Fig. 1 summarizes the flow of the study. Participants recruited on Prolific were asked to complete a short Qualtrics survey. The wording of the questions as well as the instructions given to the subjects are provided in the online appendix.

2.3.1. Preliminaries

Following the consent form, we collected basic demographics. We included an attention check among the questions presented in this

section. Subsequently, we introduced ourselves as non-partisan researchers interested in public policy. We continued by offering a concise definition of net neutrality — our issue of interest. Specifically, we highlighted that net neutrality ensures that the internet service providers (ISPs) cannot “block or favor particular content, websites, or applications”. To ensure participants’ engagement with the definition, we asked a comprehension question about its meaning.

2.3.2. Treatment

Following the issue introduction, the subjects were shown two types of information blocks: (i) partisan and (ii) non-partisan. The order of the blocks was determined by a randomly assigned treatment condition. Participants in the Partisan-2nd group encountered the non-partisan block first, followed by the partisan block. The opposite was true in the Partisan-1st group. Below, we describe both types of information in detail.

Partisan block. The partisan block is a slide show outlining the history of the net neutrality debate. Its main goal is to communicate to the viewer that Democrats support and Republicans oppose net neutrality rules. To that end, we indicated that Barack Obama introduced net neutrality legislation, while the FCC chairman Ajit Pai, appointed by Donald Trump, overturned it. Furthermore, we relied on tweets and images of politicians with nation-wide recognition to provide evidence of divergence in the party stances on the issue. In an effort to induce the feeling of substantial issue polarization, the block contains images intended to result in a strong emotional reaction. In addition to associating the debate with some of the most polarizing figures such as Donald Trump, Barack Obama, Nancy Pelosi, and Ted Cruz, we provided images of protesters meant to evoke the caricatured image of Democrats portrayed in conservative media, in order to further trigger a sense of partisan conflict. In designing the partisan block, we strove to minimize the presence of any factual information that could constitute meaningful arguments in favor of or against net neutrality.

Non-partisan block. The non-partisan block consists of videos discussing pros¹² and cons¹³ of net neutrality, focusing on evidence-based arguments. To encourage attention, we made it not possible to fast forward the videos, or continue with the survey unless they were played in full. When editing the video materials, we removed any indication of their sources (credits were given at the end of the survey) to ensure that we did not contaminate the arguments with hints of partisanship.¹⁴

¹² When viewing the discussion of the pros, the participants learned that without net neutrality, internet traffic is not treated equally. In particular, ISPs may charge companies for access to faster lanes, a move which could benefit large corporations, who can afford to pay. In the extreme, the ISPs could even deny a website access entirely. Furthermore, some services could slow down if their providers do not upgrade their plan with the ISP. This is exemplified by a case study of Netflix being slowed down by Comcast. The punchline is that the costs of purchasing higher speed could be passed to consumers, and thus users may end up paying more for their favorite services.

¹³ The discussion of cons opens with a statement that under net neutrality, ISPs would charge everyone equally regardless of how much data they send through the Internet. The viewers are reminded that smooth operation of the Internet is dependent on physical infrastructure that has limited capacity. A case study of Google and Netflix shows that the two providers occupy more than half of available bandwidth. This led to the idea that ISPs should create fast lanes – you use more, you pay more – the punchline of the video. Lastly, the discussion links the fast lanes to innovation, pointing out that paying a higher price for more usage would prompt the companies to invest in more efficient transmission technologies, benefiting everyone.

¹⁴ To prepare videos for the experiment, we relied on 0:39-1:53, 1:59-2:07, 2:16-2:25 fragments of BBC’s “What is net neutrality and how could it affect you?” (<https://www.youtube.com/watch?v=zq-2Yk5OgKc>) and 1:29-2:39 fragment of PragerU’s “What Is Net Neutrality?” (<https://www.youtube.com/watch?v=aiZ8xwvycXA>). In both cases, we edited the videos to remove the source to ensure that the participants do not learn that the videos were by the BBC or PragerU.

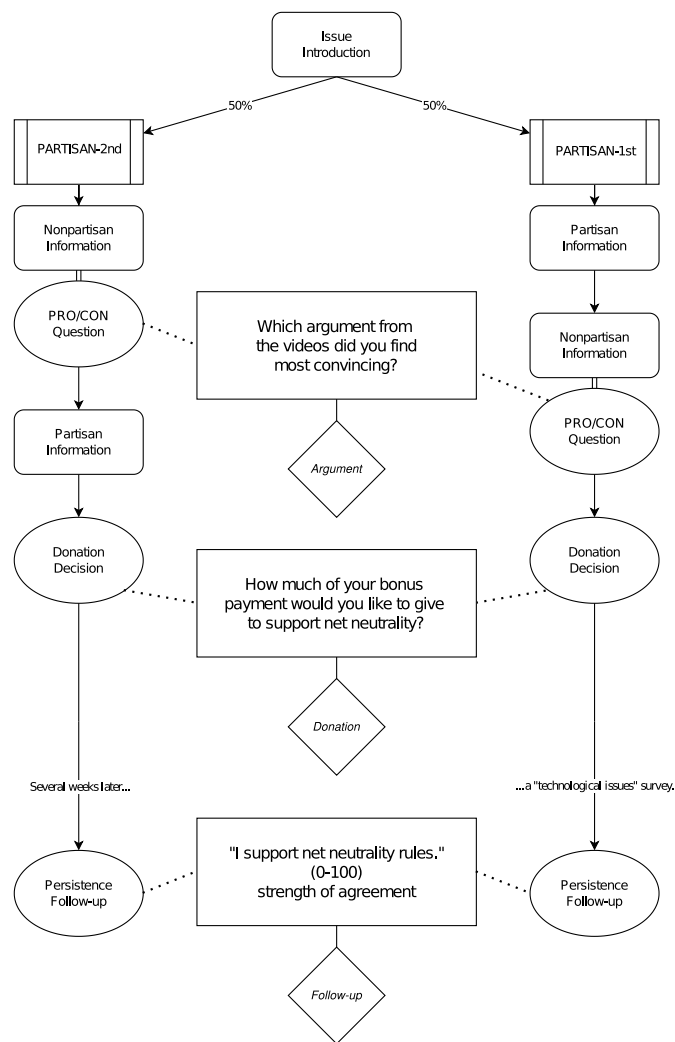


Fig. 1. The study flow. *Note:* The figure provides an overview of the study flow. First, it indicates that participants are randomly assigned to one of the two treatment groups: Partisan-1st or Partisan-2nd. Below the name of each treatment group, the figure shows the order of information blocks (rectangular shapes) and key outcomes/decisions (oval shapes). There are two information blocks. In the non-partisan block, participants watch two non-political videos on the pros and cons of net neutrality. In the partisan block, they receive information about the position of political parties (Republicans against and Democrats in favor) through a slide show about the timeline of the net neutrality debate. The figure highlights three key outcomes: (1) whether someone reported writing an argument in favor/against net neutrality after watching the videos, (2) the proportion of the bonus payment donated to the Electronic Frontier Foundation, a charity supporting net neutrality, (3) support for net neutrality in an obfuscated follow-up survey conducted a few weeks after the survey with the information intervention. In the middle of the figure, we provide wording associated with each of the outcomes.

After the participants completed the videos, we asked them which video was more convincing (pro or con). We requested that the subjects explain in their own words an argument or a reason from the videos that convinced them. Every participant had to write at least 20 words of explanation in a text box to proceed further. This exercise is a critical component of our intervention — it ensures that the participants *actively* engage with information provided and consider the merits of the policy.

An important feature of our design is that both information blocks were two-sided. The partisan block focused on both Republicans and Democrats, whose opinions on the net neutrality rules differed. Similarly, the non-partisan block contained videos arguing both pros and cons of net neutrality. This way, it was unclear to the participants which side of the argument or which political party the researchers are

likely to support. We hope that these efforts minimized the impact of experimenter demand effects on our results.

Lastly, to avoid the issue of low attention to the videos in the non-partisan block, as well as to minimize the risk of participants submitting low-quality or arbitrary arguments, we incentivized the quality of the argument with a bonus payment of \$0.6 for the 25% best comments.¹⁵ We were clear that the quality assessment is independent of the side of the debate that the participant chose to support. Importantly, we told the participants about the incentive structure before they played the videos. Lastly, for the purpose of awarding bonus payments, a research assistant graded all comments by assigning them scores from 0 to 100. We verified that there is no significant difference between the average scores by treatment. Writing arbitrary comments would certainly lower their quality score, hence there is no evidence that such behavior occurred differentially across groups.

2.3.3. Outcomes

We pre-registered two outcome variables. First, the main outcome of the study is the proportion of the bonus payment donated to the Electronic Frontier Foundation (EFF), a charity advocating for net neutrality. All participants were offered a \$0.50 bonus payment in the donation screen, which they could split with the EFF. We introduced the EFF as a non-profit organization focused on protecting net neutrality, rated 92.88 on CharityNavigator.org. We emphasized to the participants that the donation is their chance to back net neutrality, which enhances the interpretation of the outcome as an incentivized measure of support for the issue. The donation decisions were collected after the participants in the two treatment groups experienced both the partisan and the non-partisan block, albeit in a randomly assigned order. This means that the information that they received throughout the study was held constant at the point of elicitation.

Additionally, as a secondary outcome, we measured the proportion of participants who chose a pro argument after watching the non-partisan videos. Specifically, after the block was completed, we told the participants that “on the previous page, [they] explained which argument from the videos about net neutrality [they] found the most convincing”. Subsequently, we asked whether it was an argument in favor or against. It is important to note that here, unlike for the donation outcome, total information is not held constant across groups at the point of elicitation. We collect the preferred side of the argument straight after displaying the non-partisan videos (see Fig. 1 for details of the study flow), which means that the Partisan-1st group knows both partisan (shown first) and non-partisan information, while Partisan-2nd group knows only the non-partisan information. In other words, we measure the effect of knowing the party position on support for pro vs. con arguments shown in the videos.

Controls. Following the pre-registration, in the text of the paper we focus on reporting our results on the basis of regressions without any controls. However, as a robustness check, we discuss specifications where we control for the following covariates: age, household size, as well as dummy variables equal to one if someone: (i) is male, (ii) earned at least a 4-year degree, (iii) has gross household income exceeding \$70,000, (iv) is Christian, (v) is white/Caucasian. For regressions with controls reporting outcomes from the follow-up survey, we also include the number of days elapsed between the two surveys.

2.3.4. Additional components

After collecting the outcome variables, we asked the participants additional questions important for robustness analysis and further illuminating our sample. First, we elicited beliefs about support for net neutrality rules among Republicans and Democrats. Second, we asked the participants about their electoral preferences: (i) whether they

¹⁵ This also reduces the likelihood that the treatment (knowing vs. not knowing the party stance) affects attention to the non-partisan block.

intend to vote in 2022 midterm election, (ii) whether they support Democratic Party for Congress, and (iii) whether they would vote for Donald Trump in 2024 if he runs for president. Lastly, we collected measures of affective polarization via a feeling thermometer — the participants provided their feelings towards both Democrats and Republicans on a scale from 0 to 100.

2.3.5. Follow-up survey

Lastly, a few weeks after the initial survey, we recruited participants from our original sample for a follow-up study, with an obfuscated purpose. The median person took the follow-up survey 30 days after the survey with the information intervention (the minimum was 18 days and the maximum was 54 days). The follow-up study focused on three technology issues — blockchain voting, carbon capture, and net neutrality, and involved elicitation of support for these policies. The purpose of the follow-up was to test whether the treatment effect persists over time, which serves as a robustness check, helpful in refuting some of the alternative explanations of our results. Details are provided in Section 3.2.3.

3. Discussion of results

3.1. Main results

Fig. 2 provides a summary of the main results.

In Fig. 2(a), we report mean values of the main outcomes by treatment. First, we find that learning the party position reduces the subsequent effectiveness of non-partisan information in inducing support for net neutrality. In the Partisan-2nd condition, where the decision is based solely on the non-partisan videos, the proportion of individuals in favor of net neutrality was higher than in the Partisan-1st condition by 16 pp ($p < 0.001$). Participants (Republicans) in the Partisan-1st group previously viewed a slide show which hinted that Republicans oppose and Democrats support net neutrality.

Second, we report that, holding total information received in the survey constant, the order of partisan and non-partisan blocks has a significant effect on the support for net neutrality, measured through a donation to a charity advocating for it (EFF). Preempting partisan information with non-partisan videos (Partisan-2nd), was successful in increasing donations by 46.9% ($p = 0.001$) relative to the case in which the non-partisan information was shown only after the party view was clear (Partisan-1st). We explore both results in turn in the subsequent sections. Potential concerns are addressed in Section 3.2.

Moreover, Fig. 2(b) offers an alternative depiction of the donation outcome. The figure demonstrates distributions of the proportion of the bonus payment donated to the EFF by treatment. By inspecting the histograms for the Partisan-1st and the Partisan-2nd groups, we notice differences in the prevalence of the most common splits — donating nothing, donating 50%, and donating all of the bonus payment. The first one is more likely in the Partisan-1st group, while the two latter splits occur more frequently in the Partisan-2nd group. To obtain more rigorous evidence, we perform the Epps–Singleton characteristic function test of equality of two distributions (Goerg and Kaiser, 2009), which is widely applied in Economics (e.g., Cavallo et al., 2017). The test yields the p -value of 0.016, which means that we can reject the null hypothesis of equal distributions at the 5% significance level.

Lastly, we highlight an important limitation of our donation result for Republicans. When conducting the same experiment with a sample of Democrats, we do not find an effect on donations, which suggests that the ability to preempt polarization on a policy issue by early exposure to balanced non-partisan arguments might not generally apply in all political contexts. We provide a full discussion of possible explanations behind different results for Republicans and Democrats in Section 3.3.

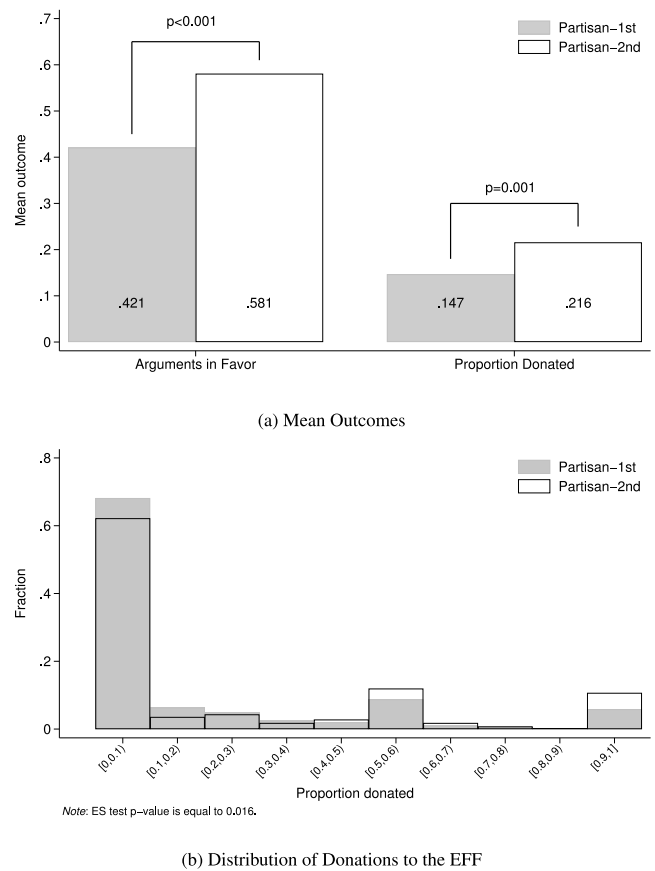


Fig. 2. Summary of main results. Note: Panel A depicts mean values of the main outcomes by treatment group (Partisan-1st and Partisan-2nd). In the Partisan-2nd group, participants (Republicans) see the non-partisan information block on net neutrality (non-political videos on its pros and cons) prior to receiving information about the position of political parties (Republicans against and Democrats in favor). In the Partisan-1st group, the order is reversed. The left bar graph of Panel A pertains to the proportion of participants who reported making an argument in favor of net neutrality after watching the non-partisan videos. We report the p -value for the regression of a dummy equal to one if a person reported an argument in favor on a dummy equal to one if they were assigned the Partisan-2nd treatment. The right bar graph in Panel A focuses on the proportion of the bonus payment donated to the Electronic Frontier Foundation (EFF), a charity supporting net neutrality. We report the p -value for the regression of the proportion of the bonus payment donated to the EFF on a dummy equal to one if the participant was assigned the Partisan-2nd treatment. In all cases, the p -values were computed using robust standard errors. Panel B depicts a histogram of the proportion of the bonus payment donated to the EFF by treatment group with ten bins of equal width (0.1). We report the p -value for the Epps–Singleton characteristic function test of equality of two distributions (denoted ES). The figure is based on a sample of $N = 807$ individuals (411 in the Partisan-1st group and 396 in the Partisan-2nd group) who chose a side of the argument on net neutrality. The proportion of the bonus payment donated to the EFF is reported for $N = 805$ (411 in the Partisan-1st group and 394 in the Partisan-2nd group).

3.1.1. Side of the argument (Intermediate outcome)

In the Partisan-2nd group, for which the videos were the only relevant information, the proportion of participants making an argument in favor of net neutrality was 58.1%. The share was equal to 42.1% in the Partisan-1st group, where the subjects previously learned about Republicans’ opposition to net neutrality (and Democrats’ support for it) from a slide show about the history of the net neutrality debate. The treatment effect size of 16 pp ($p < 0.001$), or 0.32 SD, indicates a strong impact of partisan information on the evaluation of the non-partisan materials about net neutrality, and the support for the policy itself. Overall, one could expect that Republicans’ trust in the judgment of their party can result in them interpreting the non-partisan information differently, relative to the counterfactual situation (that the Partisan-2nd group experienced) of seeing the non-partisan information without

knowing the parties' stances on the issue. While this could explain the direction of the result, its magnitude is notable given that the participants had to provide an argument rather than just declare a side. It shows people's ability to fit an argument to a position ex-ante influenced by the party stance. Ultimately, participants in both groups had access to the same reasons and examples, relating to both sides of the net neutrality debate (pros vs. cons), and managed to find *different arguments* persuasive to a remarkable degree.

Panel A of [Table 2](#) reports the regression analysis associated with the side of the argument chosen by the participants. The treatment effect is robust to including controls (15.7 pp), restricting the sample to those who passed all comprehension checks (16.1 pp), and removing pilot experiment observations (16.3 pp). Columns 5 and 6 provide insights into the effect's strength by whether individuals voted for Donald Trump in 2020, a pre-registered angle of heterogeneity. This gives an opportunity to conduct a simple descriptive moderation analysis as a previous vote for Donald Trump can serve a rough proxy for being a strong Republican. We find that the vote for Trump was associated with a 14 pp increase in the treatment effect (i.e., the effect of being assigned the Partisan-2nd group). This is consistent with the interpretation that moderate voters, who are less likely to back Trump, are less influenced by the party stance. Having said that, it is important to remember that such evidence is correlational — the heterogeneity might potentially be explained by the vote's correlation with other variables, such as the level of education. We hope that our preliminary result will encourage efforts to better understand partisanship's strength as a moderator for persuasiveness of non-partisan information.

3.1.2. Donation (Primary outcome)

We find that, in the Partisan-2nd group, where the partisan information was provided second, the proportion of the bonus payment donated to the EFF was 21.6%. In the Partisan-1st group, where the partisan block was displayed first, the donated share of the bonus equaled 14.7%. Preceding partisan information with two-sided (covering both pros and cons) non-partisan information resulted in the average treatment effect of 6.9 pp ($p = 0.001$), or 0.23 SD. This magnitude is equivalent to a 46.9% increase in the proportion of the bonus given to the EFF. Given that the total amount of the bonus that participants could split between themselves and the EFF was \$0.5, the treatment effect is equivalent to an average rise in donations by 3.45 cents. The result indicates that it is possible to preempt ideological polarization of an issue by non-partisan campaigning, even when both sides of the debate are discussed. Given our usage of materials inducing a strong sense of partisan conflict, it is surprising that the partisan information in the Partisan-2nd group did not undo prior opinion formation associated with analyzing non-partisan information. On the contrary, we report that the side chosen after watching the non-partisan videos persisted.

Panel B of [Table 2](#) offers additional regression specifications pertaining to the primary outcome — the proportion of the bonus payment donated to the EFF. Column 2 demonstrates that the effect is robust to including controls (7.4 pp). Furthermore, Column 3 indicates robustness to restricting the sample to individuals who passed all comprehension checks (8.8 pp). Column 4 shows that the treatment effect (5 pp) remains significant at the 5% level even when the observations from the pilot study are excluded (with $N = 605$). Column 5 illuminates the heterogeneity of the treatment effect with respect to whether an individual voted for Donald Trump in 2020. The point estimates indicate a weaker effect for the Trump voters, but the difference is not statistically significant. The sign is in line with the expectations that more moderate voters are less susceptible to the “party effect”, which could reverse the choice that they made after watching the videos. As we highlighted in the earlier analysis, there are other ways to interpret this type of correlation, with more moderation analysis needed for a better understanding. Lastly, Column 6 demonstrates that the effect

remains significant for the subsample of participants who reported voting for Trump in 2020.

Finally, as a part of exploratory analysis, we consider the extensive margin when investigating the effects of the intervention on donations. [Table 3](#) indicates that viewing the non-partisan information first (Partisan-2nd group) increased the share of positive donations by 6.4 pp, in comparison to the Partisan-1st group where the non-partisan information came second. We also find that the share of substantial donations (greater than 10% of the bonus payment) rose by 8.7 pp. The latter measure may be more accurate, given that we used a slider to elicit the donation amounts. In order to proceed with the survey, participants who wished to donate zero had to at least click at the thumb positioned at zero. For some individuals, it might have been easier to simply move it to a very low amount. We treat the above results as complementary to our main findings, which focus on the intensive margin (as pre-registered).

3.1.3. Mechanisms

A limitation of our study is the inability to disentangle mechanisms driving the impact of our order intervention on donations to the EFF. Our experiment was designed primarily with the policy question in mind, i.e., can early active engagement with non-partisan information inoculate people against incoming partisan signals. In light of the poor effectiveness of depolarization efforts, we focus on the conditions sufficient to prevent a partisan gap on policy support from arising.

The treatment effect is consistent with politically motivated reasoning. At the point of eliciting the donation outcome, participants in both the Partisan-1st and the Partisan-2nd groups know the position of the political parties (the total information received is held constant). However, at the time of processing information on the pros and cons of the policy, only individuals in the Partisan-1st group knew their party stance, which could have led them to differently update their beliefs about the merits of net neutrality compared to those in the Partisan-2nd group. Our result that the party position strongly affects which side of the non-partisan argument the participants indicate as more convincing makes the proposed channel plausible.

However, there are alternative mechanisms that we cannot refute using our current design. If the non-partisan block has a positive effect on the support for net neutrality, the treatment effect on the donation outcome may be driven by a preference for consistency within self.¹⁶ If one chooses their preferred side of the argument after watching the non-partisan videos, their willingness to act consistently with their previous decision may affect their donation choice. One way of testing for this channel, which we leave as a recommendation for future research, would be cross-randomizing the order treatment with whether or not the participants have to indicate which side of the argument they support after receiving the non-partisan information. It is plausible to assume that in the latter case, the preference for consistency is weaker, as participants do not have to express their opinion at the intermediate stage.

Another possibility to consider is the primacy effect — individuals become less elastic to information as they receive it, which means that in the Partisan-1st group, the negative partisan cue has a larger impact on the donation decision than in the Partisan-2nd group, as it was presented at an earlier stage. Our results are inconsistent with the opposite recency effect. Importantly, whether primacy or recency effects dominate may be context-dependent and remains an active area of research. As a suggestion for similar experiments in the future, one way to address the primacy vs. recency issue would be to calibrate the non-partisan block to have a zero average effect on the support for net neutrality. To validate the calibration, one would need to compare the impact of the non-partisan block against a pure control group.

¹⁶ The first condition may not be necessary if a mere act of focusing people's attention on the issue of net neutrality positively affects the support.

Table 2
Regression analysis for the main outcomes.

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Argument in Favor of Net Neutrality						
Partisan-2nd	0.160*** (0.035)	0.157*** (0.035)	0.161*** (0.041)	0.163*** (0.040)	0.055 (0.066)	0.195*** (0.041)
Trump in 2020					-0.195*** (0.054)	
Trump in 2020 × Partisan-2nd					0.140* (0.078)	
Constant	0.421*** (0.024)	0.591*** (0.087)	0.444*** (0.030)	0.415*** (0.028)	0.561*** (0.047)	0.366*** (0.028)
Observations	807	807	572	607	789	568
Controls	No	Yes	No	No	No	No
t-stat (Partisan-2nd)	4.595	4.487	3.887	4.065	0.836	4.743
Panel B: Proportion of the Bonus Donated to the EFF						
Partisan-2nd	0.069*** (0.022)	0.074*** (0.022)	0.088*** (0.025)	0.050** (0.026)	0.091** (0.046)	0.053** (0.024)
Trump in 2020					-0.096*** (0.033)	
Trump in 2020 × Partisan-2nd					-0.038 (0.052)	
Constant	0.147*** (0.014)	0.164*** (0.052)	0.125*** (0.015)	0.171*** (0.017)	0.218*** (0.029)	0.121*** (0.015)
Observations	805	805	572	605	788	567
Controls	No	Yes	No	No	No	No
t-stat (Partisan-2nd)	3.197	3.419	3.538	1.971	1.986	2.217
Sample	All	All	Comp.	No Pilot	All	Trump20

Note: The table reports treatment effects for the main outcomes of the paper. There are two treatment groups: Partisan-1st and Partisan-2nd. In the Partisan-2nd, participants (Republicans) see the non-partisan information block on net neutrality (non-political videos on its pros and cons) prior to receiving information about the position of political parties (Republicans against and Democrats in favor). In the Partisan-1st group, the order is reversed. **Panel A** pertains to the proportion of participants who reported making an argument in favor of net neutrality after watching the non-partisan videos. Column 1 demonstrates a regression of a dummy equal to one if the participant reported making an argument in favor of net neutrality after watching the videos. Column 2 presents the same specification with controls. The following controls were included: a dummy equal to one if a person is male, age, a dummy equal to one if they have at least a 4-year degree, a dummy equal to one if their household income exceeds \$70,000, a dummy equal to one if they are white/Caucasian, the household size capped at 6, and a dummy equal to one if their religion can be classified as Christian. Column 3 shows the regression when the sample is restricted to those who passed all comprehension questions, whereas Column 4 pertains to the specification with all pilot study observations excluded from the sample. Lastly, Column 5 contains a regression of the same dependent variable on the Partisan-2nd dummy, a dummy equal to one if the participant reported voting for Donald Trump in 2020, and their interaction. Column 6 shows the specification from Column 1 but with the sample restricted to those who reported voting for Donald Trump in 2020. **Panel B** reports the same specifications but the dependent variable is the proportion of the bonus payment donated to the Electronic Frontier Foundation (EFF), a charity supporting net neutrality. The main specification (Column 1) is based on a sample of N = 807 (411 in the Partisan-1st group and 396 in the Partisan-2nd group) for Panel A, and N = 805 (411 in the Partisan-1st group and 394 in the Partisan-2nd group) for Panel B. Robust standard errors are parenthesized. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 3
Donations to the EFF: Extensive margin.

	(1) Donation > 0	(2) Donation > 0	(3) Donation > 10%	(4) Donation > 10%
Partisan-2nd	0.064* (0.034)	0.071** (0.034)	0.087*** (0.032)	0.097*** (0.032)
Constant	0.355*** (0.024)		0.260*** (0.022)	
Observations	805	805	805	805
Sample	All	All	All	All
Controls	No	Yes	No	Yes
t-stat (Partisan-2nd)	1.852	2.092	2.701	3.007

Note: Column 1 demonstrates a regression of a dummy variable equal to one if an individual donated a positive amount to the Electronic Frontier Foundation (EFF), a charity supporting net neutrality, on a dummy equal to one if they were assigned the Partisan-2nd treatment. In the Partisan-2nd group, participants (Republicans) see the non-partisan information block on net neutrality (non-political videos on its pros and cons) prior to receiving information about the position of political parties (Republicans against and Democrats in favor). In the Partisan-1st group, the order is reversed. Column 2 shows the same specification but with controls. The following controls were included: a dummy equal to one if a person is male, age, a dummy equal to one if they have at least a 4-year degree, a dummy equal to one if their household income exceeds \$70,000, a dummy equal to one if they are white/Caucasian, the household size capped at 6, and a dummy equal to one if their religion can be classified as Christian. Column 3 depicts a regression of a dummy variable equal to one if an individual donated more than 10% of the bonus payment to the EFF on a dummy equal to one if they were assigned Partisan-2nd treatment. Column 4 shows the same specification with controls. The table is based on a sample of N = 805 (411 in the Partisan-1st group and 394 in the Partisan-2nd group). Robust standard errors are reported in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Lastly, it might be the case that viewing a highly polarizing partisan block created an emotional response that temporarily reduced participants' ability to process the arguments. This could explain lower donation levels in the Partisan-1st group, which processed non-partisan information after receiving the stressful stimulus. However, our data on answers to simple comprehension checks indicates that, in both groups, more than 97% of participants correctly answered all three checks placed in the partisan block. This shows that the block itself is unlikely to affect cognitive ability in a major way.

3.2. Robustness checks

3.2.1. Lee bounds

In Section 2.2, we consider the issue of attrition. The overall completion rate was high (94.2%) and differed by treatment group by 2.7 pp, with 92.9% finishing the survey in the Partisan-2nd condition and 95.6% in the Partisan-1st condition. To address the extent to which our results are affected by attrition during the study, we apply (Lee, 2009) bounds to our treatment effects. The relevant table is provided in the online appendix. The lower bound of the effect remains significant at the 1% level for both main outcomes: (i) the side of the argument chosen after watching the videos, and (ii) the proportion of the bonus donated to the EFF. Additionally, even when excluding all pilot observations and applying Lee bounds, the 95% confidence interval for the treatment effect does not include zero. We conclude that our main results are probably not explained by attrition.

3.2.2. Beliefs

Another concern to consider is an alternative explanation where the order of the partisan and non-partisan blocks differentially affect beliefs about Republicans' and Democrats' support for net neutrality. We provide evidence that immediately after collecting the donation outcome (after everyone received both types of information), the beliefs about own party's support do not vary by treatment group. In particular, Table 4 indicates that in the Partisan-2nd group the estimated proportion of Republicans in favor of the policy was lower than in the Partisan-1st group by just 0.5 pp ($p = 0.756$). On the other hand, we find some evidence that the order of information affected the beliefs about Democrats' support. In the Partisan-2nd group, the estimated fraction of Democrats backing net neutrality was higher by 3 pp. This result is significant at 10% level, though it is not robust to including controls. Importantly, this evidence cannot explain our main effect — donations were higher in the Partisan-2nd group, where the estimated Democrats' support was greater. It is unlikely that, holding own party's support constant, the perception that an issue is more favored by Democrats could have played a role in encouraging donations among Republicans. Lastly, it is insightful to compare constants in the regressions pertaining to the beliefs about Republicans' and Democrats' support. Ex-post, the participants considered the issue as strongly partisan — the average estimated support among Republicans was equal to 34%, with the same statistic equal to 73% for Democrats.

3.2.3. Follow-up survey

When interpreting the result, it is natural to consider the following important concern. The treatment effect can be driven by the experimenter's demand for consistency. While consistency within self is one of the key channels through which the intervention is intended to operate, we cannot immediately distinguish it from participants' desire to appear consistent in the eyes of the experimenter. A related issue is their potential unwillingness to admit that they were swayed by the party position, which they could have considered apparent to the experimenter should they fail to donate following the previous choice of a pro side argument.

To address these concerns, we conducted an obfuscated follow-up study eliciting support for three policies related to technology – net neutrality, blockchain voting, and carbon capture – presented in

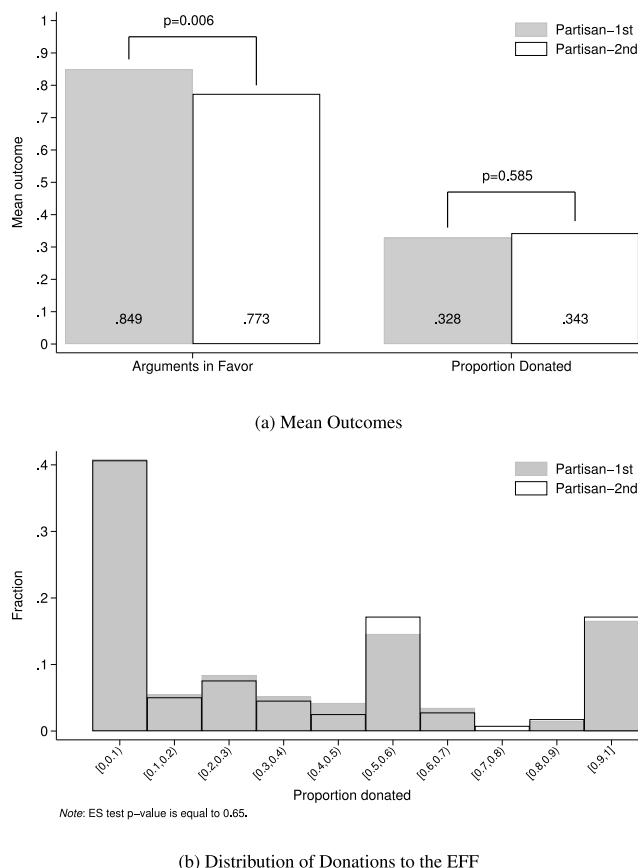


Fig. 3. Results of the additional experiment with Democrats. *Note:* Panel A depicts mean values of the main outcomes by treatment group (Partisan-1st and Partisan-2nd). In the Partisan-2nd group, participants (Democrats) see the non-partisan information block on net neutrality (non-political videos on its pros and cons) prior to receiving information about the position of political parties (Republicans against and Democrats in favor). In the Partisan-1st group, the order is reversed. The left bar graph of Panel A pertains to the proportion of participants who reported making an argument in favor of net neutrality after watching the non-partisan videos. We report the p -value for the regression of a dummy equal to one if a person reported an argument in favor on a dummy equal to one if they were assigned the Partisan-2nd treatment. The right bar graph in Panel A focuses on the proportion of the bonus payment donated to the Electronic Frontier Foundation (EFF), a charity supporting net neutrality. We report the p -value for the regression of the proportion of the bonus payment donated to the EFF on a dummy equal to one if the participant was assigned the Partisan-2nd treatment. In all cases, the p -values were computed using robust standard errors. Panel B depicts a histogram of the proportion of the bonus payment donated to the EFF by treatment group with ten bins of equal width (0.1). We report the p -value for the Epps–Singleton characteristic function test of equality of two distributions (denoted ES). The figure is based on a sample of $N = 802$ individuals (405 in the Partisan-1st group and 397 in the Partisan-2nd group) who chose a side of the argument on net neutrality. The proportion of the bonus payment donated to the EFF is reported for $N = 801$ (405 in the Partisan-1st group and 396 in the Partisan-2nd group).

random order.¹⁷ We used a different type of outcome variable, a 0–100 support scale, rather than donations, to further dissociate ourselves from the original survey. Haaland et al. (2023) and Haaland and Roth (2020) highlight merits of utilizing obfuscated follow-up surveys to address experimenter demand effects. If the treatment is only applied in the original study and the follow-up is of a substantially different format to create an impression that the two surveys are unrelated, the outcomes measured in the follow-up study should not be affected by experimenter demand in a way correlated with treatment. Our framing of the follow-up survey (technology policies with placebo items), varying the type of outcome used, as well as substantial time elapsed

¹⁷ The follow-up survey questions are provided in the online appendix.

Table 4
Ex-Post beliefs about party support for net neutrality.

	(1) Beliefs R Support	(2) Beliefs R Support	(3) Beliefs D Support	(4) Beliefs D Support
Partisan-2nd	-0.497 (1.598)	-0.300 (1.604)	2.954* (1.542)	2.457 (1.527)
Constant	33.942*** (1.156)		72.572*** (1.162)	
Observations	805	805	805	805
Sample	All	All	All	All
Controls	No	Yes	No	Yes

Note: Column 1 shows a regression of participants' estimated proportion of Republicans who support net neutrality on a dummy equal to one if they were assigned the Partisan-2nd treatment. In the Partisan-2nd group, participants (Republicans) see the non-partisan information block on net neutrality (non-political videos on its pros and cons) prior to receiving information about the position of political parties (Republicans against and Democrats in favor). In the Partisan-1st group, the order is reversed. The estimates were collected immediately after the main outcome — donation to the Electronic Frontier Foundation (EFF), a charity supporting net neutrality. Column 2 depicts the same specification with controls. The following controls were included: a dummy equal to one if a person is male, age, a dummy equal to one if they have at least a 4-year degree, a dummy equal to one if their household income exceeds \$70,000, a dummy equal to one if they are white/Caucasian, the household size capped at 6, and a dummy equal to one if their religion can be classified as Christian. Column 3 demonstrates a regression of participants' estimated proportion of Democrats who support net neutrality on a dummy equal to one if they were assigned the Partisan-2nd treatment. Column 4 presents the same specification with controls. The table is based on a sample of N = 805 (411 in the Partisan-1st group and 394 in the Partisan-2nd group). Robust standard errors are reported in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

between the two surveys, ensure that we maximize the benefits of the obfuscation methodology.

Even though the median person took the follow-up survey 30 days after the original study, we find evidence that the treatment effect persists, with the support for net neutrality higher in the Partisan-2nd group. We report null effects in the case of support for the two placebo policies. Panel A of Table 5 summarizes the regression results. In particular, the average support for net neutrality in the Partisan-1st group was equal to 66 out of 100. The score was higher in the Partisan-2nd group by 4 points ($p = 0.085$), which indicates that a significant part of the original treatment effect persisted. The standardized effect size equals 0.13 SD, which corresponds to a 44% drop from the treatment effect on the donation outcome (0.23 SD). The result is robust to including controls ($p = 0.05$). We recorded no significant treatment effects on the support for blockchain voting (-1.2 points, $p = 0.577$) and carbon capture (-2.8 points, $p = 0.201$).

Moreover, Panels B and C of Table 5 indicate that the beliefs about partisan support (Panel B – Republicans and Panel C – Democrats) for the three issues (net neutrality, blockchain voting, and carbon capture), elicited after we collected individual agreement with the policies, did not vary by treatment group.

The take-up rate of the follow-up survey – among those for whom we have a donation outcome – was equal to 85.5% (83.2% in the Partisan-2nd group and 87.8% in the Partisan-1st group). The follow-up sample is well-balanced, with no significant differences for any of the covariates (the balance table is provided in the online appendix).

To sum up, the results of the follow-up survey feature as a robustness check to verify that the main effect of the intervention on donations is unlikely to be explained away by the experimenter demand effects or people's willingness to appear unaffected by the party stance. Having said that, we believe that our follow-up results may serve as an interesting example useful for the nascent literature exploring persistence of persuasion interventions.

3.3. Additional experiment with Democrats

After learning the results of the main experiment (with a sample of N = 851 Republicans), we pre-registered and conducted an additional experiment with a sample of N = 841 Democrats. We used exactly the same protocol in both experiments to ensure that the results are directly comparable. The goal of this extension is to shed light on the overall robustness of the results and investigate potential heterogeneity of the treatment effect by party affiliation. Overall, 422 individuals in the Partisan-1st group and 419 individuals in the Partisan-2nd group

were assigned treatment, with 405 and 396 individuals completing the survey in each group respectively. We do not detect significant differential attrition. Furthermore, the sample is well-balanced on covariates (the balance table is provided in the online appendix).

Fig. 3 provides a summary of the main results. Fig. 3(a) depicts mean values of the outcomes by treatment group. In the Partisan-2nd group, the share of individuals in favor of net neutrality was higher than in the Partisan-1st group by 7.6 pp ($p = 0.006$). Thus, similarly to Republicans, the side of the argument chosen by participants is significantly impacted by the party position. The direction of the effect indicates that learning about one's party support for net neutrality increases the likelihood of reporting pro arguments in the non-partisan videos as more convincing than the arguments against. The magnitude of the treatment effect is smaller than for Republicans (7.6 pp for Democrats and 16 pp for Republicans). This may be due to the fact that in the absence of information on party preferences, the support for net neutrality among Democrats is already as high as 77.3%. The corresponding quantity for Republicans is 58.1%. Thus, there are fewer people who can be persuaded by a favorable party signal in the case of Democrats than those who can be persuaded by an unfavorable party signal in the case of Republicans.

Second, we report that, unlike for Republicans, the order of partisan and non-partisan information does not have a significant effect on the support for net neutrality, measured through a donation to the EFF. The proportion of the bonus payment donated to the charity only differed by 1.5 pp ($p = 0.585$) between the treatment groups, which is equivalent to 0.75 cents. Moreover, Fig. 3(b) shows distributions of the proportion of the bonus payment donated to the EFF by treatment. Both inspecting the histograms and conducting the Epps–Singleton test of equality of two distributions make it clear that there is no significant difference in the distributions by treatment. This finding might suggest that whether the party position is *favorable* or *unfavorable* is critical. Even with a higher support for the policy, learning that one's party also supports it, crowds out the need to provide a donation to the EFF. This may stem from the belief that the advocacy will be successfully led by the political party. This counter-force is only present in the case of a favorable party signal, which may explain the difference between the results of the study with Republicans and the study with Democrats. However, there are multiple other possible explanations. First, it might be the case that Democrats and Republicans fundamentally differ in their susceptibility to the order intervention. In particular, the cost of going against partisan signals may be higher for Democrats, which could lead to lower effectiveness of preempting polarization by displaying balanced non-partisan arguments. Furthermore, it is possible that we were simply less powered to detect the effect of the order

Table 5
Support for net neutrality in the follow-up survey.

	(1)	(2)	(3)	(4)
	Net Neutr.	Net Neutr.	Block.	Carbon
Panel A: Support for the Three Policies				
Partisan-2nd	3.993* (2.315)	4.602** (2.340)	-1.264 (2.267)	-2.837 (2.217)
Constant	66.263*** (1.654)	70.151*** (8.934)	67.986*** (1.599)	62.288*** (1.510)
Observations	689	689	689	689
Controls	No	Yes	No	No
Panel B: Beliefs about Republicans' Support				
Partisan-2nd	2.566 (1.952)	3.069 (1.945)	1.812 (2.197)	-1.102 (1.932)
Constant	49.972*** (1.396)	41.891*** (6.960)	55.778*** (1.525)	47.105*** (1.354)
Observations	686	686	686	687
Controls	No	Yes	No	No
Panel C: Beliefs about Democrats' Support				
Partisan-2nd	-1.037 (1.850)	-1.783 (1.840)	-2.162 (2.039)	1.035 (1.780)
Constant	67.474*** (1.247)	68.631*** (6.767)	49.429*** (1.416)	70.551*** (1.245)
Observations	686	686	686	687
Controls	No	Yes	No	No

Note: The table reports treatment effects for outcomes collected in the obfuscated follow-up survey. There are two treatment groups: Partisan-1st and Partisan-2nd. In the Partisan-2nd, participants (Republicans) see the non-partisan information block on net neutrality (non-political videos on its pros and cons) prior to receiving information about the position of political parties (Republicans against and Democrats in favor). In the Partisan-1st group, the order is reversed. The follow-up survey was conducted a few weeks after the survey with the information intervention. **Panel A** focuses on the support (on a scale from 0 to 100) for three technology-related policies: net neutrality, blockchain voting, and carbon capture. Column 1 demonstrates a regression of the support for net neutrality in the follow-up study on a dummy equal to one if they were assigned the Partisan-2nd treatment. Column 2 presents the same specification with controls. The following controls were included: a dummy equal to one if a person is male, age, a dummy equal to one if they have at least a 4-year degree, a dummy equal to one if their household income exceeds \$70,000, a dummy equal to one if they are white/Caucasian, the household size capped at 6, a dummy equal to one if their religion can be classified as Christian, and the time elapsed between the first survey and the follow-up survey. **Panel B** reports the same specifications but the dependent variable is the estimated proportion of Republicans supporting each policy (net neutrality for Columns 1–2, blockchain voting for Column 3, and carbon capture for Column 4). **Panel C** reports the same specifications but the dependent variable is the estimated proportion of Democrats supporting each policy. The table (Panel A specifically) is based on a sample of N = 689 individuals (361 in the Partisan-1st group and 328 in the Partisan-2nd group). Robust standard errors are parenthesized. * significant at 10%; ** significant at 5%; *** significant at 1%.

intervention on donations. This stems from the fact that the base level support for net neutrality was higher for Democrats (about 80%) than for Republicans (about 50%). Partisan signals are likely to be less effective in raising further support among Democrats above the high base level than in reducing it from around 50% level among Republicans. However, as a consequence, this gives less room to detect effects of the order intervention, which attempts to inoculate individuals against an incoming partisan signal.

We now proceed to discuss the associated regression results. Panel A of **Table 6** focuses on the side of the argument chosen by the participants following the non-partisan videos. The results in Column 1 indicate that, for the sample of Democrats, knowing party positions (Democratic party in favor of and Republican party against net neutrality) when choosing the side of the argument increases the likelihood of being in favor by 7.6 pp (p = 0.006). The treatment effect is robust to including controls (7.5 pp, p = 0.007) and restricting the sample to those who passed all comprehension checks (5.4 pp, p = 0.079).

Panel B of **Table 6** offers additional regression specifications pertaining to the primary outcome — the proportion of the bonus payment

Table 6
Regression analysis for the main outcomes (study with Democrats).

	(1)	(2)	(3)
Panel A: Argument in Favor of Net Neutrality			
Partisan-2nd	-0.076*** (0.028)	-0.075*** (0.028)	-0.054* (0.031)
Constant	0.849*** (0.018)	0.942*** (0.061)	0.862*** (0.020)
Observations	802	802	592
Controls	No	Yes	No
t-stat (Partisan-2nd)	-2.761	-2.717	-1.759
Panel B: Proportion of the Bonus Donated to the EFF			
Partisan-2nd	0.014 (0.026)	0.020 (0.025)	0.015 (0.030)
Constant	0.328*** (0.018)	0.127** (0.055)	0.308*** (0.020)
Observations	801	801	592
Controls	No	Yes	No
t-stat (Partisan-2nd)	0.546	0.773	0.487
Sample	All	All	Comp.

Note: The table reports treatment effects for the main outcomes of the paper. There are two treatment groups: Partisan-1st and Partisan-2nd. In the Partisan-2nd, participants (Democrats) see the non-partisan information block on net neutrality (non-political videos on its pros and cons) prior to receiving information about the position of political parties (Republicans against and Democrats in favor). In the Partisan-1st group, the order is reversed. **Panel A** pertains to the proportion of participants who reported making an argument in favor of net neutrality after watching the non-partisan videos. Column 1 demonstrates a regression of a dummy equal to one if the participant reported making an argument in favor of net neutrality after watching the videos. Column 2 presents the same specification with controls. The following controls were included: a dummy equal to one if a person is male, age, a dummy equal to one if they have at least a 4-year degree, a dummy equal to one if their household income exceeds \$70,000, a dummy equal to one if they are white/Caucasian, the household size capped at 6, and a dummy equal to one if their religion can be classified as Christian. Column 3 shows the regression when the sample is restricted to those who passed all comprehension questions, whereas **Panel B** reports the same specifications but the dependent variable is the proportion of the bonus payment donated to the Electronic Frontier Foundation (EFF), a charity supporting net neutrality. The main specification (Column 1) is based on a sample of N = 802 (405 in the Partisan-1st group and 397 in the Partisan-2nd group) for Panel A, and N = 801 (405 in the Partisan-1st group and 396 in the Partisan-2nd group) for Panel B. Robust standard errors are parenthesized. * significant at 10%; ** significant at 5%; *** significant at 1%.

donated to the EFF. The results in Column 1 show that, for the sample of Democrats, seeing the non-partisan information first (Partisan-2nd group) insignificantly increases the proportion of the bonus donated to the EFF (1.4 pp, p = 0.585). Column 2 demonstrates that the treatment effect remains insignificant when including controls (2 pp, p = 0.440), while Column 3 indicates the same when restricting the sample to individuals who passed all comprehension checks (1.5 pp, p = 0.626). A further discussion of results of the additional experiment, mostly based on regression analysis similar to the one reported for the main experiment, is relegated to the online appendix.

4. Conclusion

We conducted an experiment with Republicans to measure the extent to which providing non-partisan information and encouraging individuals to form their opinion before being exposed to partisan information can decrease the polarizing effect that political parties have on individual opinion formation. We find that changing the order in which partisan information (underscoring own party's opposition and rival party's support) and non-partisan information (where the participants evaluate factual arguments about the pros and cons of the policy) were displayed significantly affected an incentivized measure of support for the policy. In particular, we report that those who saw the non-partisan information first donated 46% more to a charity advocating for net neutrality — our chosen issue. This demonstrates that formation of a partisan gap, when it has not yet arisen, can be

preempted by encouraging prior active engagement with non-partisan information.

Important caveats apply. Our intervention involved several characteristics that we have shown as sufficient for preempting a partisan gap: early provision of non-partisan information, offering both arguments and counterarguments, and active engagement with provided content in a way that leads to formation of a well-informed personal opinion. We hope that future studies will focus on identifying a subset of these conditions which are necessary to achieve the intervention's objective. This will further inform real-world applications, such as optimal campaigning strategies by NGOs and issue advocacy groups. Secondly, the results of the additional experiment with a sample of Democrats indicate a null effect of the order intervention, which may suggest that its effectiveness is heterogeneous by partisan context. More work is needed to test if the results are generalizable to political parties other than Republicans, and if they apply outside U.S. politics. Moreover, an important limitation of our study is that the experiment was conducted in an artificial setting with the main outcome variable – the proportion of the bonus payment donated to the EFF – being low-stakes. We encourage future research to investigate whether the results hold in high-stakes environments and in natural settings. Lastly, more work is needed to better understand specific mechanisms that enable preempting polarization on a policy issue. Our experiment cannot directly distinguish between mechanisms such as motivated reasoning and preference for consistency within self. We hope that subsequent studies will be able to explicitly disentangle them (Section 3.1.3 suggests some ways forward).

Declaration of competing interest

none

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.jpubecon.2024.105122>.

References

- Aaroe, L., 2012. When citizens go against elite directions: partisan cues and contrast effects on citizens' attitudes. *Party Polit.* 18 (2), 215–233.
- Afrouzi, H., Arteaga, C., Weisburst, E.K., 2023. Is it the Message or the Messenger? Examining Movement in Immigration Beliefs. Technical Report, National Bureau of Economic Research.
- Akesson, J., Hahn, R.W., Metcalfe, R.D., Rasooly, I., 2022. Race and Redistribution in the United States: An Experimental Analysis. Technical Report, National Bureau of Economic Research.
- Arpan, L.M., Xu, X., Raney, A.A., Chen, C.-f., Wang, Z., 2018. Politics, values, and morals: Assessing consumer responses to the framing of residential renewable energy in the United States. *Energy Res. Soc. Sci.* 46, 321–331.
- Babcock, L., Loewenstein, G., Issacharoff, S., Camerer, C., 1995. Biased judgments of fairness in bargaining. *Am. Econ. Rev.* 85 (5), 1337–1343.
- Bae, M., 2021. The effects of message order on emotions and donation intention in charity advertising: the mediating roles of negative and positive empathy. *J. Mark. Commun.* 1–18.
- Barden, J., Petty, R.E., 2008. The mere perception of elaboration creates attitude certainty: Exploring the thoughtfulness heuristic. *J. Personal. Soc. Psychol.* 95 (3), 489.
- Baysan, C., 2021. Persistent polarizing effects of persuasion: Experimental evidence from turkey. *Amer. Econ. Rev.*
- Bechtel, M.M., Hainmueller, J., Hangartner, D., Helbling, M., 2015. Reality bites: The limits of framing effects for salient and contested policy issues. *Polit. Sci. Res. Methods* 3 (3), 683–695.
- Bénabou, R., Tirole, J., 2011. Identity, morals, and taboos: Beliefs as assets. *Q. J. Econ.* 126 (2).
- Bisgaard, M., 2019. How getting the facts right can fuel partisan-motivated reasoning. *Am. J. Polit. Sci.* 63 (4), 824–839.

- Bolsen, T., Druckman, J.N., Cook, F.L., 2014. The influence of partisan motivated reasoning on public opinion. *Polit. Behav.* 36 (2), 235–262.
- Bositis, D.A., Baer, D.L., Miller, R.E., 1985. Cognitive information levels, voter preferences, and local partisan political activity: A field experimental study on the effects of timing and order of message presentation. *Polit. Behav.* 7 (3), 266–284.
- Brashier, N.M., Pennycook, G., Berinsky, A.J., Rand, D.G., 2021. Timing matters when correcting fake news. *Proc. Natl. Acad. Sci.* 118 (5), e2020043118.
- Bursztyjn, L., Rao, A., Roth, C., Yanagizawa-Drott, D., 2023. Opinions as Facts. *Rev. Econ. Stud.* 90 (4), 1832–1864.
- Cavallo, A., Cruces, G., Perez-Truglia, R., 2017. Inflation expectations, learning, and supermarket prices: Evidence from survey experiments. *Am. Econ. J.: Macroecon.* 9 (3), 1–35. <http://dx.doi.org/10.1257/mac.20150147>.
- Chen, J.Y., 2022. Biased learning under ambiguous information. *J. Econom. Theory* 105492.
- Cohen, G.L., 2003. Party over policy: The dominating impact of group influence on political beliefs. *J. Personal. Soc. Psychol.* 85 (5), 808.
- Conlon, J.J., Mani, M., Rao, G., Ridley, M.W., Schilbach, F., 2022. Not Learning from Others. Technical Report, National Bureau of Economic Research.
- Di Tella, R., Perez-Truglia, R., Babino, A., Sigman, M., 2015. Conveniently upset: Avoiding altruism by distorting beliefs about others' altruism. *Amer. Econ. Rev.* 105 (11), 3416–3442.
- Druckman, J.N., Peterson, E., Slothuus, R., 2013. How elite partisan polarization affects public opinion formation. *Am. Polit. Sci. Rev.* 107 (1), 57–79.
- Eyal, P., David, R., Andrew, G., Zak, E., Ekaterina, D., 2021. Data quality of platforms and panels for online behavioral research. *Behav. Res. Methods* 1–20.
- Falk, A., Zimmermann, F., 2018. Information processing and commitment. *Econ. J.* 128 (613), 1983–2002.
- Gneezy, U., Saccardo, S., Serra-Garcia, M., van Veldhuizen, R., 2020. Bribing the self. *Games Econ. Behav.* 120, 311–324.
- Goerg, S.J., Kaiser, J., 2009. Nonparametric testing of distributions—the Epps-Singleton two-sample test using the empirical characteristic function. *Stata J.* 9 (3), 454–465.
- Gollwitzer, A., Martel, C., Brady, W.J., Pärnamets, P., Freedman, I.G., Knowles, E.D., Van Bavel, J.J., 2020. Partisan differences in physical distancing are linked to health outcomes during the COVID-19 pandemic. *Nat. Hum. Behav.* 4 (11), 1186–1197.
- Haaland, I., Roth, C., 2020. Labor market concerns and support for immigration. *J. Public Econ.* 191, 104256.
- Haaland, I., Roth, C., Wohlfart, J., 2023. Designing information provision experiments. *J. Econ. Lit.* 61 (1), 3–40.
- Haugtvedt, C.P., Wegener, D.T., 1994. Message order effects in persuasion: An attitude strength perspective. *J. Consum. Res.* 21 (1), 205–218.
- Joslyn, S., Demnitz, R., 2021. Explaining how long CO2 stays in the atmosphere: Does it change attitudes toward climate change? *J. Exp. Psychol.: Appl.* 27 (3), 473.
- Kahan, D.M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L.L., Braman, D., Mandel, G., 2012. The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nat. Clim. Change* 2 (10), 732–735.
- Lee, D.S., 2009. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Rev. Econom. Stud.* 76 (3), 1071–1102.
- Lind, T., Erlandsson, A., Västfjäll, D., Tinghög, G., 2022. Motivated reasoning when assessing the effects of refugee intake. *Behav. Public Policy* 6 (2), 213–236.
- Long, D., Liu, H., Nayga, R.M., 2023. Polarization in environmental donations—application to deforestation prevention donation. *Land Econom.* 99 (1), 122–140.
- Lord, C.G., Ross, L., Lepper, M.R., 1979. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *J. Personal. Soc. Psychol.* 37 (11), 2098.
- Mullainathan, S., Washington, E., 2009. Sticking with your vote: Cognitive dissonance and political attitudes. *Am. Econ. J.: Appl. Econ.* 1 (1), 86–111.
- Nyhan, B., Reifler, J., 2010. When corrections fail: The persistence of political misperceptions. *Polit. Behav.* 32 (2), 303–330.
- Satherley, N., Yogeeswaran, K., Osborne, D., Sibley, C.G., 2018. If they say “yes,” we say “no”: Partisan cues increase polarization over national symbols. *Psychol. Sci.* 29 (12), 1996–2009.
- Schaffner, B.F., Roche, C., 2016. Misinformation and motivated reasoning: Responses to economic news in a politicized environment. *Public Opin. Q.* 81 (1), 86–110.
- Schuldt, J.P., Konrath, S.H., Schwarz, N., 2011. “Global warming” or “climate change”? Whether the planet is warming depends on question wording. *Public Opin. Q.* 75 (1), 115–124.
- Schwardmann, P., Tripodi, E., Van der Weele, J.J., 2022. Self-persuasion: Evidence from field experiments at international debating competitions. *Amer. Econ. Rev.* 112 (4), 1118–1146.
- Shachar, R., 2003. Party loyalty as habit formation. *J. Appl. Econometrics* 18 (3), 251–269.
- Singh, S.P., Swanson, M., 2017. How issue frames shape beliefs about the importance of climate change policy across ideological and partisan groups. *PLoS One* 12 (7), e0181401.
- Strandberg, T., Sívén, D., Hall, L., Johansson, P., Pärnamets, P., 2018. False beliefs and confabulation can lead to lasting changes in political attitudes. *J. Exp. Psychol. [Gen.]* 147 (9), 1382.
- Thaler, M., 2023. The supply of motivated beliefs. Available at SSRN, <https://ssrn.com/abstract=4494688>.