Social learning and partisan bias in the interpretation of climate trends

Douglas Guilbeault[a], Joshua Becker[a], and Damon Centola[a,b,1]

[a]The Annenberg School for Communication, The University of Pennsylvania, Philadelphia, PA 19104; and [b]School of Engineering, The University of Pennsylvania, Philadelphia, PA 19104

Vital scientific communications are frequently misinterpreted by the lay public as a result of motivated reasoning, where people misconstrue data to fit their political and psychological biases. In the case of climate change, some people have been found to systematically misinterpret climate data in ways that conflict with the intended message of climate scientists. While prior studies have attempted to reduce motivated reasoning through bipartisan communication networks, these networks have also been found to exacerbate bias. Popular theories hold that bipartisan networks amplify bias by exposing people to opposing beliefs. These theories are in tension with collective intelligence research, which shows that exchanging beliefs in social networks can facilitate social learning, thereby improving individual and group judgments. However, prior experiments in collective intelligence have relied almost exclusively on neutral questions that do not engage motivated reasoning. Using Amazon's Mechanical Turk, we conducted an online experiment to test how bipartisan social networks can influence subjects' interpretation of climate communications from NASA. Here, we show that exposure to opposing beliefs in structured bipartisan social networks substantially improved the accuracy of judgments among both conservatives and liberals, eliminating belief polarization. However, we also find that social learning can be reduced, and belief polarization maintained, as a result of partisan priming. We find that increasing the salience of partisanship during communication, both through exposure to the logos of political parties and through exposure to the political identities of network peers, can significantly reduce social learning.

social networks | collective intelligence | motivated reasoning | polarization | science communication

Public misunderstanding of climate change is widespread (1–4). One reason for this is that climate data are noisy, and thus vulnerable to political and psychological bias (3–7). For instance, one well-known source of misinterpretation is "endpoint bias" (6, 7), in which recent fluctuations in longitudinal data are assumed to predict future trends better than long-term patterns. This tendency is stronger among politically conservative individuals, for whom motivated reasoning has been found to exacerbate the effects of endpoint bias in the interpretation of climate data (6–9). An important challenge for science communication is that this bias is often compounded by social network effects. Science communications are typically filtered through peer-to-peer social networks, both in face-to-face interactions and on social media (10–12). These social networks not only function as pathways for diffusing media communications, but also help to shape how people interpret these communications (13, 14). It is especially concerning that conservatives and liberals increasingly discuss climate change within politically homogeneous "echo chambers," where partisan bias is reinforced through repeated interactions among like-minded peers (15–17).

To counteract the biasing effects of echo chambers, studies have adopted network-based approaches for reducing belief polarization, particularly through the use of bipartisan communication networks containing equal numbers of conservatives and liberals (18–21). These structured network interventions are based on the theory that exposure to diverse beliefs enriches political knowledge and facilitates bipartisan agreement (22–25). However, studies on the effects of bipartisan communication networks show strikingly inconsistent results (18–21). While some studies find that bipartisan communication networks can enhance both political participation and the quality of deliberation (18, 21, 23, 24), other studies show that they can backfire by causing people to reinforce their existing political biases (19, 20, 25, 26).

Popular theories hold that bipartisan communication networks can fail to reduce bias by activating "biased assimilation" (27–30), a process in which people defend and reinforce their existing views when confronted by opposing beliefs (25, 26, 31). These theories are in tension with a large body of collective intelligence research, which shows that exchanging beliefs in social networks can facilitate social learning, thereby improving both consensus and accuracy among individual and group judgments (32–35). The limiting factor in prior experiments on collective intelligence in group estimation is that previous tasks rely primarily on neutral questions that do not engage motivated reasoning, raising the question of whether belief exchange in bipartisan networks can indeed facilitate social learning about politicized topics (32–37). In support of this social learning hypothesis, recent studies have begun to challenge "backfire effects" in exposure to political beliefs by showing that both conservatives and liberals are surprisingly capable of accurately interpreting peer arguments and media content about partisan topics (38–40).

This paper builds on prior work in collective intelligence and social learning (32–35) to test the hypothesis that exposure to opposing beliefs in structured bipartisan networks can facilitate

---

Significance

Scientific communications about climate change are frequently misinterpreted due to motivated reasoning, which leads some people to misconstrue climate data in ways that conflict with the intended message of climate scientists. Attempts to reduce partisan bias through bipartisan communication networks have found that exposure to diverse political views can exacerbate bias. Here, we find that belief exchange in structured bipartisan networks can significantly improve the ability of both conservatives and liberals to interpret climate data, eliminating belief polarization. We also find that social learning can be reduced, and polarization maintained, when the salience of partisanship is increased, either through exposure to the logos of political parties or through exposure to political identity markers.

---

social learning that can reduce (and even eliminate) partisan bias in the interpretation of climate trends. Our main contention is that previous experiments on bipartisan communication have been unable to clearly identify the effects of bipartisan networks on reducing partisan bias because they have had limited ability to control for the salience of partisanship during people's social interactions. For instance, research on priming effects shows that exposure to minimal partisan cues, such as party logos, can activate partisan biases in response to novel information (41–43). Moreover, recent public opinion research suggests that biased assimilation is not driven by contact with opposing beliefs but rather by political identity signaling, in which confrontation with individuals who belong to an opposing partisan group strengthens subjects' partisan biases (44–48). A key limitation of prior experiments on political polarization and belief exchange in bipartisan networks is that the effects of partisan priming and political identity cannot be excluded from face-to-face discussion groups, in which people are exposed to opposing views and partisan cues simultaneously (18–20, 25, 26). Prior studies have thus been unable to determine whether bipartisan networks fail to reduce partisan bias as a result of the effects of participants' exposure to opposing beliefs or as a result of their exposure to partisan cues that prime their political bias.

Here, we report the results of an online experiment, where a web interface was used to control the salience of partisanship while conservatives and liberals exchanged interpretations of climate data in structured, bipartisan social networks. Based on earlier work, we predicted that conservatives would exhibit significant endpoint bias at baseline when interpreting climate data, creating the potential for increased polarization through biased assimilation (6, 7). However, we also hypothesized that structured bipartisan social networks could reduce belief polarization in the interpretation of climate data through social learning. Importantly, our theory of social learning suggests that exposure to opposing viewpoints does not generate polarization but rather improves individual and collective judgments among both liberals and conservatives, eliminating belief polarization. Nevertheless, we also expected that these social learning effects could be impeded by increasing the salience of political partisanship during participants' interactions. The results show that there were strong effects of partisan priming on reducing social learning. We found significant effects of both (*i*) revealing the political identities of players in the network and (*ii*) minimal priming, via exposure to party logos, on limiting social learning and sustaining belief polarization in bipartisan social networks.

Experimental Design

Two thousand four hundred unique participants were recruited from Amazon's Mechanical Turk for this study. In each trial, subjects were randomized into one of four experimental conditions: (*i*) a control group of participants with the same political ideology, (*ii*) a structured social network in which an equal number of conservatives and liberals were shown the average estimate of their network "neighbors" in the absence of any partisan cues, (*iii*) a structured social network in which an equal number of conservatives and liberals were shown the average estimates of their network neighbors while being exposed to the logos of the Democratic and Republican parties, and (*iv*) a structured social network in which an equal number of conservatives and liberals were shown the average estimate of their network neighbors along with information about the political identity of each of their neighbors (all conditions are shown in *SI Appendix*, Fig. S2).

In each trial, each condition contained 40 individuals, such that each experimental trial contained 200 individuals. We conducted 12 independent trials of this design. (Additional robustness trials using an additional 1,077 subjects were conducted to test the effects of homogeneous echo chambers on social learning

and to identify any differences between showing subjects their neighbors' individual answers versus showing them the average of their neighbors' answers. All results were consistent with our main findings, as discussed below.)

In the control condition, subjects were isolated and not embedded in social networks. In the network conditions, subjects were randomly assigned to a single location in a decentralized social network with a uniform degree distribution, in which every subject had four network neighbors (*SI Appendix*). Using a uniform degree distribution ensured that no one had greater power over the communication dynamics of the network (35). Maintaining the same topology across network conditions allowed us to isolate the effects of partisan cues on the exchange of estimates by holding the structure of the communication networks constant across conditions.

In all trials, subjects were presented with NASA's graph (Fig. 1) and asked to forecast the amount of Arctic Sea ice in 2025. To identify endpoint bias in participants' interpretations of NASA's climate data, we considered whether participants' estimates corresponded to the correct trend identified by NASA, that is, above or below the final point on the graph. In this case, the correct trend is downward from the endpoint in the graph (6, 7).

In every condition, subjects were given three rounds to provide estimates. In round 1, all subjects in every condition provided an independent estimate using the same interface. In round 2 and round 3, control subjects were permitted to revise their answers using only independent reflection, without any information about their peers or their judgments. In the networks without partisan cues, subjects were only shown the average answers of their network neighbors, while they were permitted to revise their responses. For rounds 2 and 3 in the networks with minimal partisan priming, the logos of the Democratic and Republican parties appeared below the group average on the interface. Finally, in the network condition with political identity revealed, subjects were shown the usernames and political ideologies of the four peers connected to them in the social network (*Materials and Methods*). With this design, any differences in how liberals and conservatives interpreted the climate data between the control and network conditions can be attributed to their exposure to social information. Moreover, because peer estimates were presented in an identical fashion in all three network
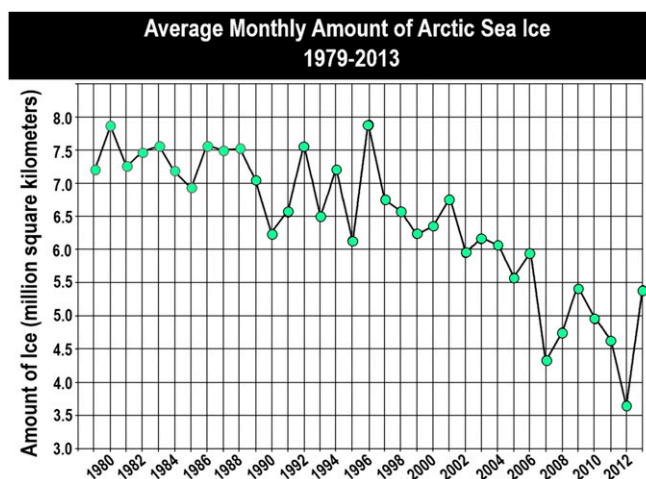

Fig. 1. NASA graph used as a stimulus in the online experiment. This graph was adapted from NASA's 2013 public communications about climate change. The graph has been found to produce misinterpretations about the scientific information it communicates because its final data point indicates an increase in the amount of Artic Sea ice in the opposite direction of the overall trend that NASA intended to communicate. Data from ref. 6.

conditions (*SI Appendix*, Fig. S2), any differences in the interpretation of the climate data across network conditions can be attributed to the effects of the displayed partisan cues.

For each trial, we measured the change in subjects' accuracy between round 1 and round 3 in terms of the percentage of subjects in each condition who predicted the correct trend in the climate data. We refer to this as "trend accuracy," which is the primary variable of interest for climate communications (6–8) and the primary measure of interest here. This is distinct from "point estimate accuracy," which reflects precise correspondence with NASA's projections. (Additional analyses for point estimate accuracy are provided in the *SI Appendix* and are consistent with the results for trend accuracy.) According to this measure, an increase (or decrease) in trend accuracy occurs if the percentage of subjects who predicted the correct trend at round 3 was higher (or lower) than at round 1. We compared the changes in trend accuracy across each of the experimental conditions.

Results Overview

We begin our analysis by confirming that at baseline (round 1), conservatives exhibited significant bias in their predictions and were significantly less likely to correctly interpret the climate change trend (6–9). After two rounds of revision, each of the network conditions showed significantly different outcomes. In structured bipartisan networks without partisan cues, we find that participants exhibited strong social learning effects. In these networks, both liberals and conservatives showed significant improvements in their trend accuracy, resulting in the elimination of partisan bias in the interpretation of the climate data. In bipartisan networks with political identity markers, exposure to opposing beliefs still facilitated significant improvements in trend accuracy among subjects, although social learning in this condition was reduced, and moderate polarization was maintained. Finally, in bipartisan networks in which subjects were exposed to party logos during communication, social learning was prevented, and baseline levels of polarization were maintained.

Baseline Partisan Bias in the Interpretation of Climate Trends.
To begin our analyses, we measured baseline differences in trend accuracy among liberals and conservatives by identifying how many subjects reported the correct trend in round 1. All subjects in round 1, regardless of experimental condition, provided an independent estimate. Thus, for this baseline measurement, all subjects constituted independent observations.

In round 1, conservatives were significantly less likely to interpret NASA's graph correctly. Fig. 2 shows that 60.8% of conservative subjects estimated the correct trend at baseline, compared with 73.9% of liberal subjects ($n = 2,164$; $P < 0.001$, χ^2 test). This result is consistent with previous findings that conservative political alignments are correlated with a significant bias toward misinterpreting climate change data (2–7).

Exposure to Opposing Beliefs in Bipartisan Social Networks Facilitated Social Learning.
We began by comparing individual learning in the control condition with social learning in the network conditions. We measured the change in trend accuracy between round 1 and round 3 for the liberal and conservative control groups, the networks without partisan cues, the networks with party logo primes, and the networks with political identity markers. Since the 40 subjects within each network were not independent observations, we measured outcomes at the trial level, such that each trial of 200 subjects (40 subjects per condition) provides five observations. Thus, 12 experimental trials yield 12 observations for each condition and 60 observations in total. This approach permits direct comparisons between outcomes of the control and network conditions, as well as identification of the causal effects of the network conditions on collective outcomes.
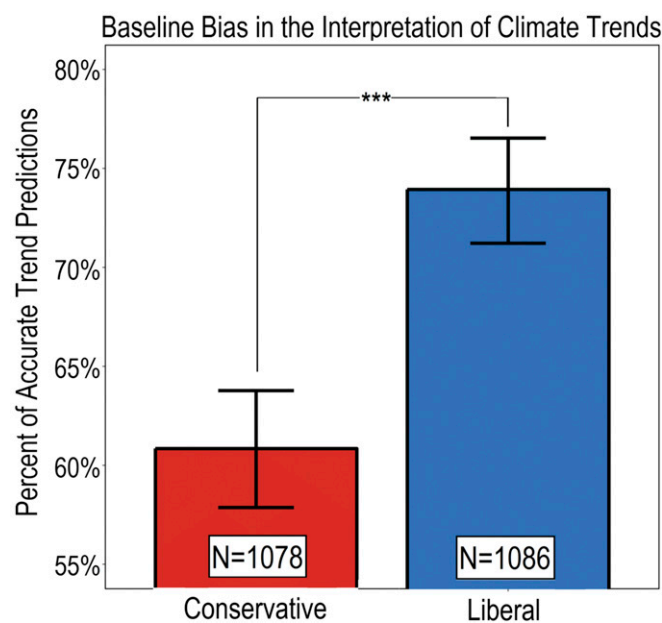

Fig. 2. Baseline trend accuracy for conservatives and liberals across all experimental conditions. Data are from round 1 of the task for all 12 trials. The error bars show 95% confidence intervals. Each individual produces one observation. All observations are independent. One hundred twenty-two conservatives and 114 liberals did not successfully input responses at both round 1 and round 3 (***P < 0.001).

Fig. 3*A* shows the average trend prediction for subjects in each experimental condition in round 1 and round 3. The results show strong social learning effects in the bipartisan networks without partisan cues. In particular, at round 1, there were no significant differences between liberal control groups and bipartisan networks without partisan cues ($n = 24$; $P = 0.32$, Wilcoxon rank sum test); however, by round 3, bipartisan networks without partisan cues showed significantly greater trend accuracy than liberal control groups (by 9.4 percentage points: $n = 24$; $P < 0.01$, Wilcoxon rank sum test). More generally, by round 3, subjects in bipartisan networks without partisan cues achieved significantly greater trend accuracy than subjects in every other condition, including the conservative control groups (by 24.1 percentage points: $n = 24$; $P < 0.01$, Wilcoxon rank sum test), the networks with minimal partisan priming (by 13.3 percentage points: $n = 24$; $P < 0.01$, Wilcoxon rank sum test), and the networks with political identity markers (by 8.2 percentage points; $n = 24$; $P < 0.01$, Wilcoxon rank sum test).

Social learning was reduced in bipartisan networks with political identity markers; however, exposure to network peers with opposing beliefs still facilitated improvements in the interpretation of the climate data. In particular, at round 1, liberal control groups had significantly more accurate trend predictions than bipartisan networks with political identity markers (by 6.3 percentage points: $n = 24$; $P = 0.02$, Wilcoxon rank sum test); however, by round 3, there was no significant difference in trend accuracy between these two conditions ($n = 24$; $P = 0.45$, Wilcoxon rank sum test). Exposure to party logos during subjects' interactions had a stronger effect on dampening social learning in bipartisan social networks. Subjects in bipartisan networks with minimal priming via party logos were not significantly different from liberal control groups at either round 1 ($n = 24$; $P = 0.08$, Wilcoxon rank sum test) or round 3 ($n = 24$; $P = 0.50$, Wilcoxon rank sum test) of the study.

Fig. 3*B* shows the magnitude of the change in accuracy for all conditions from round 1 to round 3. In the control condition, conservative subjects exhibited significant but small improvements
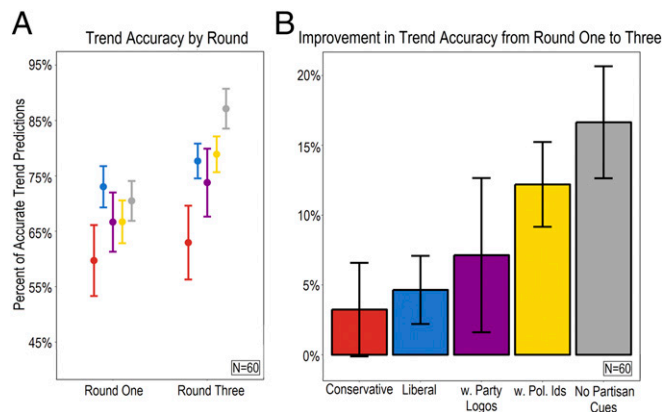
Fig. 3. Changes in average trend accuracy across experimental conditions. (*A*) Round 1 and round 3 estimates from all 12 experimental trials, where each trial provides one observation for each condition. All conditions are independent. The error bars show 95% confidence intervals. (*B*) Bars display the improvement in accuracy from round 1 to round 3, in percentage points. Ids, identities; Pol., political; w., with.

in trend accuracy based on individual reflection (3.2 percentage points: $n = 12$; $P = 0.05$, Wilcoxon signed rank test). Likewise, liberals in control groups also showed significant but small improvements in trend accuracy based on individual reflection (4.6 percentage points: $n = 12$; $P < 0.01$, Wilcoxon signed rank test). As an additional analysis of the results from the control condition, we tested for improvements to trend accuracy using paired individual-level analyses, which provide greater statistical power. The results were consistent with the trial-level measures presented here (*SI Appendix*).

Comparing these results with the network conditions, we find that improvements in trend accuracy were significantly greater in bipartisan networks without partisan cues (16.6 percentage points: $n = 24$; $P < 0.01$, Wilcoxon rank sum test). Subjects in bipartisan networks without partisan cues showed significantly greater improvement than conservative controls (by 13.4 percentage points: $n = 24$; $P < 0.01$, Wilcoxon rank sum test) and liberal controls (by 12 percentage points: $n = 24$; $P < 0.01$, Wilcoxon rank sum test). Moreover, these subjects also showed significantly greater improvement in the interpretation of climate trends than subjects in bipartisan networks with party logos (by 9.5 percentage points: $n = 24$; $P = 0.01$, Wilcoxon rank sum test). Bipartisan networks with party logos did not improve significantly more than either the conservative control groups ($n = 24$; $P = 0.37$, Wilcoxon rank sum test) or the liberal control groups ($n = 24$; $P = 0.47$, Wilcoxon rank sum test). By contrast, subjects in bipartisan networks with political identity markers showed greater improvement than both the conservative control subjects (by 13.4 percentage points: $n = 24$; $P < 0.01$, Wilcoxon rank sum test) and liberal control subjects (by 12 percentage points: $n = 24$; $P < 0.01$, Wilcoxon rank sum test). Overall improvement in trend accuracy in the identity marker condition was surprisingly strong, showing no statistical difference from bipartisan networks without partisan cues ($n = 24$; $P = 0.09$, Wilcoxon rank sum test).

Exposure to Opposing Beliefs Eliminated Partisan Bias in Bipartisan Social Networks Without Partisan Cues. To see whether improvements to trend accuracy in the network conditions had implications for belief polarization, we calculated trial-level outcomes for conservatives and liberals within each condition separately.

Fig. 4 shows the trend accuracy of each political group in each condition at round 1 and round 3. As expected, at round 1, liberals were significantly more accurate than conservatives in every condition: control groups ($n = 24$; $P < 0.01$, Wilcoxon rank sum

test), bipartisan networks with minimal partisan priming ($n = 24$; $P = 0.02$, Wilcoxon signed rank test), bipartisan networks with identity markers ($n = 24$; $P = 0.01$, Wilcoxon signed rank test), and bipartisan networks without partisan cues ($n = 24$; $P = 0.03$, Wilcoxon signed rank test). There were no significant baseline differences in trend accuracy among conservatives across all conditions ($n = 48$; $P = 0.37$, Kruskal–Wallis H test) or among liberals across all conditions ($n = 48$; $P = 0.84$, Kruskal–Wallis H test).

By round 3, however, conservatives in bipartisan networks without partisan cues had significantly more accurate trend predictions than conservatives in the control groups (by 25.4 percentage points: $n = 24$; $P < 0.01$, Wilcoxon rank sum test). On average, 88.3% of conservatives in the networks without partisan cues finished with accurate trend predictions, compared with only 62.9% of conservatives in the control condition. Remarkably, by the end of the study, conservatives in the networks without partisan cues also had significantly more accurate trend predictions than liberals in the control condition (by 10.6 percentage points; $n = 24$; $P < 0.01$, Wilcoxon rank sum test).

The benefits of social learning were not limited to conservatives. Liberals also improved in networks without partisan cues, finishing with significantly higher trend accuracy than liberals in the control condition (by 7.9 percentage points: $n = 24$; $P = 0.01$, Wilcoxon rank sum test). By the end of the study, in bipartisan networks without partisan cues, there were no longer significant
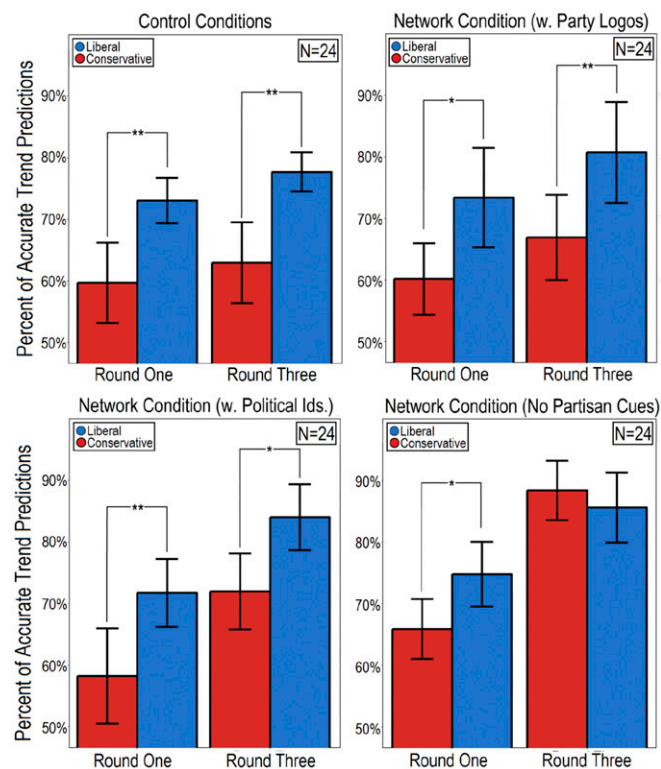
Fig. 4. Change in average trend accuracy, by political group, for each experimental condition. The overall performance of each political group, averaging across all 12 trials, is displayed for both control and network conditions. Network conditions in each trial contain 20 liberals and 20 conservatives, whereas each control condition contains 40 liberals and 40 conservatives, respectively. The trend accuracy of liberals and conservatives in the network conditions was measured by computing the percentage of correct trends estimated by each subgroup, within each network, thus producing two group-level observations for each network and 24 in total for each condition. The error bars display 95% confidence intervals. All observations are at the group level, and are therefore independent (*$P < 0.05$, **$P \leq 0.01$). Ids., identities; w., with.

differences in trend accuracy between the liberals and conservatives ($n = 24$; $P = 0.28$, Wilcoxon signed rank test). In this condition, partisan bias was eliminated from subjects' interpretation of NASA's climate data. Moreover, conservative subjects in this condition ended the study with a slightly higher average trend accuracy than their liberal counterparts.

In all other conditions, significant partisan bias was maintained, with liberals consistently outperforming conservatives. At round 3, liberals exhibited significantly better trend accuracy than conservatives in (*i*) the control condition (by 14.7 percentage points: $n = 24$; $P < 0.01$, Wilcoxon rank sum test), (*ii*) bipartisan networks with party logos (by 13.7 percentage points: $n = 24$; $P < 0.01$, Wilcoxon signed rank test), and (*iii*) bipartisan networks with political identity markers (by 12.4 percentage points: $n = 24$; $P = 0.02$, Wilcoxon signed rank test).

To identify the effects of partisan priming on social learning among conservatives, we compared the final trend accuracy of conservatives across all of the network conditions. Conservatives in bipartisan networks without partisan cues improved significantly more than conservatives in networks with party logos (by 15.7 percentage points: $n = 24$; $P < 0.01$, Wilcoxon rank sum test) and conservatives in networks with political identity markers (by 8.3 percentage points: $n = 24$; $P = 0.03$, Wilcoxon rank sum test). These findings suggest that partisan priming significantly reduced social learning; however, we were surprised to find that it did not entirely eliminate social learning. In bipartisan networks with political identity markers, conservatives still showed significantly more accurate trend predictions at round 3 than conservatives in the control condition (by 9.4 percentage points: $n = 24$; $P = 0.04$, Wilcoxon rank sum test). Similarly, liberals in bipartisan networks with political identity markers completed the study with a higher average trend accuracy than liberals in the control groups (by 7.3 percentage points: $n = 24$; $P = 0.04$, Wilcoxon rank sum test), indicating that both conservatives and liberals benefitted from belief exchange in politically mixed networks, even with identity markers present. The most significant impediment to social learning came from exposure to party logos. Improvements in bipartisan networks with party logos were not statistically different from individual learning in the control condition, for both conservatives ($n = 24$; $P = 0.28$, Wilcoxon rank sum test) and liberals ($n = 24$; $P = 0.66$, Wilcoxon rank sum test), suggesting that this condition offered a form of partisan priming that prevented social learning.

Robustness. We conducted two additional robustness tests to identify the scope conditions of our results. Because this study focused on bipartisan networks, an open question is whether social learning can reduce partisan bias in social networks that are not bipartisan in composition. As a robustness test, we conducted additional trials where belief exchange without partisan cues was facilitated in structured echo chambers [i.e., networks of 40 people consisting exclusively of liberals or conservatives (*SI Appendix*)]. Baseline trend accuracy for conservatives and liberals in these trials was not statistically different from that for participants in all other trials (conservatives: $n = 53$; $P = 0.54$, Kruskal–Wallis H test and liberals: $n = 53$; $P = 0.92$, Kruskal–Wallis H test). We found that social learning was significantly reduced in echo chamber networks compared with bipartisan networks without partisan cues. By round 3, there were no significant differences in the trend accuracy between conservative subjects in echo chamber networks and conservative subjects in the control condition ($n = 17$; $P = 0.26$, Wilcoxon rank sum test). By contrast, at round 3, conservatives in bipartisan networks without partisan cues had significantly more accurate trend predictions than conservatives in echo chamber networks (by 17 percentage points: $n = 17$; $P = 0.04$, Wilcoxon rank sum test). These robustness tests suggest that politically diverse social networks benefit social learning

more than politically homogeneous social networks when the salience of partisanship is minimized in both cases.

We also conducted a set of robustness trials to examine whether political identity markers would disrupt social learning when subjects were exposed to the individual answers of their network peers, as opposed to the average answer of their peer neighborhood. Unlike all of the foregoing trials in this study, in which we studied the network dynamics of peer influence in social networks, this final robustness test was designed to determine whether individuals respond differently to partisan cues when they can see each of their individual neighbors' responses versus when they can only see the average of their neighbors' responses. The results from these robustness tests showed that there were no significant effects of providing individual responses, as opposed to average responses, on subjects' social learning. Subjects' improvement in these trials was consistent with the outcomes observed in the bipartisan networks with political identity markers, in which exposure to peers' political identities reduced social learning (*SI Appendix*).

Discussion

Prior experiments have found that structured bipartisan communication networks can fail to reduce political polarization, and can even lead to the reinforcement of existing partisan biases. These results have been used to argue that exposure to opposing views drives belief polarization in bipartisan networks (19, 20). To the contrary, we find that exposure to opposing beliefs in structured bipartisan networks can eliminate partisan bias. These results are consistent with recent experiments (35), which show that exposure to diverse beliefs can facilitate improvements among less accurate individuals through a process of social learning. Our findings also show that when the salience of political partisanship is increased, even through minimal partisan priming, this social learning effect can be reduced and belief polarization can be sustained. Our findings thus offer a cautionary conclusion. Politically diverse communication networks can indeed eliminate partisan bias in the interpretation of climate data, but increasing the salience of partisanship can significantly limit the effectiveness of social learning.

Our findings contribute to a growing literature on the challenges of communicating scientific findings about climate change in the context of partisan bias and motivated reasoning. Earlier studies of the effects of bias on science communications focused on individual-level responses to messages, and how these responses are altered by message design and framing effects (1–7). We complement this work by showing how partisan bias can be mediated by structured social networks that act as a filter on judgments, with the capacity to enhance individual and collective understanding of scientific information. Our results suggest that minimizing the salience of partisanship in structured bipartisan networks can offer a useful strategy for improving public understanding of contentious scientific information in settings where polarizing issues can lead to biased interpretations. We anticipate that these findings will provide a useful point of departure for future work that aims to identify how social media networks can be employed to accelerate public understanding of science communications in settings where partisan bias can impede their acceptance.

Materials and Methods

This research was approved by the Institutional Review Board at the University of Pennsylvania, where the study was conducted. All subjects provided their political ideology and informed consent during registration. Conservative and liberal subjects were balanced across a number of demographic traits, including gender and age (*SI Appendix*). Upon arriving at the study website, participants viewed instructions on how to play the estimation game. When a sufficient number of subjects arrived, subjects were randomized to a condition and the trial would begin. In all trials, subjects were presented with NASA's graph (Fig. 1) and asked to forecast the amount of Arctic Sea ice in 2025. Subjects in all conditions were awarded monetary

prizes based on the accuracy of their final estimates. Accuracy was determined by how close their final answer was to NASA's projections (*SI Appendix*). Because the participants in every experimental condition were equally incentivized to focus on accuracy, economic motivation for accuracy cannot account for any differences in performance across experimental conditions.

Each condition in every trial contained 40 individuals, such that each experimental trial contained 200 individuals. If placed into a network condition, subjects were randomly assigned to one node in the network, and they maintained this position throughout the experiment. The network conditions employed a fixed random network where every node had the same number of connections. We constructed a network with four edges per node, and we employed the same network topology across all network conditions to minimize variance. We used random decentralized networks because previous experiments illustrate that this topology is reliable for generating social learning in online collective estimation tasks (35). Network conditions varied in whether they displayed additional information on the interface alongside the average of players' network neighbors, where the identity marker condition displayed the political ideologies of network peers and the minimal priming condition displayed an image of party logos for the Democratic and Republican parties (*SI Appendix*).

Each politically mixed network had an equal number of liberals and conservatives. We sorted subjects based on political ideology because political ideology has been shown to be a highly salient dimension for partisan bias in the domain of climate change (1–7), and because people have been found to strongly define their sense of political identity on the basis of political ideology (41–46). In the identity marker condition, usernames were masked for all players, which preserved anonymity among all participants and prevented communication outside of the experiment. In the priming condition, the logos for the Republican and Democratic parties were used as a minimal prime based on recent studies (42, 43), which show that these logos are effective at priming implicit bias based on both party membership and political ideology (*SI Appendix*).

Data Availability. The data analyzed in this paper are included as supporting information and are also publicly available for download from the laboratory website of the Network Dynamics Group (https://ndg.asc.upenn.edu).

1. National Academies of Sciences, Engineering, and Medicine; Committee on the Science of Science Communication; Division of Behavioral and Social Sciences and Education (2017) *Communicating Science Effectively: A Research Agenda* (The National Academies Press, Washington, DC).
2. Kahan D, et al. (2012) The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nat Clim Chang* 2:732–735.
3. McCright A, Dunlap R (2010) Anti-reflexivity: The American conservative movement's success in undermining climate science and policy. *Theory Cult Soc* 27:100–133.
4. McCright A, Dunlap R (2011) The politicization of climate change and polarization in the American public's view of global warming, 2001–2010. *Sociol Q* 52:155–194.
5. Bosetti V, et al. (2017) COP21 climate negotiators' responses to climate model forecasts. *Nat Clim Chang* 7:184–191.
6. Jamieson K, Hardy B (2014) Leveraging scientific credibility about Arctic Sea ice trends in a polarized political environment. *Proc Natl Acad Sci USA* 111:13598–13605.
7. Hardy B, Jamieson K (2017) Overcoming endpoint bias in climate change communication: The case of Arctic sea ice trends. *Environ Commun* 11:205–217.
8. Kunda Z (1990) The case for motivated reasoning. *Psychol Bull* 108:480–498.
9. Kahan D (2013) Ideology, motivated reasoning, and cognitive reflection. *Judgm Decis Mak* 8:407–424.
10. Katz E, Lazarsfeld P (1955) *Personal Influence: The Part Played by People in the Flow of Mass Communications* (Free Press, New York).
11. Pearce W, Holmberg K, Hellsten I, Nerlich B (2014) Climate change on Twitter: Topics, communities and conversations about the 2013 IPCC Working Group 1 report. *PLoS One* 9:e94785.
12. Gough A, et al. (2017) Tweet for behavior change: Using social media for the dissemination of public health messages. *JMIR Public Health Surveill* 3:e14.
13. Watts D, Dodds P (2007) Influentials, networks, and public opinion formation. *J Consum Res* 34:441–458.
14. Bakshy E, Eckles D, Yan R, Rosenn I (2012) Social influence in social advertising: Evidence from field experiments. *Proceedings of the 13th ACM Conference on Electronic Commerce* (ACM, New York), pp 146–161.
15. Williams H, McMurray J, Kurz T, Lambert H (2015) Network analysis reveals open forums and echo chambers in social media discussions of climate change. *Glob Environ Change* 32:126–138.
16. Farrell J (2015) Politics: Echo chambers and false certainty. *Nat Clim Chang* 5:719–720.
17. Jasny L, Waggle J, Fisher D (2015) An empirical examination of echo chambers in US climate policy networks. *Nat Clim Chang* 5:782–786.
18. Fishkin J (2009) *When the People Speak: Deliberative Democracy & Public Consultation* (Oxford Univ Press, Oxford).
19. Sunstein C (2002) The law of group polarization. *J Polit Philos* 10:175–195.
20. Sunstein C (2009) *Going to Extremes* (Oxford Univ Press, Oxford).
21. Price V, Cappella J, Nir L (2002) Does disagreement contribute to more deliberative opinion? *Polit Commun* 19:95–112.
22. Gutmann A, Thompson D (1996) *Democracy and Disagreement* (Harvard Univ Press, Cambridge, MA).
23. Mutz D, Martin P (2001) Facilitating communication across lines of political difference: The role of mass media. *Am Polit Sci Rev* 95:97–114.
24. Mutz D (2002) The consequences of cross-cutting networks for political participation. *Am J Pol Sci* 46:838–855.
25. Kaplan M (1977) Discussion polarization effects in a modified jury decision paradigm: Informational influences. *Sociometry* 40:262–271.
26. Isenberg D (1986) Group polarization: A critical review and meta-analysis. *J Pers Soc Psychol* 50:1141–1151.

27. Lord C, Ross L, Lepper M (1979) Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *J Pers Soc Psychol* 37:2098–2109.
28. Corner A, Whitmarsh L, Xenias D (2012) Uncertainty, skepticism and attitudes toward climate change: Biased assimilation and attitude polarization. *Clim Change* 114:463–478.
29. Munro G, Ditto P (1997) Biased assimilation, attitude polarization, and affect in reactions to stereotype-relevant scientific information. *Pers Soc Psychol Bull* 23:636–653.
30. Dandekar P, Goel A, Lee DT (2013) Biased assimilation, homophily, and the dynamics of polarization. *Proc Natl Acad Sci USA* 110:5791–5796.
31. Ditto P, Munro G, Apanovitch G, Marie A, Lockhart L (1998) Motivated sensitivity to preference-inconsistent information. *J Pers Soc Psychol* 75:53–69.
32. Acemoglu D, Dahleh MA, Lobel I, Ozdaglar A (2011) Bayesian learning in social networks. *Rev Econ Stud* 78:1201–1236.
33. Golub B, Jackson M (2010) Naive learning in social networks and the wisdom of crowds. *Am Econ J Microecon* 2:112–149.
34. Woolley AW, Chabris CF, Pentland A, Hashmi N, Malone TW (2010) Evidence for a collective intelligence factor in the performance of human groups. *Science* 330:686–688.
35. Becker J, Brackbill D, Centola D (2017) Network dynamics of social influence in the wisdom of crowds. *Proc Natl Acad Sci USA* 114:E5070–E5076.
36. Lorenz J, Rauhut H, Schweitzer F, Helbing D (2011) How social influence can undermine the wisdom of crowd effect. *Proc Natl Acad Sci USA* 108:9020–9025.
37. Navajas J, Niella T, Garbulsky G, Bahrami B, Sigman M (2018) Aggregated knowledge from a small number of debates outperforms the wisdom of large crowds. *Nat Hum Behav* 2:126–132.
38. Wood T, Porter E (January 16, 2018) The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Polit Behav*, 10.1007/s11109-018-9443-y.
39. Pennycook G, Rand D (2018) Crowdsourcing judgments of news source quality. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3118471. Accessed July 27, 2018.
40. Pennycook G, Rand D (June 20, 2018) Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 10.1016/j.cognition.2018.06.011.
41. Malka A, Lelkes Y (2010) More than ideology: Conservative–liberal identity and receptivity to political cues. *Soc Justice Res* 23:156–188.
42. Iyengar S, Westwood S (2014) Fear and loathing across party lines: New evidence on group polarization. *Am J Pol Sci* 59:690–707.
43. Slothus R, de Vreese C (2010) Political parties, motivated reasoning, and issue framing effects. *J Polit* 72:630–645.
44. Iyengar S, Sood G, Lelkes Y (2012) Affect, not ideology: A social identity perspective on polarization. *Public Opin Q* 76:405–431.
45. Mason L (2018) Ideologues without issues: The polarizing consequences of ideological identities. *Public Opin Q* 82:280–301.
46. Devine C (2014) Ideological social identity: Psychological attachment to ideological in-groups as a political phenomenon and a behavioral influence. *Polit Behav* 37:1–27.
47. Otten S, Moskowitz G (2000) Evidence for implicit evaluative in-group bias: Affect-biased spontaneous trait inference in a minimal group paradigm. *J Exp Soc Psychol* 36:77–89.
48. Hart S, Nisbet E (2012) Boomerang effects in science communication: How motivated reasoning and identity cues amplify opinion polarization about climate mitigation policies. *Commun Res* 39:701–723.