A Randomized Online Experiment during the 2020 US-Iran Crisis shows that Exposure to Common Enemies can Increase Political Polarization

Eaman Jahani[1], Natalie Gallagher[2], Friedolin Merhout[3], Nicolo Cavalli[4,5], Douglas Guilbeault[6], Yan Leng[7], Christopher A. Bail[8,9]

[1]Institute for Data, Systems and Society, Massachusetts Institute of Technology, 50 Ames St, Cambridge, MA 02142

[2]Department of Psychology, Northwestern University, 2029 Sheridan Road, Evanston, IL 60208

[3]Department of Sociology, University of Copenhagen, 1353 Copenhagen K, Denmark

[4]Carlo F. Dondena Centre, Bocconi University, 1 Via Guglielmo Röntgen, Milan, Italy 20136

[5]Nuffield College and Department of Sociology, Oxford University, 1 New Road, Oxford, United Kingdom OX1 1NF

[6]Haas School of Business, University of California, Berkeley, 2220 Piedmont Ave, Berkeley, CA 94720

[7]McCombs School of Business, University of Texas at Austin, 300 MLK Jr., Austin, TX 78712

[8]Department of Sociology, Duke University, 254 Soc. Psych Hall, Durham, NC 27708

[9]Sanford School of Public Policy, Duke University, Durham, NC 27708

Keywords: polarization, social learning, online bot experiment, partisan priming, identity threat

ABSTRACT

A longstanding theory indicates that the threat of a common enemy can mitigate conflict between members of rival groups. We tested this hypothesis in a pre-registered experiment where 1,670 Republicans and Democrats in the United States were asked to complete an online social learning task with a bot that was labeled as a member of the opposing party. Prior to this task, we exposed respondents to primes about a) a common enemy (involving Iran and Russia); b) a patriotic event; or c) a neutral, apolitical prime. Though we observed no significant differences in the behavior of Democrats as a result of priming, we found that Republicans—and particularly those with very strong conservative views—were significantly *less* likely to learn from Democrats when primed about a common enemy. Because our study was in the field during the 2020 Iran Crisis, we were able to further evaluate this finding via a natural experiment— Republicans who participated in our study after the crisis were even less influenced by the beliefs of Democrats than those Republicans who participated before this event. These findings indicate common enemies may not reduce inter-group conflict in highly polarized societies, and contribute to a growing number of studies that find evidence of asymmetric political polarization in the United States. We conclude by discussing the implications of these findings for research in social psychology, political conflict, and the rapidly expanding field of computational social science.

INTRODUCTION

Political polarization— or the tendency for members of rival political groups to adopt increasingly distant opinions about how to solve social problems— is pervasive in many Western Democracies today[1–4]. In the United States, for example, 59.3% of Democratic voters believe federal aid to the poor should be increased compared to only 20.2% of Republicans voters. Conversely, 68.9% of Republicans believe immigration to the United States should be decreased, compared to 21.9% of Democrats[1]. Such discrepancies extend beyond policy issues into the attitudes of Republicans and Democrats towards each other. The proportion of Americans identified with a political party who would be uncomfortable if their child married someone of the opposite party has risen from less than 10% in 1960 to at least 33% in 2010[2]. These trends show no signs of slowing in the wake of the recent divisive presidential elections and the impeachment trials of President Donald Trump.

Though social scientists have offered many explanations for the recent growth of political polarization, relatively little research has identified solutions to this increasingly urgent social problem.[3] One possibility is that members of rival groups will set aside their differences if they face a shared threat from a common enemy[4–6]. This theory— which has roots in social psychology, social network analysis, and philosophy that date back as far as a Sanskrit treatise on warfare from the 4th century B.C.— has been invoked to explain the consolidation of rival factions and even the emergence of the modern nation-state in Western Europe and many other places[7-11]. Indeed, this

theory is so well-established that it has reached widespread prominence in lay beliefs of intergroup dynamics, as captured by the popular proverb: “the enemy of my enemy is my friend” [9].

Different proposals have been offered to explain this phenomenon. Within the common ingroup identity model[10], this effect occurs because members of rival factions come to realize they have more in common with each other than their shared enemy. In the United States, for example, some argue that the threat of the Soviet Union prevented political polarization throughout the Cold War because it fostered a sense of shared fate or national identity that inspired Americans to set aside their differences in the face of a formidable enemy[11–14]. Yet it is also possible that simply priming national identity might have a similar effect, if the mechanism of depolarization is reminding rival factions about their similarities to each other[7,10]. Another possibility is that the common enemy effect is driven by fear of out-groups, or some combination of this process and ingroup favoritism[11-14].

However, a series of recent studies have provided empirical evidence that runs contrary to the theory that common enemies - by activating shared superordinate identities - bring rival groups together. Dach-Gruschow and Hong (2006)[15], for instance, find that identifying with the common superordinate identity of “American” failed to unite white and Black Americans in the United States in the aftermath of Hurricane Katrina. Klar (2018)[16] builds on this by showing experimentally that priming Republican and Democratic women to identify with each other as “women” (the superordinate gender category) actually led to the amplification of cross-party biases. Klar explains these counterintuitive results using a theory put forward by Rutchik & Eccleston (2010)[17], who argue that “when there is a perception that subgroups do not have a shared conception of the superordinate group, appeals to the common ingroup identity made by outgroup members are likely to backfire” (111). This theory identifies a plausible mechanism for why exposure to common enemies may fail to unite rival groups.

Importantly, there are several reasons to suspect that the backfire effects of common enemy priming are especially likely to hold in the current political landscape of the U.S., which is characterized by high cross-party animosity. We expect that priming superordinate identities via common enemy priming can backfire when animosity among rival political groups is high, since these are conditions in which rival groups are especially likely to disagree on how they conceptualize their shared superordinate identity, as well as who is considered as belonging to it. In particular, we maintain that when rival groups are sufficiently antagonistic, they may dislike and distrust each other as much if not more than the common enemy, such that the common ingroup identity linking rival groups may feel like a threat to the existing partisan identity; under such conditions, priming a common enemy threat may inadvertently decrease social learning and cooperation across rival groups[7,16,17]. This framework is especially well-suited for characterizing the current tensions between the Republican and Democratic parties, given recent survey research[18] indicating that both parties consistently view each other as “un-American” and as a “threat to the

nation". For this reason, exposure to a common enemy that threatens Democrats and Republicans equally as "Americans" may backfire by decreasing cross-party social learning and cooperation.

Furthermore, there are two key reasons to expect that the backfire effects of common enemies may not be uniformly distributed across rival groups, and that they may be especially strong among Republicans in the U.S. context. First, a fairly recent nationally representative survey indicates that Republicans are significantly more likely to identify Democrats as "un-American" and as a "threat to the nation" than the reverse[18]; specifically, 27% of Democrats viewed Republicans as a threat to the nation's well-being, whereas 36% of Republicans viewed Democrats as a threat to the nation's well-being, marking a sizable 9 percentage point difference[18]. Second, these survey results are consistent with a broader body of work demonstrating that Republicans react more strongly to threats concerning partisan and national identities, which prior studies account for through a variety of psychological mechanisms, including Republicans' greater propensity toward patriotism[19,20]. Relatedly, a number of studies document asymmetric polarization, whereby inter-group animosity appears to be driven more by Republicans than Democrats across a range of contexts[21–24]. If, as Rutchik & Eccleston (2010) and Kar (2018) propose, attempts to prime unifying identities can backfire when rival groups differ substantially in their views of their shared identities, then it follows that these backfire effects may be particularly strong among Republicans, who are especially prone to characterizing Democrats as thinking differently about American identity and as ultimately not belonging to it.

Studying whether common enemies reduce political polarization presents numerous methodological challenges. To begin, external threats are not randomly distributed across countries or historical contexts and rigorous causal inference is thus not possible with observational data. Similarly, field experiments that expose members of rival groups to common enemies would either be unethical, logistically impossible, or both. Simple survey experiments, however, lack the external validity necessary to demonstrate whether exposure to a common enemy shapes anything more than ephemeral attitudes or dispositions. In this paper, we adopt a hybrid research design in which we recruited a large group of Democrats and Republicans to participate in a social learning task[23,25] which we developed using the online platform Empirica[26] to study how different primes about collective identity influence how partisans exchange information to collaboratively solve an estimation problem when financial incentives are at stake.

METHODS

This research was approved by the Institutional Review Board at Northwestern University, where the study was conducted, and it included informed consent by all participants. All methods were carried out in accordance with the relevant guidelines and regulations specified by Northwestern's Institutional Review Board. Figure 1 (below) describes our research design. From October 2019 to January 2020, we recruited 1,670 self-identified Republicans or Democrats who live in the United States from an online panel to complete a brief survey about their political preferences. Participants

were randomized into one of three conditions. In the first condition, respondents were asked to read a neutral or apolitical article about early human drawings that were recently discovered by archaeologists in South Africa. This condition serves as our control population. In the second condition, we study the effect of priming ingroup identity alone by asking respondents to read an article about Fourth of July celebrations in several U.S. cities; the ingroup identity being primed by this article is the shared superordinate identity of "American". In the third condition, we exposed respondents to a common enemy prime in which they read an article about how Russia, Iran, and China were conspiring to attack U.S. military and political interests. As our Supplementary Materials describe, these articles were carefully selected from a group of 43 candidate primes from *Reuters.com* that we pretested in order to ensure they created the expected priming effect among both Republicans and Democrats. We elected to use articles from *Reuters.com* because previous studies indicate it is equally trusted and well-respected by Republicans and Democrats[24].

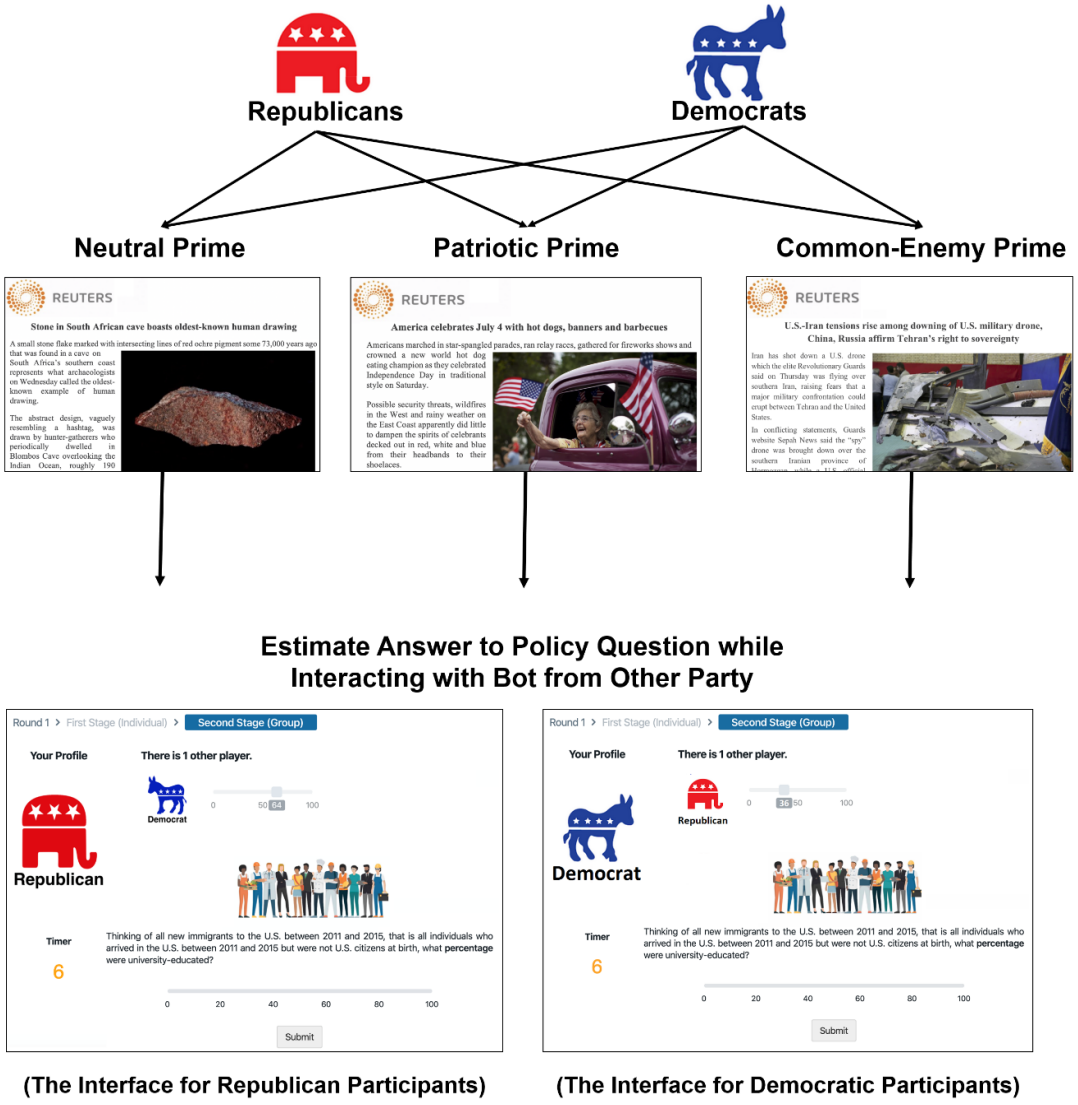


Figure 1: Schematic illustrating the experimental design. 1,670 Republicans and Democrats were randomized into one of three experimental conditions: (1) the neutral prime condition, where they read an article about early human

carvings in South Africa; (2) a patriotic prime condition, where they read an article about July 4th celebrations; and (3) a common-enemy prime condition, where they read an article about the combined threat of Iran, China, and Russia. After reading the article, each participant was offered financial incentives to estimate the answer to a question about a political issue and told that their compensation would increase according to the accuracy of their response. After submitting their first estimate, participants were shown the estimate of a bot impersonating a member of the opposing party. By measuring how often members of each party revise their answers towards the bot in the subsequent round of estimation, we measure how much members of each party learn from the opposing party within each treatment condition.

To collect a behavioral measure of political polarization, we told each respondent they could receive additional pay for providing more accurate estimates to a question about a political issue: "What percentage of immigrants between 2011 and 2015 were college educated?" Respondents first estimated the answer to this question themselves, with no further input. After making their own estimate, they were exposed to the guess of a bot which impersonated a member of the opposing political party who was also involved in the estimation task. The bot always provided an initial estimate that was about 50 percentage points away from the participant's initial guess. After viewing this response, participants were invited to revise their estimate. Below, we report how much participants revised their predictions after being exposed to an estimate that they believed was from a member of the opposing political party. The extent to which each respondent updated their estimate towards that of the bot describes the degree to which participants learned from a member of the opposing party and were willing to cooperatively incorporate their views into their own, when financial incentives are at stake.

To further validate our proposed mechanism - namely that the common enemy article primed the salience of the polarized superordinate identity of "American" - we conducted an exit survey across all conditions that participants completed immediately after the experiment, in which participants were asked to identify the extent to which they identify as being "American", as well as the extent to which they identify with their own and the opposing party

We gained additional information about the effects of common enemy priming via a natural experiment that occurred during our fieldwork.[27] On January 3rd, 2020, United States special forces in Iraq assassinated Qassim Suleimani, an influential Iranian general. This triggered a major geopolitical crisis that many people believed might have caused the outbreak of war between the two countries. This event occurred in the middle of our fieldwork, which began in October 2019 and concluded in late January 2020. Since our common enemy prime involved discussion of US-Iran relations, this unanticipated exogenous event gave us additional leverage to test how increasing the salience of a common enemy interacts with extant partisan tensions. If the saliency of a common enemy can exacerbate polarization - and particularly among Republicans - in highly polarized contexts, we should find that this exogenous shock - which increased the salience of Iran as a common enemy - should similarly exacerbate these backfire effects. Since numerous features of this natural exogenous shock were beyond our methodological control, we are limited in our ability to exploit this shock to highlight particular mechanisms driving changes in participants' responses; such mechanisms may, for instance, include biases concerning the way in which this shock was covered in the media, along with selection biases in which audiences were most likely to encounter these media narratives. Identifying the key channels through which this shock influenced participants' responses is outside the scope of this study; for our purposes, the Iran Crisis served as a natural instrument for validating the robustness of our predictions concerning the backfire effects of common enemy priming. In other words, this theory was falsifiable with respect

to this natural shock: if the Iran crisis had no effect or the opposite effect on the willingness for Democrats and Republicans to influence each other's beliefs, then this would speak against the prediction that increasing the salience of common enemies in polarized contexts exacerbates polarized beliefs and behaviors.

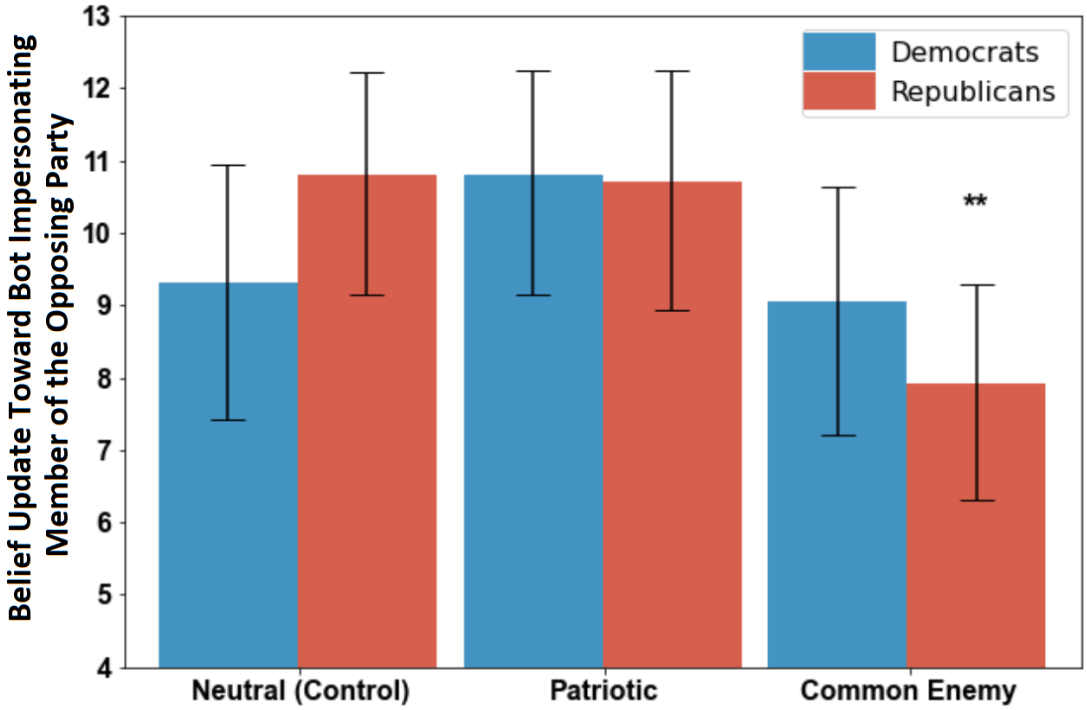


Fig. 2. The extent that participants updated their beliefs toward the opinion of the bot impersonating a member of the opposing party during the collaborative online task, shown across experimental conditions and differentiated by political party. Vertical axis describes the post-stratified average belief update (in percentage points) for participants in each priming condition, where strata are defined by gender, political knowledge, the accuracy of initial guess, and awareness of bot's membership in the opposing party. Larger values indicate that the participant updated their beliefs to become closer to the bot's opinions, demonstrating greater receptivity to social influence from the other party. The neutral condition provides a baseline of comparison or a "control" condition. Error bars display 95% confidence intervals. ** $p < 0.01$. See *SI* for description of our post-stratification methodology.

RESULTS

We estimate the effect of exposure to a common enemy and patriotic prime, respectively, by comparing whether people in these two conditions updated their beliefs more or less towards the bot impersonating the opposing party than those in the neutral, control condition. As Figure 2 shows, we observed no significant differences in the willingness of Democrats to update their estimates towards the bot impersonating a Republican respondent in the study across each priming condition ($p > 0.05$, N = 530). In contrast, Republican participants were significantly less likely to be influenced by the bot - which they believed to be a Democrat - after exposure to the common enemy prime, as compared to Republicans in the control condition ($p < 0.01$, N = 314).

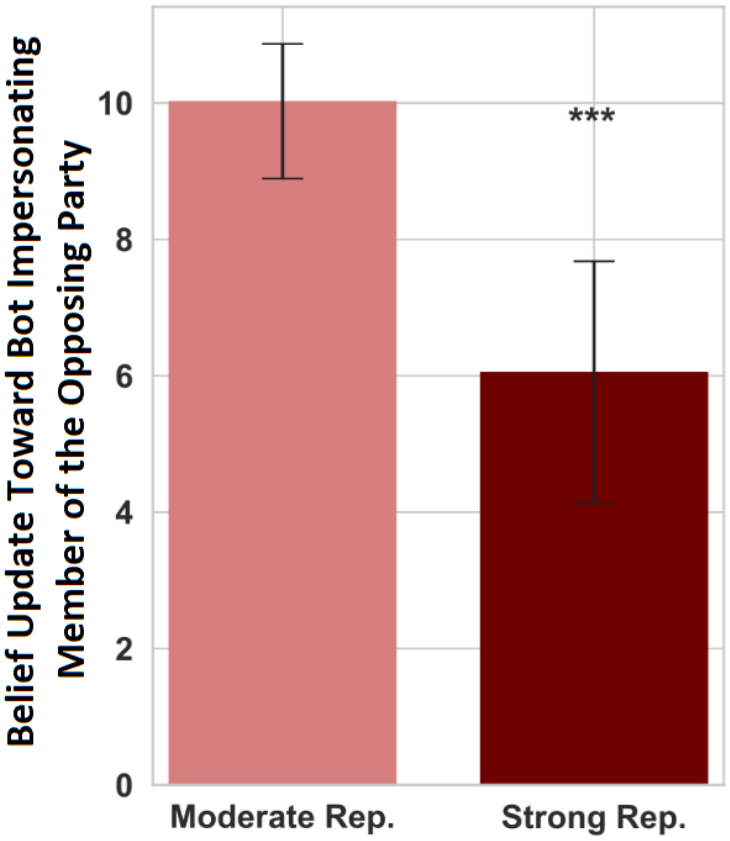


Fig. 3. The extent that Republicans updated their beliefs toward the opinion of the Democrat bot during the online task across experimental conditions, split by the strength of partisanship. The data are collapsed across conditions. Vertical axis describes the post-stratified average belief update (in percentage points), where strata are defined by gender, political knowledge, the accuracy of initial guess, and awareness of bot's membership in the opposing party. Larger values indicate that the participant updated their beliefs to become closer to the bot's opinions, demonstrating greater receptivity to social influence from the other party. Strong Republicans (Rep.) are defined as those who are in the top 10th percentile of partisan identification measured by feeling thermometers. Moderate Republicans (Rep.) are those in all other percentiles. These "moderate republicans" provide a baseline of comparison. Error bars display 95% confidence intervals. *** $p < 0.001$. See *SI* for description of our post-stratification methodology.

What is more, the strength of partisanship among Republicans exacerbated this effect. As Figure 3 shows, Republican respondents who were in the top 10th percentile of a thermometer-based measure of ingroup favoritism were less likely to update towards the bot impersonating a Democrat than those with less strong partisan views ($p < 0.001$, N = 484) (see *SI* for details on this thermometer measure). We observed no significant differences among Democrats using the same strength of partisanship measure.[1]

[1] An alternative method to measure the strength of partisanship is to use the raw measure of feeling toward the ingroup party, instead of its difference with the out-group party. We can then treat the participants in the top 10th percentile of this measure as strongly partisan and the rest as moderately partisan. The 10th percentile cutoff for Republican participants in our data corresponds to 99 out of 100 percentage points for feeling toward Republicans. As there is only

Supplementary analyses support our proposed mechanism (*SI*). Our exit survey indicates that Republicans exposed to the common enemy prime identified significantly more strongly with being "American" than Republicans in the control condition (fig. S6*A*). Yet this increase in identification with being "American" among Republicans did not include a significant change in Republicans' willingness to identify with Democrats, suggesting an exclusive conception of American identity (fig. S6*B*). By contrast, Democrats exposed to the common enemy article were less likely to identify as being "American" than Democrats in the control condition (fig. S7). These results lend further support to our findings regarding asymmetric polarization, which suggest that Republicans and Democrats differed in their conception of American identity and its relation to the common enemy threat, and that an exclusive conception of American identity was activated among Republicans in particular, consistent with recent data from a nationally representative survey[18] indicating that Republicans are especially likely to identify Democrats as un-American and as threats to the nation.

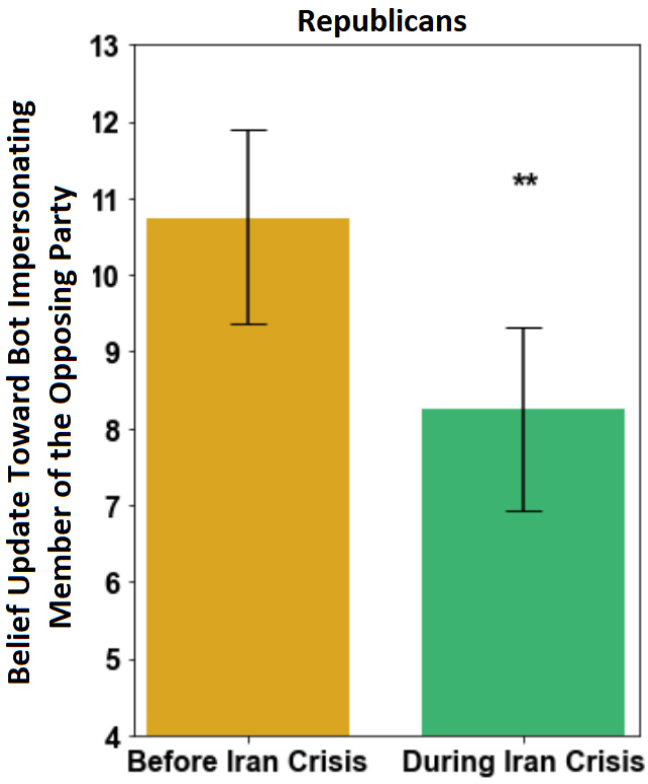


Fig. 4. The extent that Republicans updated their beliefs toward the opinion of the Democrat bot during the collaborative online task across experimental conditions, before and after the 2020 Iran crisis. The data are collapsed across conditions. Vertical axis describes the post-stratified average belief update (in percentage points), where strata are defined by gender, political knowledge, the accuracy of initial guess, and awareness of bot's membership in the opposing party. Larger values indicate that the participant updated their beliefs to become

about 50% overlap between these two different measures of strong partisanship (based on the aforementioned measure or feeling thermometers toward one's own party alone), we expect this alternative measure to act as a robustness check to our result based on the normalized feeling thermometer measure analyzed in the main text. Figure S5 in the appendix uses this alternative measure of ingroup favoritism.

closer to the bot's opinions, demonstrating greater receptivity to social influence from the other party. The "Before Iran Crisis" outcomes provide a baseline of comparison. Error bars display 95% confidence intervals. ** $p < 0.01$. See *SI* for description of our post-stratification methodology.

Lastly, we examine the willingness of participants to use information provided by a member of the other party after a natural shock which occurred during our fieldwork and was expected to amplify the salience of our common enemy prime. In the middle of our fieldwork, on January 3rd, 2020, United States special forces in Iraq assassinated Qassim Suleimani, an influential Iranian general, triggering widespread panic about the possibility of war. Since our common enemy prime depicted Iran as a common enemy, this unanticipated exogenous event gave us additional leverage to test how increasing the salience of a common enemy interacts with extant partisan tensions. As Figure 4 shows, Republicans were even less likely to be influenced by Democrats after the assassination than before this event, regardless of which priming condition they received ($p < .01$, $N = 485$). This finding suggests that Republicans were even less likely to be influenced by the views of Democrats when the salience of the common enemy threat was increased. Our exit survey results are similarly consistent with these behavioral outcomes, since Republicans were found to significantly increase the extent to which they identified with being American during the Iran crisis, whereas no change was observed in Democrats' degree of national identification amid the crisis (fig. S8). Our supplementary appendix shows that all of the results above are highly robust to a myriad of statistical tests and methods.

DISCUSSION

To the best of our knowledge, this study provides the first experimental analysis of the behavioral effects of exposure to a common enemy in the United States during a period of extreme political polarization. Contrary to widespread belief, we found that threats from a common enemy either led to no changes in social learning among partisans or— in the case of Republicans— led their beliefs to be even less influenced by Democrats. These findings are consistent with a growing number of studies that document asymmetric polarization— or patterns of inter-group animosity that appear to be driven more by Republicans than Democrats[21–24]. These results suggest that political narratives about global, combative conflicts—which politicians often invoke rally patriotic support—may have the unintended consequence of increasing polarization within a nation.

Our study provides several important contributions to the study of political polarization and computational social science more broadly. First, our study contributes to a growing body of work on "backfire effects" in political communication[16,17,22], where exposure to the attitudes and beliefs of a rival political group have been shown to exacerbate partisan bias. A number of recent studies have found asymmetric backfire effects where partisan bias is particularly amplified among Republicans as a result of cross-party interaction. For example, one recent study of cross-party communication over Twitter found that Republicans were more likely to increase their partisan bias in response to exposure to social media messages from opinion leaders from the opposing party[22]. This result is consistent with social learning experiments which show that partisan priming can lead Republicans to be significantly less cooperative than Democrats when discussing climate change[23]. Since cross-party interaction has been found to consistently entrench partisan bias, a number of studies have proposed that exposure to a common enemy may encourage cross-party influence and cooperation and thereby reduce political tensions[9-14]. However, the results from this study suggest that exposure to a shared enemy may not be sufficient to eliminate partisan

boundaries to information sharing and cooperation, and may even amplify political tensions—particularly among Republicans.

More generally, our findings are consistent with the burgeoning theory that in societies that experience extreme polarization such as the United States today, partisan tensions may be high enough that political rivals are perceived as more closely connected to the external enemy than the nation itself [5,7,15-18,28,29]. Under such conditions, the threat of a common enemy may increase political tensions among rival groups. Thus, our finding that Republicans learn less from Democrats after exposure to a common enemy may simply reflect their perception that their local opponents are somehow sympathetic to the shared enemy, consistent with recent nationally representative surveys of political attitudes[18]. This sentiment could be observed during the Iran crisis, when many prominent Republican leaders accused democrats of unduly lamenting the death of a dangerous Iranian general; consistent with our theory, we find that the backfire effects of common enemy priming were stronger after the Iranian crisis, and particularly among Republicans, suggesting that the Iran crisis may have intensified Republican perceptions of Democrats as un-American and as threats to the nation comparable to their foreign enemy. Together, these findings provide critical insight into the behavioral dynamics of political polarization in highly polarized societies such as the United States.

REFERENCES

1. American National Election Study. 2016 Time Series Study.
2. Iyengar, S., Sood, G. & Lelkes, Y. Affect, Not Ideology: A Social Identity Perspective on Polarization. *Public Opin. Q.* **76**, 405–431 (2012).
3. Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N. & Westwood, S. J. The Origins and Consequences of Affective Polarization in the United States. *Annu. Rev. Polit. Sci.* **22**, 129–146 (2019).
4. Heider, F. Attitudes and Cognitive Organization. *J. Psychol.* **21**, 107–112 (1946).
5. Simmel, G. *Conflict / The Web Of Group Affiliations*. (Free Press, 1955).
6. Coser, L. *Lewis a. Coser-The Functions of Social Conflict-Free Press*. (The Free Press, 1956).
7. Brewer, M. Superordinate Goals versus Superordinate Identity as Bases of Intergroup Cooperation. in *Social Identity Processes: Trends in Theory and Research* (2000).
8. Brubaker, R. Ethnicity, Race, and Nationalism. *Annu. Rev. Sociol.* 35, 21–42 (2009).
9. Heider, F. *The Psychology of Interpersonal Relations*. (Psychology Press, 1958).
10. Gaertner, S. L., Dovidio, J. F., Anastasio, P. A., Bachman, B. A. & Rust, M. C. The Common Ingroup Identity Model: Recategorization and the Reduction of Intergroup Bias. *Eur. Rev. Soc. Psychol.* **4**, 1–26 (1993).
11. Desch, M. C. War and strong states, peace and weak states? *Int. Organ.* **50**, 237–268 (1996).
12. Walt, S. The Case Against Peace – Foreign Policy. *Foreign Policy* (2016).
13. Skrentny, J. The Effect of the Cold War on African-American Civil Rights: America and the World Audience, 1945-1968. *Theory Soc.* **27**, 237–285 (1998).
14. Sullivan, D. L., Landau, M. J. & Rothschild, Z. K. An Existential Function of Enemyship: Evidence That People Attribute Influence to Personal and Political Enemies to Compensate for Threats to Control. *J. Pers. Soc. Psychol.* **98**, 434–449 (2010).
15. Dach-Gruschow, Karl, and Ying-yi Hong. "The Racial Divide in Response to the Aftermath of Katrina: A Boundary Condition for Common Ingroup Identity Model." *Analyses of Social*

Issues and Public Policy 6, no. 1 (2006): 125–41. https://doi.org/10.1111/j.1530-2415.2006.00110.x.
16. Klar, Samara. "When Common Identities Decrease Trust: An Experimental Study of Partisan Women." *American Journal of Political Science* 62, no. 3 (2018): 610–22. https://doi.org/10.1111/ajps.12366.
17. Rutchick, Abraham M., and Collette P. Eccleston. "Ironic Effects of Invoking Common Ingroup Identity." *Basic and Applied Social Psychology* 32, no. 2 (May 19, 2010): 109–17. https://doi.org/10.1080/01973531003738353.
18. Pew Research Center. "Political Polarization in the American Public," 2014. https://www.pewresearch.org/politics/2014/06/12/political-polarization-in-the-american-public/
19. Jost, John T., Jack Glaser, Arie W. Kruglanski, and Frank J. Sulloway. "Political Conservatism as Motivated Social Cognition." *Psychological Bulletin* 129, no. 3 (2003): 339–75. https://doi.org/10.1037/0033-2909.129.3.339.
20. Jost, John T. "The End of the End of Ideology." *The American Psychologist* 61, no. 7 (October 2006): 651–70. https://doi.org/10.1037/0003-066X.61.7.651.
21. Grossmann, M., D. Hopkins. *Asymmetric Politics: Ideological Republicans and Group Interest Democrats.* (Oxford University Press, 2016).
22. Bail, C. A., L. P. Argyle, T. W. Brown, J. Bumpus, H. Chen, M. B. Fallin Hunzaker, J. Lee, M. Mann, F. Merhout, and A. Volfovsky. "Exposure to Opposing Views on Social Media Can Increase Political Polarization." *Proceedings of the National Academy of Sciences of the United States of America* 115, no. 37 (11 2018): 9216–21.
23. Guilbeault, D., J. Becker, and D. Centola. "Social Learning and Partisan Bias in the Interpretation of Climate Trends." *Proceedings of the National Academy of Sciences* 115, no. 39 (September 25, 2018): 9714–19.
24. Budak, C., S. Goel, J. Rao. Fair and Balanced? Quantifying Media Bias through Crowd-sourced Content Analysis. *Public Opin. Q.* 80, 250–271 (2016).
25. Guilbeault, Douglas, Samuel Woolley, and Joshua Becker. "Probabilistic Social Learning Improves the Public's Judgments of News Veracity." *PLOS ONE* 16, no. 3 (March 9, 2021): e0247487. https://doi.org/10.1371/journal.pone.0247487.
26. Almaatouq, Abdullah, Joshua Becker, James P. Houghton, Nicolas Paton, Duncan J. Watts, and Mark E. Whiting. "Empirica: A Virtual Lab for High-Throughput Macro-Level Experiments." *Behavior Research Methods* 53, no. 5 (October 1, 2021): 2158–71. https://doi.org/10.3758/s13428-020-01535-9.
27. Muñoz, J., A. Falcó-Gimeno, E. Hernández. Unexpected Event during Survey Design: Promise and Pitfalls for Causal Inference. *Polit. Analysis* 28, 186–206 (2020).
28. Molden, D. *Understanding priming effects in social psychology: An overview and integration in Understanding priming effects in social psychology*. (Guilford Press, 2014).
29. Gibler, D. M. Outside-In: The Effects of External Threat on State Centralization. *J. Confl. Resolut.* **54**, 519–542 (2010).

Data availability: All data underlying this study will be made publicly available upon publication, here: (github link to be provided).

Code availability: All code underlying this study will be made publicly available upon publication, here: (github link to be provided).

Supplementary Information for:

A Randomized Online Experiment during the 2020 US-Iran Crisis shows that Exposure to Common Enemies can Increase Political Polarization

This file includes:

SUPPLEMENTARY METHODS

Recruitment and Full Set of Measures. We conducted the recruitment for this study using a panel of 3,177 self-identified Democrats and Republicans collected via CloudResearch during spring 2019. Cloud Research (formerly known as TurkPrime) is an online survey company that enables the collection of high-quality survey panels from Amazon Mechanical Turk users using advanced screeners to remove low-quality respondents and inauthentic survey respondents. Though this source does not provide samples that are representative of the United States population, recent research has demonstrated that samples drawn from CloudResearch capture more variation among important demographic and political belief covariates than in-person lab samples (*1,2*). During the profiling process, each of these respondents were asked a series of questions about their demographic characteristics and political beliefs. Between October 2019 and January 2020, we invited 3,162 participants from the online panel to participate in the study. The invitations and study description specified that participants would be compensated $2 for participation with an additional bonus of up to $1 based upon their accuracy in the estimation task (answering the question about U.S. immigration described in the main text of our article). Since we expected performance in the estimation task to depend on the feedback participants received from the bots—and since bot feedback in the first round was based on initial participant input (participant input +/- 50 percentage points)—we ultimately compensated all respondents with the full bonus amount regardless of the accuracy of their answers.

1,692 out of the 3,162 people invited to join the study initially agreed to participate for a raw response rate of 53.5%. As Table S1 below shows, we observed no significant differences in response rates between Republicans and Democrats. We ran additional models (not shown) that examined whether age, gender, race and ethnicity, level of education, or income level were significant predictors of non-response. Of these, only age was a significant predictor of non-response: older people were slightly more likely to participate ($p < 0.001$). Our sample also compares favorably to the national population on each of these measures. The mean age of respondents in our sample is 42.2 compared to the national average of 37.84 reported by the 2016 American Community Survey. Our respondents were 52.2% female, compared to the national average of 51%. Our sample contained fewer racial

minorities than the national average, however. Our sample was 82.2% white or caucasion compared to 70% (the national average).

Study Response Rates by Party

	Did not Participate in Study	Participated in Study	Raw Response Rate
Democrats	569	655	53.5
Democrat Leaners	205	219	51.7
Republicans	549	651	54.3
Republican Leaners	147	167	53.2

Table S1: Study Response Rate by Party Identification.

Before they were forwarded to the online platform where they performed the collaborative task with a bot impersonating a member of the opposing party, respondents were sent a link that redirected them towards an interface that asked them to provide their MTurk "worker id" (a unique identifier given to each Mturk worker by Amazon). On that platform, participants first read an informed consent dialogue detailing the study process, the conditions for compensation, and their rights as study participants. If they consented to participate in the study, respondents were then directed to the study's main online platform where they first were shown a profile populated with the age, gender, and party identification they listed for themselves when they first joined the panel used to recruit respondents to the study. Participants were offered the opportunity to change these details before proceeding to the next stage which introduced them to the estimation task.

Before entering the priming conditions, subjects participated in an example of the estimation task to ensure they understood the instructions. During this example task, subjects were shown the experimental estimation interface (figure 1 in the main text of our article), but with a different political question: "What percentage of food stamp (SNAP) recipients do you believe are employed full-time or part-time?" After completing the example estimation task, participants were randomized into one of three priming conditions (the neutral, patriotic, or common-enemy prime) conditions. Subjects were presented with a prompt stating: "Please read the following article taken from Reuters, a non-partisan news outlet. On the next screen, we are going to ask you several questions about the article. You will need to answer those correctly to advance to the game." Next we asked respondents to answer three questions about facts presented in the article. Subjects were able to retry answering these questions until they answered each of them correctly. Once each question was answered correctly, subjects entered the estimation task, where they were asked to provide an estimate to the following political question described in the main text of our article: "Thinking of all new immigrants to the U.S. between 2011 and 2015, that is all individuals who arrived in the U.S. between 2011 and 2015, but were not U.S. citizens at birth, what percentage were university-educated?" At Round One, all subjects regardless of condition provided an independent estimate, without exposure to the estimate from the bot impersonating a member of the opposing party.

At round two of the collaborative online task on our platform, subjects were given an opportunity to revise their estimate after being exposed to the estimate of a bot described as a member of the opposing party. This bot was programmed to initially provide an estimate that was exactly 50 percentage-points away from the participant's estimate. This manipulation allowed us to test the effect of each priming condition on the willingness of participants to revise their estimates toward the estimate provided by the bot. For each bot guess after round two, the bot either stuck close to its initial guess (the stubborn bot) or moved its guess closer to the participant's guess (the friendly bot updated in a memoryless Bayesian fashion, such that the prediction for bot in round $t+1$ is $x_{bot,t+1} = \frac{3}{4}\, x_{bot,t} + \frac{1}{4}\, x_{participant,t} + Uni(-3, 3)$). Participants were randomly assigned to interact with the stubborn or friendly bot for all rounds of the game.

After completing five rounds of the game, participants completed an exit survey with several measures in the following order. First, participants were asked a political knowledge question: "Which party has a majority of seats in the Senate?" Next, they were asked to identify the political party of the other player in the game as a manipulation check. They then estimated the political knowledge of the other person (i.e. the bot) in the game. Participants indicated their level of identification with five different groups using a single-item measure of identification: Americans, people in their state, people in their neighborhood, Democrats, and Republicans (*3*). They completed a mood measure (the PANAS), and feeling thermometers towards ingroups (their political party, and all americans, and outgroups (the opposing political party, Russians, Iranians, and Chinese), and two distractor groups (Texans and Californians) (*3*). Finally, participants responded to several questions about their experience in the game, and were debriefed and compensated for their time. A video that illustrates the entire participant experience is available online.

Selection of News Articles for Treatment Conditions. In order to identify neutral, patriotic, and common-enemy primes, we conducted multiple rounds of pre-testing of news articles to ensure that each one created a) no response; b) a prideful response; or c) a fearful or threat response. Our pre-testing effort was further designed to ensure that neither Democrats nor Republicans were significantly more likely to respond to one of these primes. We chose to select articles from Reuters because previous studies indicate it as the most centrist news organization at the time of this writing (*4*). We gathered and pre-tested a total of 42 articles from Reuters (12 neutral articles, 12 articles that prime patriotism, and 18 that describe a common enemy). To ensure comparable length — and to expunge explicit partisan cues from the text of the prime — we edited the original Reuters articles before pre-testing them to an average length of 400 words. Each article was associated with a single image, sourced from the original publication.

Between March 29th and April 1st, 2019 we pretested the 42 articles by asking 566 respondents (347 Democrat and 219 Republicans) recruited from MTurk, to complete a survey on Qualtrics about their attitudes and feelings. This survey presented respondents with one article, randomly selected out of the set of 42 total articles, and asked the same set of questions used in the exit survey of the main study. We used an iterative process to select the three final articles. First, we searched for the article in each category (neutral, patriotic, or common enemy) that created equal shifts in party identification, identification with America, and emotional responses among both Republicans and Democrats. To ensure that respondents viewed the potential primes as realistic news stories, our pre-testing survey also included open-ended questions designed to determine whether respondents

suspected the articles had been edited or fabricated. We received no indications from respondents that they expected such editing or manipulation.

All of the articles used as primes in this experiment are provided below. The prime for the control condition is shown as fig. S1, for the patriotic condition as fig. S2, and for the common enemy condition as fig. S3. Our analysis suggested the best "neutral" prime was an article titled "Stone in African cave boasts oldest-known human drawing" (Table S2). The article focused on details of an archaeological excavation in Blombos Cave, near the Indian ocean. To select the ingroup prime, for each candidate article, we compared self-reported importance of being American to MTurkers after reading the article with the levels recorded among respondents randomly selected to read the candidate neutral article. We found only one article that effectively primed patriotism, by generating a significant difference in self-reported American identity among Republicans and Democrats (n=173, p>|z|=0.073, Wilcoxon rank sum test, Table S3). The article was entitled "America celebrates July 4th with hot dogs, banners and barbecues," and discussed Fourth of July celebrations in various sites throughout the U.S., including both cities with a large population of Democrats (e.g. Austin, Texas) and Republicans (e.g. Gainesville, Florida). A similar article was used to prime patriotic responses in a previous study (*5*).

Selecting the article that created an equally strong sense of threat or fear among both Republicans and Democrats proved more challenging. In our initial round of testing, we discovered no suitable article— though the issue was mostly related to the *amount* of threat experienced by respondents, not the consistency of this feeling across members of both parties. This led us to select four articles for a second round of pre-testing with an expanded sample on Qualtrics as well as a small sample on the main study platform. We used this second round of pre-testing to get more precise estimates of the identification measures described above and to assess the viability of the articles in the context of the experimental platform. To ensure that our respondents would feel sufficiently threatened, we also created and tested a new common enemy prime by combining content from three news stories from Reuters about possible threats from Iran, China, and Russia into a single article entitled "U.S.-Iran tensions rise among downing of U.S. Military Drone, China, Russia affirm Tehran's right to sovereignty." By combining real-world news stories, the article discusses a scenario where Iran has shot down a U.S. drone, thereby raising fears of a military confrontation; meanwhile, the article suggests that China and Russia both independently sided with Iran in the conflict, thus suggesting the growing threat of a joined threat from all three enemies of the U.S. We found that this article significantly primed American identity compared to the neutral article (Table S4). Therefore, we selected it as our common enemy prime.

	Identity		*Emotions*										
	American	Party	Distressed	Afraid	Upset	Scared	Nervous	Enthusiastic	Proud	Strong	Inspired	Excited	*n*
Neutral primes (compared to baseline)													
Alaska	0.731	0.144	0.253	0.069^{+}	0.128	0.207	0.007^{**}	0.109	0.094^{+}	0.009^{**}	0.048^{*}	0.034^{*}	76
Cave	0.513	0.178	0.072^{+}	0.103	0.282	0.231	0.043^{*}	0.213	0.051^{+}	0.023^{*}	0.719	0.713	73
Dog Show	0.202	0.309	0.005^{**}	0.002^{**}	0.017^{*}	0.004^{**}	0.003^{**}	0.034^{*}	0.014^{*}	0.054^{+}	0.018^{*}	0.006^{**}	77
Ethiopian Airline	0.829	0.745	0.013^{*}	0.480	0.011^{*}	0.935	0.442	0.001^{***}	0.006^{**}	0.026^{*}	0.025^{*}	0.001^{***}	76
Luke Perry	0.498	0.178	0.253	0.050^{+}	0.976	0.128	0.060^{+}	0.000^{***}	0.000^{***}	0.001^{***}	0.000^{***}	0.000^{***}	73
Lunar New Year	0.919	0.094^{+}	0.558	0.103	0.255	0.039^{*}	0.039^{*}	0.003^{**}	0.003^{**}	0.009^{**}	0.007^{**}	0.005^{**}	73
Mozambique	0.421	0.172	0.021^{*}	0.666	0.026^{*}	0.665	0.345	0.000^{***}	0.000^{***}	0.008^{**}	0.005^{**}	0.001^{**}	74
Myanmar	0.204	0.047^{*}	0.147	0.814	0.034^{*}	0.549	0.743	0.000^{***}	0.000^{***}	0.003^{**}	0.000^{***}	0.000^{***}	70
New Species	0.717	0.782	0.318	0.983	0.207	0.477	0.194	0.293	0.318	0.815	0.565	0.703	76
New Zealand	0.980	0.259	0.000^{***}	0.012^{*}	0.000^{***}	0.018^{*}	0.333	0.001^{***}	0.000^{***}	0.002^{**}	0.000^{***}	0.000^{***}	79
Oscars	0.827	0.117	0.033^{*}	0.054^{+}	0.316	0.101	0.015^{*}	0.003^{**}	0.001^{**}	0.002^{**}	0.006^{**}	0.075^{+}	73
Vienna	0.948	0.357	0.021^{*}	0.002^{**}	0.017^{*}	0.022^{*}	0.015^{*}	0.015^{*}	0.014^{*}	0.025^{*}	0.076^{+}	0.035^{*}	77

Table S2. Wilcoxon Rank Sum Tests for Identity and Emotions scores for candidate neutral articles, compared to a baseline. Selected article in bold.

	Identity		*Emotions*										
	American	Party	Distressed	Afraid	Upset	Scared	Nervous	Enthusiastic	Proud	Strong	Inspired	Excited	*n*
Ingroup primes (compared to pooled neutral primes)													
Alabama	0.571	0.532	0.001^{**}	0.001^{**}	0.007^{**}	0.006^{**}	0.000^{***}	0.682	0.570	0.158	0.689	0.823	177
D-Day	0.161	0.106	0.101	0.057^{+}	0.065^{+}	0.024^{*}	0.036^{*}	0.912	0.000^{***}	0.625	0.226	0.468	180
Economy	0.114	0.246	0.032^{*}	0.152	0.013^{*}	0.207	0.639	0.885	0.806	0.448	0.121	0.627	182
Gold Medals	0.825	0.634	0.239	0.119	0.013^{*}	0.432	0.279	0.013^{*}	0.000^{***}	0.090^{+}	0.002^{**}	0.026^{*}	179
Hot Dog	0.073^{+}	0.845	0.134	0.165	0.044^{*}	0.658	0.226	0.011^{*}	0.000^{***}	0.028^{*}	0.012^{*}	0.033^{*}	179
July 4	0.990	0.241	0.046^{*}	0.300	0.057^{+}	0.153	0.374	0.036^{*}	0.028^{*}	0.149	0.059^{+}	0.244	183
Memorial Day	0.636	0.945	0.389	0.162	0.128	0.188	0.223	0.222	0.525	0.631	0.259	0.123	178
Pyongyang	0.913	0.174	0.025^{*}	0.077^{+}	0.006^{**}	0.319	0.556	0.081^{+}	0.005^{**}	0.300	0.215	0.099^{+}	180
Super Soldiers	0.776	0.322	0.237	0.129	0.112	0.782	0.602	0.047^{*}	0.051^{+}	0.073^{+}	0.011^{*}	0.053^{+}	179
Thanksgiving	0.769	0.828	0.015^{*}	0.103	0.050^{+}	0.145	0.051^{+}	0.853	0.208	0.909	0.704	0.284	185
Tornados	0.161	0.923	0.007^{**}	0.076^{+}	0.012^{*}	0.034^{*}	0.003^{**}	0.262	0.071^{+}	0.736	0.114	0.189	185
Veterans	0.211	0.837	0.003^{**}	0.963	0.115	0.804	0.260	0.539	0.710	0.824	0.861	0.739	178

Table S3. Wilcoxon Rank Sum Tests for Identity and Emotions scores for patriotic, compared to neutral articles. Selected article in bold.

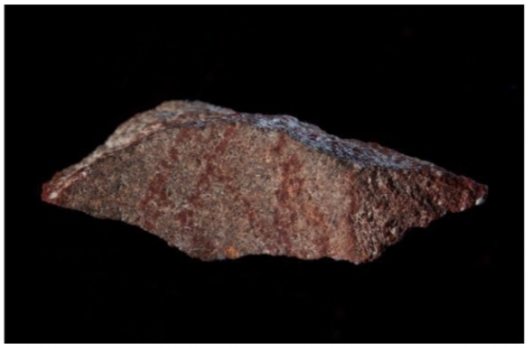
	Identity		*Emotions*										
	American	Party	Distressed	Afraid	Upset	Scared	Nervous	Enthusiastic	Proud	Strong	Inspired	Excited	*n*
Outgroup primes													
First pre-test (compared to pooled neutral primes)													
Isis:													
‗ Rise and fall	0.608	0.234	0.115	0.745	0.963	0.997	0.848	0.539	0.915	0.550	0.426	0.122	178
‗ Organ harvesting	0.657	0.140	0.002^{**}	0.016^{*}	0.001^{**}	0.009^{**}	0.006^{**}	0.059^{+}	0.152	0.942	0.169	0.059^{+}	181
‗ Slavery	0.064^{+}	0.858	0.008^{**}	0.982	0.000^{***}	0.828	0.526	0.000^{***}	0.001^{***}	0.009^{**}	0.001^{***}	0.003^{**}	182
‗ Defeat	0.230	0.486	0.436	0.594	0.867	0.522	0.275	0.512	0.853	0.576	0.298	0.843	177
‗ Iraq	0.406	0.947	0.188	0.058^{+}	0.746	0.254	0.099^{+}	0.005^{**}	0.022^{*}	0.065^{+}	0.073^{+}	0.121	176
‗ Cyber attack	0.364	0.724	0.849	0.480	0.751	0.399	0.781	0.783	0.767	0.746	0.275	0.110	179
‗ Homegrown	0.074^{+}	0.673	0.204	0.211	0.437	0.309	0.061^{+}	0.619	0.756	0.145	0.182	0.169	176
‗ Palmyra	0.718	0.281	0.024^{*}	0.293	0.017^{*}	0.077^{+}	0.217	0.356	0.327	0.637	0.202	0.975	181
Iran:													
‗ Gulf	0.472	0.354	0.976	0.593	0.540	0.812	0.447	0.619	0.203	0.265	0.122	0.179	180
‗ Death to US	0.877	0.646	0.419	0.189	0.420	0.319	0.129	0.508	0.430	0.981	0.189	0.569	181
China:													
‗ Trade War	0.373	0.604	0.315	0.928	0.816	0.633	0.879	0.049^{*}	0.033^{*}	0.360	0.188	0.257	178
‗ Tech race	0.409	0.371	0.083^{*}	0.018^{*}	0.179	0.001^{**}	0.008^{**}	0.095^{+}	0.113	0.129	0.350	0.041^{*}	176
‗ Cyber attack	0.739	0.416	0.007^{**}	0.033^{*}	0.014^{*}	0.006^{**}	0.008^{**}	0.046^{*}	0.011^{*}	0.213	0.006^{**}	0.037^{*}	179
Russia:													
‗ Weapons	0.218	0.972	0.024^{*}	0.003^{**}	0.302	0.000^{***}	0.005^{**}	0.009^{**}	0.006^{**}	0.068^{+}	0.004^{**}	0.057^{+}	177
‗ Cuba help	0.742	0.251	0.908	0.036^{*}	0.420	0.039^{*}	0.007^{**}	0.553	0.552	0.279	0.689	0.495	178
Others:													
‗ Al Qaeda	0.170	0.351	0.466	0.827	0.519	0.277	0.517	0.005^{**}	0.123	0.229	0.007^{**}	0.018^{*}	179
‗ Pearl Harbor	0.378	0.274	0.659	0.235	0.179	0.461	0.272	0.045^{*}	0.018^{*}	0.612	0.740	0.048^{*}	179
‗ Cuba	0.916	0.501	0.959	0.937	0.609	0.737	0.751	0.081^{+}	0.677	0.325	0.416	0.545	178
Second pre-test (compared to selected neutral prime)													
Organ harvesting	0.859	0.003^{*}	0.000^{***}	0.000^{***}	0.000^{***}	0.000^{***}	0.000^{***}	0.000^{***}	0.000^{***}	0.309	0.000^{***}	0.000^{***}	229
Rise and fall	0.264	0.048^{*}	0.000^{***}	0.012^{*}	0.000^{***}	0.012^{*}	0.000^{***}	0.000^{***}	0.001^{***}	0.465	0.000^{***}	0.000^{***}	221
Homegrown	0.410	0.192	0.000^{***}	0.000^{***}	0.000^{***}	0.000^{***}	0.000^{***}	0.000^{***}	0.000^{***}	0.023^{*}	0.000^{***}	0.000^{***}	229
Superenemy	0.040^{*}	0.849	0.000^{***}	0.000^{***}	0.000^{***}	0.000^{***}	0.000^{***}	0.000^{***}	0.000^{***}	0.025^{*}	0.000^{***}	0.000^{***}	226

Table S4. Wilcoxon Rank Sum Tests for Identity and Emotions scores for outgroup articles. Selected article in bold.

REUTERS

Stone in South African cave boasts oldest-known human drawing

A small stone flake marked with intersecting lines of red ochre pigment some 73,000 years ago that was found in a cave on South Africa's southern coast represents what archaeologists on Wednesday called the oldest-known example of human drawing.



The abstract design, vaguely resembling a hashtag, was drawn by hunter-gatherers who periodically dwelled in Blombos Cave overlooking the Indian Ocean, roughly 190 miles (300 km) east of Cape Town, the researchers said. It predates the previous oldest-known drawings by at least 30,000 years.

While the design appears rudimentary, the fact that it was sketched so long ago is significant, suggesting the existence of modern cognitive abilities in our species, Homo sapiens, during a time known as the Middle Stone Age, the researchers said.

The cross-hatched design drawn with ochre, a pigment used by our species dating back at least 285,000 years ago, consists of a set of six straight lines crossed by three slightly curved lines. The coarse-grained stone flake measures about 1-1/2 inches (38.6 mm) long and 1/2-inch (12.8 mm) wide.

"The abrupt termination of all lines on the fragment edges indicates that the pattern originally extended over a larger surface. The pattern was probably more complex and structured in its entirety than in this truncated form," said archaeologist Christopher Henshilwood of the University of Bergen in Norway and the University of the Witwatersrand in South Africa, who led the research published in the journal Nature.

"We would be hesitant to call it art. It is definitely an abstract design and it almost certainly had some meaning to the maker and probably formed a part of the common symbolic system understood by other people in this group," Henshilwood added.

Other Blombos Cave artifacts of similar age included ochre pieces engraved with abstract patterns resembling the one drawn on the stone as well as ochre-covered shell beads. Blombos Cave artifacts dating from 100,000 years ago included a red ochre-based paint.

"All these findings demonstrate that early Homo sapiens in the southern Cape used different techniques to produce similar signs on different media," Henshilwood said. "This observation supports the hypothesis that these signs were symbolic in nature and represented an inherent aspect of the advanced cognitive abilities these early African Homo sapiens, the ancestors of all of us today."

Homo sapiens first appeared more than 315,000 years ago in Africa, later trekking to other parts of the world.

Figure S1: The article used in the control condition.

REUTERS

America celebrates July 4 with hot dogs, banners and barbecues

Americans marched in star-spangled parades, ran relay races, gathered for fireworks shows and crowned a new world hot dog eating champion as they celebrated Independence Day in traditional style on Saturday.

A woman waves an American flag as she rides in an antique pickup truck through Barnstable Village on Cape Cod, during the annual Fourth of July Parade celebrating the country's Independence Day, in Barnstable, Massachusetts, July 4, 2015. REUTERS/Mike Segar

Possible security threats, wildfires in the West and rainy weather on the East Coast apparently did little to dampen the spirits of celebrants decked out in red, white and blue from their headbands to their shoelaces.

Crowds at Boston's Old State House erupted in applause and cannons shot out tri-color confetti after the annual July Fourth reading of the Declaration of Independence.

Meskie Hyman, 11, exuded patriot pride in a star-spangled shirt and a hairband with two American flags that fluttered under cloudy skies in Maplewood, New Jersey.

"I love that it's a free country and we have the right to speak. It lets us see everyone's potential and find our heroes," she said.

The crowd in Washington, D.C. watched a parade of brass bands, law enforcement motorcycle units, high school drum majorettes, antique cars and police and military detachments.

The National Mall, an open area west of the Capitol and site of a fireworks display expected to draw hundreds of thousands, was soggy from steady rain that had stopped by the time the parade started. Among them was Kearston Andrews, 26, who had travelled with her family from near Gainesville, Florida, and said security concerns had not affected them.

Even after promising an increased presence at weekend events, Massachusetts State Police maintained a holiday spirit, tweeting with a colloquial Boston phrase early Saturday "Have a wicked safe 4th of July!"

In a stunning upset on Coney Island, 23-year-old newcomer Matt Stonie grabbed the famed mustard-colored champion's belt after gobbling 62 hot dogs in 10 minutes in the annual Nathan's International Hot Dog Eating Contest.

Stonie, of San Jose, California, ousted long-time champion Joey "Jaws" Chestnut, who still holds a world record of 69 franks eaten in 2013. Miki Sudo hung onto the women's title, downing 38 hotdogs in 10 minutes.

A celebration in Austin, Texas featured country music legend Willie Nelson at an all-day picnic replete with music and drinking. And in Oklahoma, the record to beat at the "Watermelon Seed Spittin' World Championship" in Pauls Valley, about 60 miles (100 km) southeast of Oklahoma City, was 66 feet, 11 inches (20.41 meters), set in 1989 by Jack Dietz of Chicago, organizers said.

Figure S2: The article used in the patriotic priming condition.

REUTERS

U.S.-Iran tensions rise among downing of U.S. military drone, China, Russia affirm Tehran's right to sovereignty

Iran has shot down a U.S. drone which the elite Revolutionary Guards said on Thursday was flying over southern Iran, raising fears that a major military confrontation could erupt between Tehran and the United States.

The purported wreckage of a U.S. military drone is seen displayed by the Islamic Revolution Guards Corps (IRGC) in Tehran, Iran. Tasnim News Agency/Handout via REUTERS

In conflicting statements, Guards website Sepah News said the "spy" drone was brought down over the southern Iranian province of Hormozgan. while a U.S. official said a U.S. Navy drone had been shot down in international airspace over the Strait of Hormuz. The U.S. official's statement resonated with earlier comments by Navy Captain Bill Urban who said no U.S. aircraft were flying over Iran on the day of the downing.

In response to the U.S.'s assertion, Russian Security Council Secretary Nikolai Patrushev stated that Russia had military intelligence showing the U.S. drone in Iranian airspace when it was shot down. This intelligence, Patrushev said, proved Iran was justified in shooting down the drone to protect its sovereignty over Iranian airspace.

Reaffirming Iran's resolve, a senior Iranian security official declared that, "our airspace is our red line and Iran has always responded and will continue to respond strongly to any country that violates it."

Tensions between the two countries have spiked since last year, when the United States reimposed sanctions on Iran after withdrawing from a 2015 nuclear deal between Iran and major world powers.

Among these escalating tensions, China offered strong support to Iran on Friday, with its top diplomat telling Iran's foreign minister that China opposes unilateral U.S. sanctions and acknowledges Tehran's rights to safeguard its interests.

Meeting in Beijing, Chinese State Councillor Wang Yi told Iranian Foreign Minister Mohammad Javad Zarif that given the important and rapidly-evolving situation, the two needed to strengthen communication and coordination.

"China resolutely opposes the U.S. implementation of unilateral sanctions, understands the current situation and concerns of the Iranian side, and affirms Iran's right to safeguard its legitimate interests," China's Foreign Ministry paraphrased Wang as saying.

Tensions might escalate further if the remaining parties to the nuclear deal cannot deliver the promised economic relief. In this case, Iran said it would start enriching uranium at higher levels than agreed to under the terms of the deal.

Figure S3: Article Viewed by Respondents in the Common Enemy Priming Condition

Post-Stratification Modeling Strategy. In this section we explain the methodology we used to obtain results in figures 2, 3 and 4. In both post-stratification and other robustness models, the dependent variable is the amount the participants updated their estimates in response to the estimate made by the bot impersonating a member of the opposing party. If the participant updates their first estimate to move closer to the bot's estimate, the dependent variable is positive. In contrast, if the participant moves away from the bot, the dependent variable is negative. The equation below formalizes the definition of our dependent variable:

$$y_{\mathrm{t}} = \begin{cases} +|x_{\mathrm{p},2} - x_{\mathrm{p},1}| & \text{if } |x_{\mathrm{p},2} - x_{\mathrm{b},1}| \leq |x_{\mathrm{p},1} - x_{\mathrm{b},1}| \\ -|x_{\mathrm{p},2} - x_{\mathrm{p},1}| & \text{otherwise} \end{cases}$$

where y is the dependent variable corresponding to the update and $x_{p,1}$, $x_{p,2}$, $x_{b,1}$ correspond to the player's first and second estimates and bot's first estimate respectively.

To generate our main results, we estimated the mean update toward the bot, accounting for the variation in composition of participants by gender, political knowledge, their initial estimate and whether they correctly identified the political affiliation of the other user on the platform within each prime condition. Effectively, we treated the full sample of participants in our study as the reference population and used post-stratification to compute mean updates that have controlled for variation in strata size across treatments.

Different strata were defined for each possible combination of covariates mentioned above. Next, we discuss the rationale for controlling for each of these variables. A key assumption of our research design is that respondents realize they are collaborating with a member of the other party during the estimation task in our online platform. Though we used strong visual cues to communicate this to respondents, our exit-survey also included a question that asked respondents whether they remembered the party affiliation of the other user on the online platform. Hence, our post-stratification analysis used a binary indicator of whether respondents correctly identified the political affiliation of the other user (i.e. the bot) on the platform, as one dimension of each strata definition.

Because the initial estimates of our bots were programmed to always be 50 percentage points plus a small random noise away from our respondents (on a 0-100 scale), this response may be more or less credible depending upon the respondent's initial guess. For example, respondents who guessed that only 10% of recent immigrants hold a college degree would view an estimate of 60% from the bot, which would be more plausible than another respondent who guessed 48%, which would provoke a guess of 98% from the bot. To adjust for the initial estimates, we converted respondents' first estimates to categorical variables by binning them into intervals of length 25, and used this four-level categorical variable as another dimension of strata definition.

The exit survey included a question about which party has a majority of seats in the Senate. This question was used to test whether the respondent had high political knowledge. Our post-stratification estimate thus includes a binary indicator of whether the respondents correctly answered this question as another factor in strata definition. Finally, there is some evidence in the literature on the existence of the gender gap in political partisanship (*6*). Therefore, we control for the gender of the respondent as a binary variable in our analysis.

The equation below shows how the mean update within each prime treatment can be estimated using post-stratification:

$$\overline{Y}_t = \sum_s \frac{N_s}{N} \overline{y}_{t,s}$$
$$= \sum_s \frac{N_s}{N} \frac{1}{n_{t,s}} \sum_{i \in P_{t,s}} y_i$$

where $\overline{Y}_t$ is the post-stratified estimate of mean update in prime treatment t, N_s is the number of participants in strata s across all treatments, N is the total number of participants across all primes treatments. $y_{t,s}$ is the average update of participants in prime treatment t and strata s. $P_{t,s}$ and $n_{t,s}$ are the set and size of such participants respectively, and finally y_i is the amount participant i updated their initial estimate after observing the bot's estimate. As noted above, the estimate $\overline{Y}_t$ treats all recruited participants as the reference population and weighs the average of each strata based on frequency of that strata in the population.

The point estimates from the equation above are shown in figure 2. The point estimates shown in figures 3 and 4 are computed in a similar fashion with the difference that t would correspond to level of ingroup favoritism or timing with respect to the assassination. The variance of $\overline{Y}_t$ can be estimated using equation below (*7*):

$$\widehat{Var}(\overline{Y}_t) \approx \frac{N - n_t}{n_t N} \sum_s \frac{N_s}{N} S^2_{s,t} + \frac{N - n_t}{n_t^2 (N-1)} \sum_s \frac{N - N_s}{N} S^2_{s,t}$$

where n_t is the number of participants in treatment t and $S^2_{s,t}$ is the sample variance of updates in prime treatment t and strata s. Given the estimate of variance above, we can construct confidence intervals as shown in figures 2, 3, and 4 in the main text. We can further perform post-stratification adjusted t-tests to compare mean updates between various conditions. We used the R package survey to perform the two-sample t-tests between treatments (*7*).

To produce the results in figure 3 of the main text, we created a measure that describes how much Republicans identify with their own party or their ingroup bias. More specifically, we calculated the difference between each respondent's ingroup and out-group feeling thermometer. We then labeled those participants in the top 10th percentile of this measure as strongly partisan with high ingroup bias and the remaining individuals as moderate partisans. An alternative method to measure the strength of partisanship is to use the raw measure of feeling toward the ingroup party. We can then treat the participants in the top 10th percentile of this measure as strongly partisan and the rest as moderately partisan. The 10th percentile cutoff for Republican participants in our data corresponds to 99 out of 100 percentage points for feeling toward Republicans. We use this alternative measure as a robustness check to our result based on the normalized feeling thermometer measure analyzed in the main text. Figure S4 shows the post-stratified difference between these two groups of Republicans.

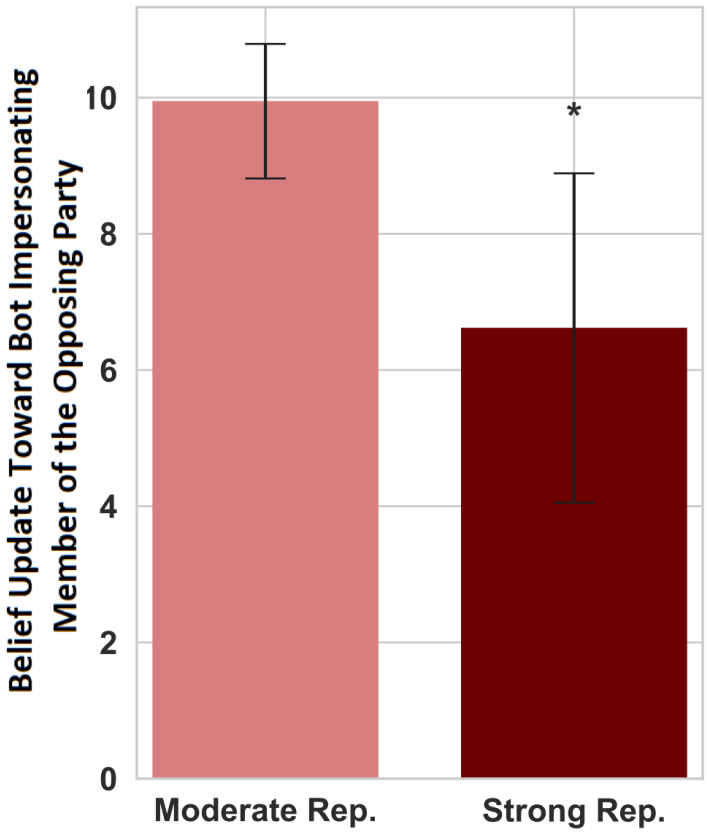


Fig. S4 The extent that Republicans updated their beliefs toward the opinion of the Democrat bot across all prime conditions with a bot impersonating a Democrat during the online task by an alternative measure for strength of partisanship. Vertical axis describes the post-stratified average belief update (in percentage points), where strata are defined by gender, political knowledge, the accuracy of initial guess, and awareness of bot's membership in the opposing party. Strong republicans are defined as those who are in the top 10th percentile of an ingroup feeling thermometer measure. Error bars display 95% confidence intervals. * $p < 0.05$.

SUPPLEMENTARY ANALYSES

Full Results from Models and Robustness Checks. In this section we report full results of the models summarized in the main text of our paper as well as a series of robustness checks for Republicans and Democrats.

Robustness Check on Republicans' Updates: Model 1 – 4

	Dependent variable: first update			
	Model 1	Model 2	Model 3	Model 4
Intercept	15.691***	12.534***	15.808*	15.649***
	(2.131)	(1.586)	(8.743)	(2.311)
Patriotic prime	-0.306	-0.276	-0.028	-0.241
	(1.343)	(1.314)	(1.332)	(1.347)
Common enemy prime	-3.081**	-2.816**	-2.898**	-2.891**
	(1.373)	(1.335)	(1.364)	(1.375)
Manipulation check (1=yes, 0=no)	-1.984	-2.232	-1.871	-2.009
	(1.522)	(1.479)	(1.51)	(1.525)
Male (1=yes, 0=no)	-1.593		-1.503	-1.568
	(1.122)		(1.118)	(1.124)
Respondent's initial estimate [25, 50)	-0.517			
	(1.185)			
Respondent's initial estimate [50, 75)	3.233*			
	(1.88)			
Respondent's initial estimate [75, 100]	5.891			
	(5.018)			
Respondent's initial estimate [0, 10)			-2.164	
			(8.625)	
Respondent's initial estimate [10, 20)			-0.514	
			(8.525)	
Respondent's initial estimate [20, 30)			1.854	
			(8.535)	
Respondent's initial estimate [30, 40)			-1.367	
			(8.547)	
Respondent's initial estimate [40, 50)			-2.115	
			(8.599)	
Respondent's initial estimate [50, 60)			-2.763	
			(8.927)	
Respondent's initial estimate [60, 70)			3.66	
			(8.743)	
Respondent's initial estimate [70, 80)			17.483*	
			(9.593)	
Respondent's initial estimate [90, 100]			3.063	
			(14.705)	
Respondent's initial estimate: distance to 50				0.013
				(0.045)
High political knowledge (1=yes, 0=no)	-3.356**		-3.485**	-3.529**
	(1.49)		(1.478)	(1.489)
Observations	485	512	485	485
R2	0.044	0.014	0.085	0.034
Adjusted R2	0.028	0.009	0.058	0.022
Residual Std. Error	12.101(df=476)	12.2(df=508)	11.916(df=470)	12.141(df=478)
F Statistic	2.739*** (df=8;476)	2.477* (df=3;508)	3.109*** (df=14;470)	2.773** (df=6;478)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table S5. Robustness check of Republicans' belief updates toward the opposing party bot across experimental conditions. Models 1 to 4.

Robustness Check on Republicans' Updates: Model 5 – 8

	Dependent variable: first update			
	Model 5	Model 6	Model 7	Model 8
Intercept	15.084***	15.761***	15.453***	15.775***
	(2.168)	(3.008)	(2.22)	(2.166)
Patriotic prime	0.287	-0.454	-0.29	-0.437
	(1.36)	(1.355)	(1.344)	(1.359)
Common enemy prime	-3.081**	-2.856**	-3.055**	-3.198**
	(1.387)	(1.38)	(1.376)	(1.389)
Age		-0.011		
		(0.044)		
Manipulation check (1=yes, 0=no)	-1.51	-2.41	-1.972	-2.019
	(1.541)	(1.53)	(1.523)	(1.538)
High issue knowledge (1=yes, 0=no)	-0.263			
	(1.475)			
Male (1=yes, 0=no)	-1.568	-1.579	-1.952*	-1.608
	(1.135)	(1.14)	(1.124)	(1.136)
Respondent's initial estimate [25, 50)	-0.165	-0.467	-0.524	-0.548
	(1.195)	(1.192)	(1.186)	(1.2)
Respondent's initial estimate [50, 75)	3.233*	3.42*	3.228*	3.218*
	(1.872)	(1.895)	(1.881)	(1.893)
Respondent's initial estimate [75, 100]	6.609	6.258	5.968	5.821
	(5.435)	(4.962)	(5.027)	(5.043)
High political knowledge (1=yes, 0=no)	-3.731**	-2.52*	-3.349**	-3.398**
	(1.496)	(1.516)	(1.492)	(1.507)
Stubborn bot [reference: friendly bot]			0.427	
			(1.103)	
Observations	469	467	485	478
R2	0.046	0.043	0.044	0.044
Adjusted R2	0.027	0.024	0.026	0.027
Residual Std. Error	11.985(df=459)	11.947(df=457)	12.112(df=475)	12.157(df=469)
F Statistic	2.44** (df=9;459)	2.258** (df=9;457)	2.447*** (df=9;475)	2.67*** (df=8;469)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table S6. Robustness check of Republicans' belief updates toward the opposing party bot across experimental conditions. Models 5 to 8.

Robustness Check on Democrats' Updates: Model 1 – 4

	Dependent variable: first update			
	Model 1	Model 2	Model 3	Model 4
Intercept	8.387***	7.053***	-3.628	3.823*
	(2.302)	(1.827)	(6.659)	(2.307)
Patriotic prime	-0.294	0.002	-0.348	-0.509
	(1.313)	(1.31)	(1.321)	(1.305)
Common enemy prime	1.054	1.061	1.11	1.096
	(1.327)	(1.303)	(1.336)	(1.321)
Manipulation check (1=yes, 0=no)	2.682	2.154	2.621	2.482
	(1.814)	(1.778)	(1.813)	(1.804)
High issue knowledge (1=yes, 0=no)				
Male (1=yes, 0=no)	0.898		0.791	0.854
	(1.077)		(1.082)	(1.071)
Respondent's initial estimate [25, 50)	-2.097*			
	(1.235)			
Respondent's initial estimate [50, 75)	-4.208***			
	(1.532)			
Respondent's initial estimate [75, 100)	-3.015			
	(4.034)			
Respondent's initial estimate [0, 10)			12.727*	
			(6.834)	
Respondent's initial estimate [10, 20)			11.978*	
			(6.314)	
Respondent's initial estimate [20, 30)			12.396*	
			(6.313)	
Respondent's initial estimate [30, 40)			10.354	
			(6.293)	
Respondent's initial estimate [40, 50)			8.214	
			(6.334)	
Respondent's initial estimate [50, 60)			7.261	
			(6.45)	
Respondent's initial estimate [60, 70)			9.724	
			(6.449)	
Respondent's initial estimate [70, 80)			7.501	
			(6.871)	
Respondent's initial estimate: distance to 50				0.152***
				(0.046)
High political knowledge (1=yes, 0=no)	-0.283		-0.27	-0.42
	(1.455)		(1.461)	(1.445)
Observations	530	554	530	530
R2	0.022	0.004	0.034	0.028
Adjusted R2	0.007	-0.001	0.01	0.017
Residual Std. Error	12.357(df=521)	12.507(df=550)	12.34(df=516)	12.298(df=523)
F Statistic	1.479 (df=8;521)	0.805 (df=3;550)	1.411 (df=13;516)	2.492** (df=6;523)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table S7. Robustness check of Democrats' belief updates toward the opposing party bot across experimental conditions. Models 1 to 4.

Robustness Check on Democrats' Updates: Model 5 – 8

	Dependent variable: first update			
	Model 5	Model 6	Model 7	Model 8
Intercept	3.823*	8.655***	11.227***	8.7***
	(2.307)	(2.341)	(3.208)	(2.361)
Patriotic prime	-0.294 -0.004	-0.506	-0.281	-0.245
	(1.325)	(1.338)	(1.314)	(1.319)
Common enemy prime	1.054	0.923	1.073	1.372
	(1.351)	(1.361)	(1.329)	(1.333)
Age		-0.063		
		(0.053)		
Manipulation check (1=yes, 0=no)	2.409	2.501	2.75	2.405
	(1.845)	(1.834)	(1.818)	(1.843)
High issue knowledge (1=yes, 0=no)	-1.447			
	(1.419)			
Male (1=yes, 0=no)	0.898	0.458	0.86	0.75
	(1.089)	(1.12)	(1.08)	(1.084)
Respondent's initial estimate [25, 50)	-2.299*	-2.167*	-2.142*	-2.347*
	(1.246)	(1.257)	(1.238)	(1.246)
Respondent's initial estimate [50, 75)	-4.639***	-4.155***	-4.214***	-4.218***
	(1.553)	(1.557)	(1.533)	(1.533)
Respondent's initial estimate [75, 100)	-3.102	-1.891	-3.189	-3.1
	(4.018)	(4.256)	(4.047)	(4.023)
High political knowledge (1=yes, 0=no)	-0.079	-0.158	-0.29	0.128
	(1.484)	(1.478)	(1.456)	(1.459)
Stubborn bot [reference: friendly bot]			-0.655	
			(1.083)	
Observations	514	510	530	520
R2	0.028	0.024	0.023	0.023
Adjusted R2	0.01	0.007	0.006	0.008
Residual Std. Error	12.294(df=504)	12.381(df=500)	12.364(df=520)	12.313(df=511)
F Statistic	1.595 (df=9;504)	1.379 (df=9;500)	1.354 (df=9;520)	1.504 (df=8;511)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table S8. Robustness check of Democrats' belief updates toward the opposing party bot across experimental conditions. Models 5 to 8.

Model 1: Full Results from Figure 2 in Main Text. Model 1 reports the full results of the first two models of updating behavior among Republicans (Table S5) and Democrats (Table S7). This model includes binary indicators that describe the patriotic and common enemy primes. A key assumption of our research design is that respondents realize they are collaborating with a member of the other party during the estimation task in our online platform. Though we used strong visual primes to communicate this to respondents, our exit-survey also included a question that asked respondents whether they remembered the party affiliation of the other user on the online platform. Our models thus include a binary indicator of whether respondents correctly identified the political affiliation of the other user on the platform.

Model 2: Sensitivity of Results to Respondent's Initial Estimate. Model 2 assesses the sensitivity of our findings to heterogeneity in the initial estimates provided to the question task on our online platform (Republicans, Table S5; Democrats, Table S7). Because the initial estimates of our bots were programmed to always be 50 points away from our respondents (on a 0-100 scale), this response may be more or less credible depending upon the bot's initial guess. For example, respondents who guessed that only 10% of recent immigrants hold a college degree would view an estimate of 60% from the bot, which would be more plausible than another respondent who guessed 48%, which would provoke a guess of 98% from our bots. In order to assess the sensitivity of our results to such bot behavior, we binned respondents according to their initial estimate range and reported these categorical measures in Model 2 in Tables S5 and S7. As these models show, the effects discussed in the main text of our article remain despite the inclusion of these additional indicators in this model.

Model 3: Pre-Existing Knowledge of Issue in Estimation Task. In our study's exit-survey, we asked whether the players are familiar with the issue in the estimation task (immigration). We collected this measure in order to determine whether people with more knowledge about the issue may be less likely to update their estimates in response to the bot regardless of priming condition or political party. Model 3 in tables S5 and S7 adds a binary indicator of familiarity with the issue, showing that the inclusion of this additional indicator does not substantively alter the results we report in the main text of our article.

Model 4: Age Effects. There is some evidence in the scholarly literature on political polarization that political beliefs become more resistant to change across the life course. For this reason, we included a continuous measure of respondent age that we report in Model 4 in Tables S5 and S7. Once again, the results we report in the main text of our article hold despite the inclusion of this additional indicator.

Model 5: Variation in Bot Behavior. In order to capture real-world variation in how members of opposing parties might interact with each other on our platform, our study randomized respondents to perform the estimation with a "friendly" bot or a "stubborn bot." While the former always updated towards the respondent's estimate, the latter held fast to its own initial estimate. To examine whether the behavior of the bot interacts with our treatment effects, Model 5 in tables S6 and S8 reports interactions between the type of bot each respondent interacted with and each prime. We observed no significant effects of this interaction for any of our models and the main effects of the common enemy prime remain consistent among Republicans, as before.

Model 6: Strength of Partisanship. Model 6 reports the full model reported in figure 3 in the main text of our article, which analyzes heterogeneous treatment effects according to the strength of partisanship among Republicans and Democrats. We measured the strength of partisanship using the feeling thermometer measures we collected in our study's exit-survey. The binary indicator reported in Model 6 describes respondents whose feelings towards their own party are in the top 10% of favorability ratings (on a 0-100 point scale). As Tables S6 and S8 show, we observe significant negative updating behavior (i.e., less cooperation with the bot impersonating a member of the opposing party) among Republicans, but not Democrats. This effect is robust to other cut-off points, though the effect becomes weaker when strong partisans are defined as those in the top 20 or 30% of respondents (in terms of strength of affection towards their own party).

Model 7: Inconsistent Reporting of Party Identification. Because we collected respondents' political identification from the panel data originally used to identify respondents as well as on the

platform itself, we were able to identify respondents who inconsistently reported their party identification. Though it is possible that some respondents changed political parties in the interim period, it is considerably more likely that these respondents were either responding expressively or were not paying attention to the online platform's onboarding questions. As an additional robustness check, Model 7 in Tables S6 and S8 removes these respondents from the model, producing results that are nearly identical to those we report in the main text of our article.

Model 8: Main controls, where we keep the last position of the slider if the player did not enter for the second round (Republicans, Table S6; Democrats, Table S8). For those players who do not enter the second round, we use the last position of the slider as their responses for 64 Republicans and 66 Democrats. After we add these individuals, the effect of enemy priming on the republicans disappears.

Updating Behavior Beyond the First Round. Thus far we have only discussed the updating behavior of respondents after they viewed the estimate of the bot impersonating a member of the opposing party for the first time. This is because the most consequential effect of exposure to a member of the opposing party is likely to occur during the initial exposure. Yet our study allowed respondents to update their estimates during two subsequent rounds of updating in response to the bot's estimates. Even though exposure to party affiliation during these rounds is redundant, we analyzed updating behavior among Republicans and Democrats across each of these rounds but observed negligible or small treatment effects in so doing. Between the second and third round of updating, we observed no significant differences in the updating behavior of Republican respondents ($p > 0.05$, $N = 464$) and a small negative updating effect among Democrats who were exposed to the common enemy prime ($p < 0.05$, $N= 504$) but not the patriotic prime ($p > 0.05$, $N = 504$). Between the third and fourth round of updating, we observed no statistically significant differences in updating behavior among either Republicans ($p > 0.05$, $N = 441$) or Democrats ($p > 0.05$, $N = 487$).

Robustness check for the findings about strong partisans. As an additional robustness check for the findings reported in figure 3, we employ regression models with variations on the number of bins. Similar to the results we reported about the effect of partisanship using two different measures of party identification in figures 3 and S4, Table S9 below also shows that the effect of partisanship is consistently negative and significant for our two different measures of ingroup bias. The effect is stronger with 10 bins in the initial estimate than 4 bins, however.

Robustness Check for the Strength of Partisanship as in Figure 3 (Dependent variable: First update)

	Republicans' feel		In-group bias	
	(4 bins)	(10 bins)	(4 bins)	(10 bins)
Manipulation check (1=yes)	-2.012	-1.976	-2.073	-2.042
	(1.528)	(1.516)	(1.529)	(1.518)
Male (1=yes)	-1.908*	-1.807	-2.035*	-1.956*
	(1.121)	(1.116)	(1.125)	(1.119)
Initial estimates [25,50]	-0.809		-0.830	
	(1.200)		(1.201)	
Initial estimates [50,75]	2.444		2.550	
	(1.902)		(1.893)	
Initial estimates [75,100]	6.876		6.576	
	(5.063)		(5.045)	
Initial estimates [0,10]		-1.118		-3.288
		(8.631)		(8.660)
Initial estimates [10,20]		0.653		-1.409
		(8.538)		(8.558)
Initial estimates [20,30]		2.512		0.545
		(8.546)		(8.580)
Initial estimates [30,40]		-0.458		-2.663
		(8.556)		(8.600)
Initial estimates [40,50]		-1.798		-3.741
		(8.607)		(8.658)
Initial estimates [50,60]		-3.092		-5.096
		(8.926)		(8.982)
Initial estimates [60,70]		4.168		2.313
		(8.750)		(8.792)
Initial estimates [70,80]		18.514*		16.041*
		(9.615)		(9.634)
Initial estimates [90,100]		7.833		2.258
		(14.812)		(14.754)
High political knowledge (1=yes)	-3.336**	-3.380**	-3.270**	-3.374**
	(1.491)	(1.477)	(1.491)	(1.478)
Strong partisan by:				
In-group bias (1=yes)			-3.475*	-3.824**
			(1.857)	(1.833)
Republicans' feel (1=yes)	-3.044*	-3.52**		
	(1.694)	(1.678)		
Constant	15.318***	14.663*	15.347***	16.783*
	(2.031)	(8.735)	(2.030)	(8.808)
Observations	484	484	484	484
R^2	0.038	0.082	0.039	0.081
Adjusted R^2	0.024	0.056	0.025	0.056

SE in parentheses: *p<0.10, **p<0.05, *** p<0.01

Table S9. Robustness check for the strength of partisanship as in figure 3.

Robustness check for the effect of the Iran crisis. Table S10 below presents a robustness check on the results reported in figure 4. The binary indicator reported in Table S10 is an indicator for whether the experiment is conducted before or after the Iran Crisis. As Table S10 shows, we observe Republicans were significantly less likely to cooperate with Democrats after the assassination than before this event with 10 bins. The effect is weaker with 4 bins, however.

Robustness Check for the Effect of the Iran Crisis as in Figure 4
(Dependent variable: First update)

	4 bins	10 bins
After Iran crisis (1=yes)	-2.006*	-2.342**
	(1.125)	(1.111)
Manipulation check (1=yes)	-1.82	-1.771
	(1.523)	(1.510)
Male (1=yes)	-2.130*	-2.085*
	(1.130)	(1.124)
Initial estimates [25,50]	-0.479	
	(1.187)	
Initial estimates [50,75]	3.132*	
	(1.883)	
Initial estimates [75,100]	6.647	
	(5.046)	
Initial estimates [0,10]		-3.033
		(8.64)
Initial estimates [10,20]		-1.268
		(8.543)
Initial estimates [20,30]		1.023
		(8.551)
Initial estimates [30,40]		-2.126
		(8.567)
Initial estimates [40,50]		-2.978
		(8.616)
Initial estimates [50,60]		-4.397
		(8.94)
Initial estimates [60,70]		3.254
		(8.75)
Initial estimates [70,80]		16.840*
		(9.601)
Initial estimates [90,100]		4.388
		(14.703)
High political knowledge (1=yes)	-2.974**	-3.026**
	(1.500)	(1.485)
Constant	15.444***	16.724*
	(2.046)	(8.793)
Observations	485	485
R^2	0.038	0.082
Adjusted R^2	0.024	0.056

*SE in parentheses: *p<0.10, **p<0.05, *** p<0.01*

Table S10. Robustness check for the effect of the Iran crisis as shown in Figure 4.

Supporting our Mechanism via Exit Survey Results. Immediately after the main experiment, participants from all conditions were invited to complete an exit survey, in which they were asked to indicate the extent to which they identified with (i) being American, with (ii) Republicans, and with (iii) Democrats. Specifically, they were asked to answer the following questions.

How strongly do you agree or disagree with the following statements?

-I identify with Americans
Options: Strongly Disagree, Disagree, Neither Disagree or Agree, Agree, Strongly Agree

-I identify with Democrats
Options: Strongly Disagree, Disagree, Neither Disagree or Agree, Agree, Strongly Agree

-I identify with Republicans
Options: Strongly Disagree, Disagree, Neither Disagree or Agree, Agree, Strongly Agree

These exit survey questions provide a supplementary set of outcome measures that allow us to provide additional evidence for our proposed mechanism. This proposed mechanism suggests that the common enemy article primes American identity, but that due to polarization, Republicans and Democrats differ in their representation of American identity, which is exclusive toward each other. Specifically, we predict that Republicans will be especially prone toward a view of American identity that increases, rather than decreases, receptivity toward and inclusion of Democrats, given a recent nationally representative survey which shows that Republicans are significantly more likely to view Democrats as un-American and as threats to the nation, compared to Democrats' attitudes toward Republicans (*8*). These survey measures allow us to validate the plausibility of this mechanism by measuring (i) whether the common enemy article primed American identity, (ii) whether this activation of American identity was similar or different among Republicans and Democrats, and (iii) whether this activation of American identity translated into the inclusion or exclusion of increased identification with the other party.

First, fig. S5 shows that across all conditions Republicans identified with being "American" significantly more strongly than Democrats (N=893, $p<0.0001$, $\chi 2$ test), consistent with prior research showing that Republicans exhibit higher overall levels of patriotism (*9,10*). Next, fig. S6*A* shows that Republicans identified significantly more strongly with being American after reading the common enemy article as compared to Republicans who read the neutral control article (N=274, $p=0.05$, $\chi 2$ test). Yet, fig. S6*B* shows that this increase in Republicans' identification with being American did not co-occur with an increase in the extent to which they identified with Democrats (N=274, $p=0.62$, $\chi 2$ test); instead, Republicans appear to have increased their sense of national identity in an exclusive manner that did not include increased cross-party identification. Meanwhile, fig. S7 shows that Democrats were significantly less likely to identify with being American after reading the common enemy article as compared to Democrats who read the neutral control article (N=311, $p<0.01$, $\chi 2$ test), suggesting that the common enemy article had the opposite effect on Democrats' identification with the national superordinate category as compared to Republicans. These results are consistent with our proposed mechanism, as quoted in the main text from Rutchik & Eccleston (2010) (*11*), who argue that "when there is a perception that subgroups do not have a shared conception of the superordinate group, appeals to the common ingroup identity made by outgroup members are likely to backfire" (pg. 111). In particular, we find evidence that Republicans and Democrats have different (and in some respects, opposing) conceptions of the superordinate category of "American" and its relation to the common enemy threat in our experiment; moreover, our results suggest that the activation of this superordinate category – and particularly among Republicans – did not lead to cross-party inclusion and identification, as popular theory suggests. Instead, our findings are consistent with the view that in highly polarized contexts, the activation of a superordinate identity may not lead to cross-party identification, and may actually exacerbate cross-party tensions (*11,12*).

This mechanism is further supported by the finding - presented in fig. S8 - that Republicans identified significantly more strongly as Americans after the assassination of Suleimani, an event that likely increased the salience of Iran as a common enemy (N=428, p=0.05, $\chi 2$ test). There was no significant difference in Democrats' strength of identification with Americans before and during the Iran crisis (N=465, p=0.12, $\chi 2$ test). These findings are consistent with popular news accounts during the Iran crisis, which documented many prominent Republican leaders who accused Democrats of unduly lamenting the death of a dangerous Iranian general, insinuating an alliance or sympathy between Democrats and this common enemy. It is thus plausible that this Iranian threat could have increased the salience of Republicans' sense of American identity, which excluded Democrats and framed them as being just as threatening as Iran, and indeed as potentially complicit in this common enemy threat. This is also consistent with our behavioral outcomes in fig. 4 from the main text, which show that after the Iran crisis, Republicans became even less willing to incorporate information from Democrats in a cooperative estimation task.

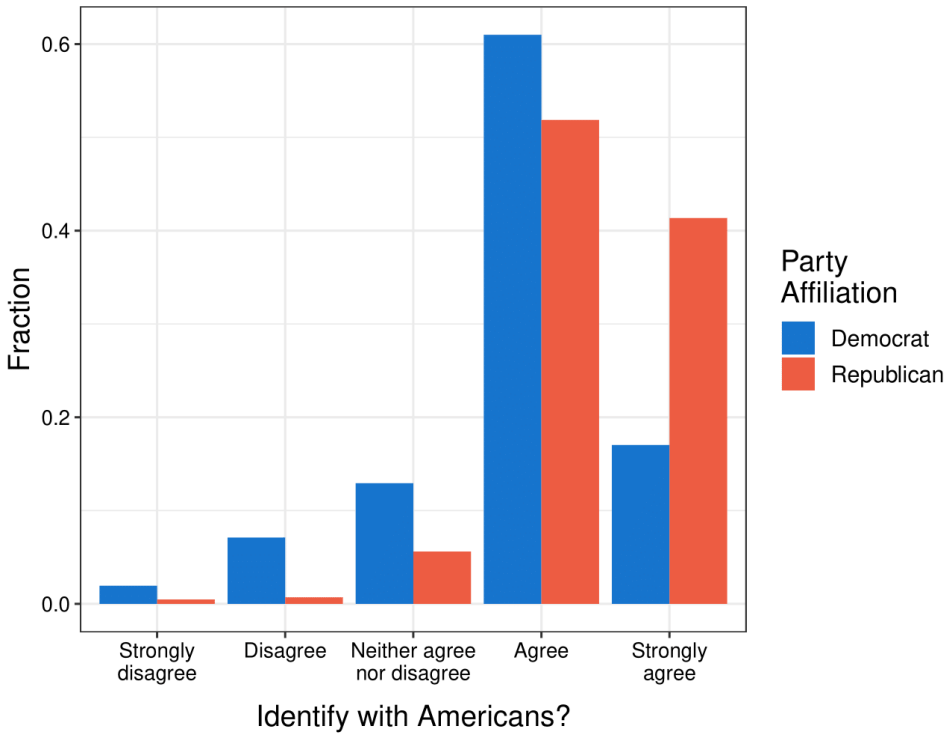


Fig. S5. Exit survey results showing the strength of identification with being American for both Democrats and Republicans; results are collapsed across conditions.

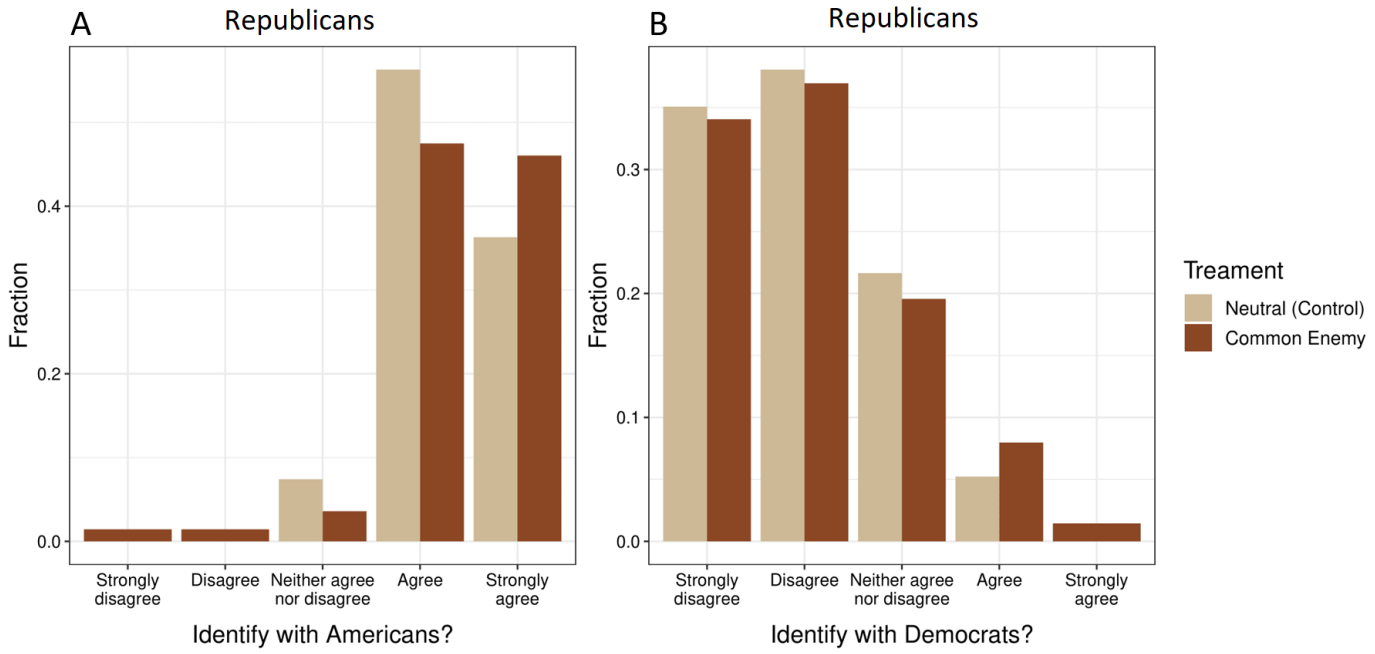


Fig. S6. Exit survey results showing Republicans' strength of identification with (A) Americans and (B) Democrats, within both the neutral (control) condition and the common enemy condition.

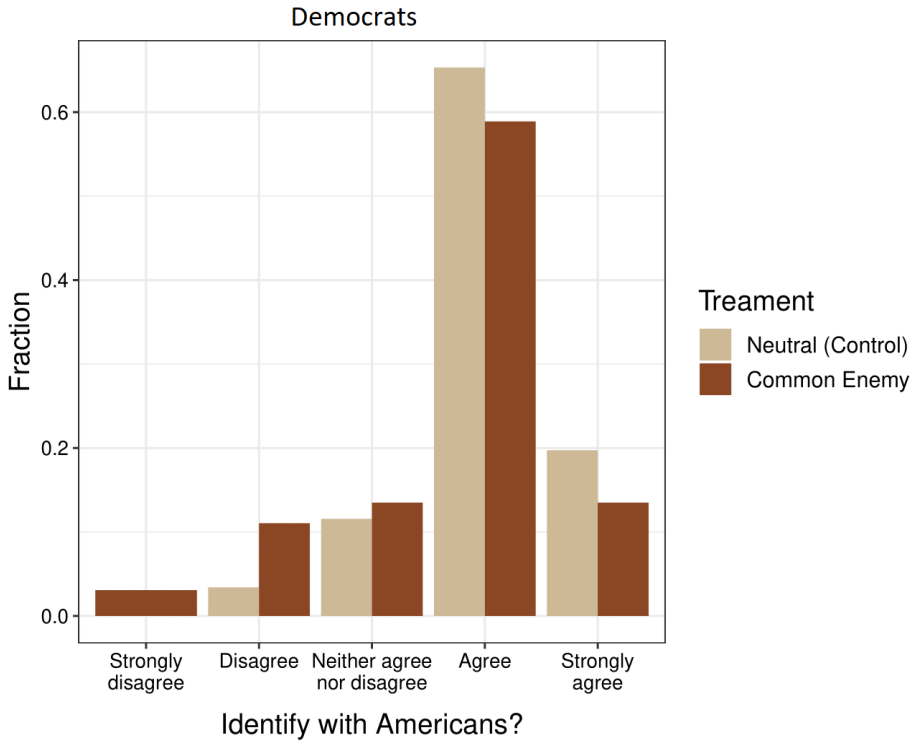


Fig. S7. Exit survey results showing Democrats' strength of identification with Americans within both the neutral (control) condition and the common enemy condition.

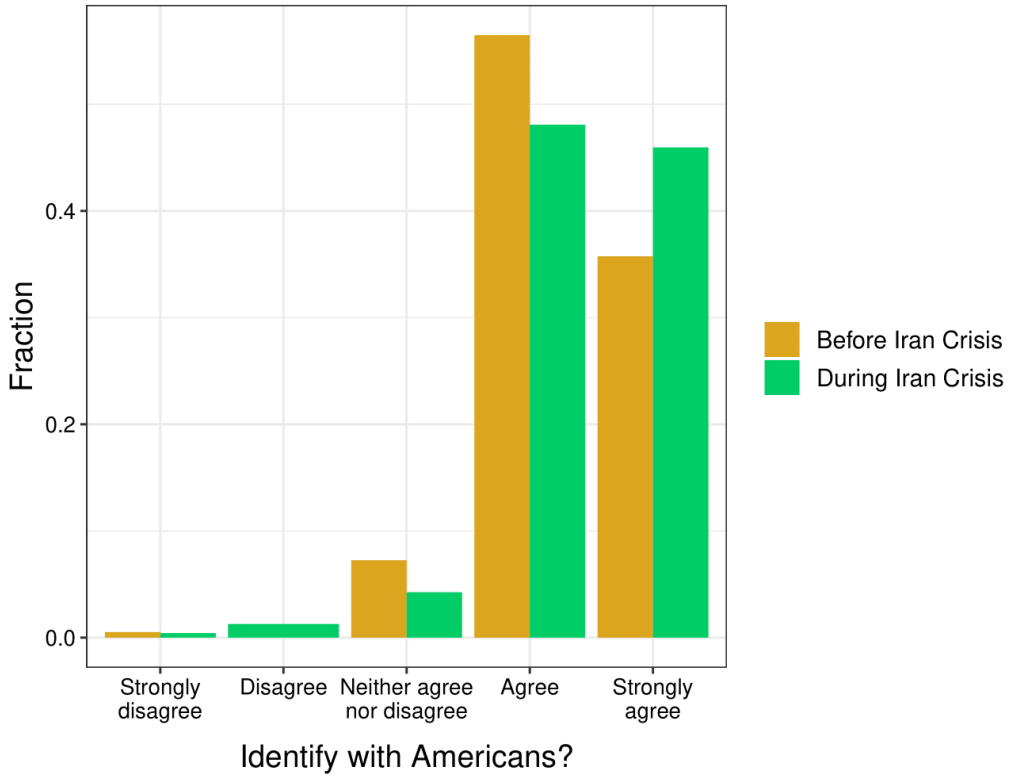


Fig. S8. Exit survey results showing Republicans' strength of identification with Americans before and during the Iran crisis. Results are collapsed across priming conditions.

SUPPLEMENTARY DISCUSSION

Deviations from Preregistration. The full text of our pre-registration is available via the Open Science Framework here: https://osf.io/eupbv/?view_only=686c0ad6d24d4cf498156c4710ad47c0. Though we focused upon our pre-registered hypotheses about asymmetric polarization in the main text of our article, some of our predictions about this process were not supported by our analysis. For example, we initially hypothesized we would observe asymmetric polarization across both of our treatment conditions (patriotic and common enemy priming), but we only found evidence of this process in the latter condition. We also pre-registered rival hypotheses that were not supported by our findings. Based on the common ingroup identity model, we highlighted the possibility that participants in our common enemy and patriotic priming conditions would incorporate more information from the outgroup bot; this hypothesis was not supported ($N=1015$, $p > 0.05$). We made a separate set of predictions about how people would behave in response to different types of behavior by our bot. More specifically, we expected participants to cooperate more with the friendly bot (that updates its estimates towards respondents) than the stubborn bot (which does not). We did not find consistent evidence of this behavior. Since it takes more than one round for the player to discern the bot's behavior (as either friendly or stubborn), we compare the difference in updates between the first and final estimating rounds via models that include the same controls as those reported in figure 2 from the main text of our article. These models indicate Republican participants are more likely to collaborate with the stubborn bot than the friendly bot ($p < 0.01$, $N = 471$), but we observed no significant

differences in the updating behavior of democrats based on the type of bot they encountered on our online platform ($p > 0.05$, $N = 521$). In contrast, our pre-registration statement predicted two effects of partisanship. First, we tested competing hypotheses about whether Democrats or Republicans were more accommodating of the stubborn bot. However, there was no difference between Republicans and Democrats in updating towards the stubborn bot (total update from the first to final round; $t(974)=3.17$, $p > 0.05$).

Our analyses of the data also deviated from our pre-registration statement in several significant ways. In this document, we stated that we would evaluate our hypotheses by examining updating behavior within each round of estimating and between the first and last round. Yet we did not realize that a) information about the party affiliation of the bot impersonating a member of the opposing party would be redundant after the first round of updating; and b) people's experience of the bot's initial guess would be more consequential than subsequent rounds because of the considerable heterogeneity in respondents' initial estimates. For these reasons, we elected to focus upon the updating behavior of respondents between the first two rounds and reported the other models in our supplementary materials (above). In our pre-registration document, we also stated that we would control for the type of bot (friendly or stubborn) in our analyses. Because we did not observe significant interactions between bot behavior and our priming conditions— and because we adopted the aforementioned binned measure of respondent's initial estimate, which interacts with bot behavior— we elected not to include this measure in our models. We also pre-registered a mediation analysis. More specifically, we included measures of personal affect (PANAS), and ingroup and outgroup identification as potential mediators of any effect to the patriotic and common enemy primes. When conducting mediation analyses, these mediators did not predict cooperation with the outgroup bot. In retrospect, we concluded that these measures may not have been effective because they were administered after the prime and estimating task had been completed, rather than immediately after the prime. Finally, we did not pre-register our analyses of the pre and post-Iran crisis effects of the common enemy prime because it was not possible to anticipate such unprecedented events at the time we wrote our pre-registration statement.

Additional Study Limitations. Despite its many important contributions, our study also has several important limitations. First, we only analyzed the effect of common enemy priming in one country, and during a single historical period. It is possible that common enemies may reduce political polarization in other less polarized periods, or in countries that are not dominated by two parties. Second, our measure of social learning was limited to an anonymous, online setting where respondents could not experience shared emotions about the common enemies they were confronted with. It is possible that the realization of shared fear is a key mechanism of depolarization among members of rival political groups. Finally— though we attempted to identify a common enemy prime that was least likely to prime partisan positions— it is possible that other shared threats, such as large-scale public health pandemics, might create different effects. Despite these limitations, we hope that our study will provide a helpful basis for future research on political polarization, social identity, and the fledgling field of computational social science which routinely employs the types of online experiments used in this study to advance our understanding of large-scale human behavior.

SUPPLEMENTARY REFERENCES

1. Litman, Leib, Jonathan Robinson, and Tzvi Abberbock. "TurkPrime.Com: A Versatile

Crowdsourcing Data Acquisition Platform for the Behavioral Sciences." *Behavior Research Methods* 49, no. 2 (April 1, 2017): 433–42. https://doi.org/10.3758/s13428-016-0727-z.

2. Hauser, David J., and Norbert Schwarz. "Attentive Turkers: MTurk Participants Perform Better on Online Attention Checks than Do Subject Pool Participants." *Behavior Research Methods* 48, no. 1 (March 1, 2016): 400–407. https://doi.org/10.3758/s13428-015-0578-z.
3. Watson, David, Lee Anna Clark, and Auke Tellegen. "Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales." *Journal of Personality and Social Psychology* 54 (1988): 1063–70. https://doi.org/10.1037/0022-3514.54.6.1063.
4. Budak, Ceren, Sharad Goel, and Justin M. Rao. "Fair and Balanced? Quantifying Media Bias through Crowdsourced Content Analysis." *Public Opinion Quarterly* 80, no. S1 (January 1, 2016): 250–71. https://doi.org/10.1093/poq/nfw007.
5. Levendusky, Matthew S. "Americans, Not Partisans: Can Priming American National Identity Reduce Affective Polarization?" *The Journal of Politics* 80, no. 1 (January 1, 2018): 59–70. https://doi.org/10.1086/693987.
6. Box-Steffensmeier, Janet M., Suzanna De Boef, and Tse-Min Lin. "The dynamics of the partisan gender gap." *American Political Science Review* 98.3 (2004): 515-528.
7. Oberski, Daniel. 2014. "survey: An R package for complex survey analysis of structural equation models." *Journal of statistical software* 57(1): 1-27.
8. Pew Research Center. "Political Polarization in the American Public," 2014. https://www.pewresearch.org/politics/2014/06/12/political-polarization-in-the-american-public/
9. Jost, John T., Jack Glaser, Arie W. Kruglanski, and Frank J. Sulloway. "Political Conservatism as Motivated Social Cognition." *Psychological Bulletin* 129, no. 3 (2003): 339–75. https://doi.org/10.1037/0033-2909.129.3.339.
10. Jost, John T. "The End of the End of Ideology." *The American Psychologist* 61, no. 7 (October 2006): 651–70. https://doi.org/10.1037/0003-066X.61.7.651.
11. Rutchick, Abraham M., and Collette P. Eccleston. "Ironic Effects of Invoking Common Ingroup Identity." *Basic and Applied Social Psychology* 32, no. 2 (May 19, 2010): 109–17. https://doi.org/10.1080/01973531003738353.
12. Klar, Samara. "When Common Identities Decrease Trust: An Experimental Study of Partisan Women." *American Journal of Political Science* 62, no. 3 (2018): 610–22. https://doi.org/10.1111/ajps.12366.