Closed-Mindedness Prevails and Cascades in Online Political Spaces

Michalis Mamakos
University of Warsaw

William Brady
Northwestern University and IPR

Steven Franconeri
Northwestern University and IPR

Galen Bodenhausen
Northwestern University

Eli Finkel
Northwestern University and IPR

Version: February 6, 2026

DRAFT

Please do not quote or distribute without permission.

Abstract

The unwillingness to consider alternative viewpoints hinders the integration of differing perspectives required to resolve contentious problems. Here, the authors conducted the first large-scale behavioral study of closed-mindedness, examining its prevalence and the reactions it receives in online political spaces. They first developed a text classifier of closed-mindedness by fine-tuning a large language model, then applied it to Reddit comments. Closed-mindedness was significantly more prevalent in political than nonpolitical communities. Within political communities, closed-mindedness was more common among comments expressing partisan views, and it predicted social disapproval more strongly when these views misaligned with the community's lean. Closed-minded comments were likelier to elicit replies, and those replies were especially likely to be closed-minded, suggesting a cascade of closed-minded exchanges that hinder constructive discourse. These findings underscore closed-mindedness as a core and self-reinforcing feature of online political discourse, thereby compromising the active consideration of diverse perspectives critical for healthy democratic functioning.

Significance statement

Social media platforms function as the modern public sphere for moral and political discourse. However, constructive discourse is attainable only to the degree that participants are willing to entertain differing perspectives. This report demonstrates that closed-mindedness is prevalent in online political discussions, that it is socially sanctioned when expressing disagreeable views, and that it attracts closed-minded replies. These properties perpetuate a cycle of closed-minded interactions that render online discourse far less likely to lead to solutions for contentious issues. These observations highlight a need to rethink the design of social media platforms to promote receptivity to diverse perspectives.

Closed-mindedness prevails and cascades in online political spaces

Healthy democratic functioning requires the integration of diverse perspectives (Habermas, 1985; Bohman, 2006). When citizens engage with differing views, they can identify common ground on divisive issues (Guttman & Thompson, 1996) or at least gain insight into the reasons behind ongoing disagreements (Price et al., 2002). Deliberative engagement with diverse perspectives can also enhance wisdom and improve collective decision-making (Young, 1996). A prerequisite for such salutary outcomes is that individuals are open to changing their prior views when exposed to contradictory information (Rawls, 1951; Barabas, 2004). However, in partisan political contexts, people tend to be motivated less by curiosity and a desire to learn than by a tendency to reject information that deviates from their initial attitudes (Bolsen et al., 2014), a tendency that can prevent the psychological discomfort of cognitive conflict and uncertainty (Festinger, 1957) and increase status within the ingroup (Mercier & Landemore, 2012). The normative ideal for deliberative democracy is violated when discussions are characterized by closed-mindedness: An unwillingness to have one's knowledge confronted by alternative opinions or inconsistent evidence (Webster & Kruglanski, 1994). The present research asks: How prevalent is closed-mindedness in political discourse, and what are the downstream reactions associated with it?

In the 21st century, social media platforms have become primary public squares (Yarchi et al., 2021), significantly influencing real-world outcomes, such as increasing political polarization and voter turnout (Allcott et al., 2020; Lorenz-Spreen et al., 2023). Despite their impact, it is unclear whether these platforms host discourse that meets the conditions necessary for deliberative democracy to flourish. To our knowledge, no previous research has examined the closed-mindedness of online platforms behaviorally at scale.

The present report provides the first large-scale behavioral test examining the prevalence of closed-mindedness in online political discourse and the downstream reactions it elicits. Previous research suggested that exposure to differing views can even fuel polarization (Bail et al., 2018), thereby contradicting the echo-chamber hypothesis—that partisan animosity on social media is driven by engagement exclusively with like-minded views (Sunstein, 2018). Indeed, online users who opt into discussions with the other side are especially likely to be uncivil, indicating that their intent in engagement may not be constructive discourse (Mamakos & Finkel, 2023). These observations suggest that online political spaces may be strongly characterized by closed-mindedness. We therefore hypothesized that closed-mindedness is prevalent in political (vs. nonpolitical) communities.

The conditions for deliberative democracy suggest that people should be open to differing perspectives even when dialogue falls along partisan lines. However, displays of partisanship often reflect the growing animosity toward the other side (Huddy et al., 2015; Kalmoe & Mason, 2022), suggesting that such displays may also express an unwillingness to consider alternative perspectives. We hypothesized that comments in online political communities are more likely to display closed-mindedness when those comments also display a partisan lean.

How do other people react to closed-mindedness? Such reactions can indicate social approval or disapproval, and may depend on the views that closed-minded comments express. Groups strive to maintain consistency by punishing norm violators (Schachter, 1951; Jetten & Hornsey, 2014). Thus, we hypothesized that when comments display a partisan lean *misaligned* with a community's lean, their closed-mindedness may highlight deviance, *increasing* the likelihood of social sanction (disapproval). Conversely, because groups tend to approve of norm-conformist behavior (Kruglanski & Webster, 1991; Kruglanski et al., 2006), we hypothesized that when comments display a partisan lean *aligned* with a

community's lean, their closed-mindedness may highlight compliance, *decreasing* the likelihood of social sanction.

In addition to social approval, closed-mindedness could predict differences in the likelihood of receiving closed-minded replies. Previous research showed that divisive (Rathje et al., 2021) and dismissive (Heltzel & Laurin, 2024) social media content receives higher engagement. We hypothesized that closed-minded comments are more likely to receive replies, and that these replies may themselves aim to display closed-mindedness. Therefore, closed-mindedness could cascade, leading to a self-perpetuating cycle of closed-minded exchanges that hinder constructive discourse.

In summary, the present report seeks to examine the following hypotheses:

H1. Closed-mindedness is more prevalent in political than nonpolitical communities.

H2. Closed-mindedness is more prevalent among comments displaying a partisan lean.

H3. For comments displaying a partisan lean, closed-mindedness is associated with (a) a *higher* likelihood of social sanction when the lean *misaligns* with the community's lean and (b) a *lower* likelihood of social sanction when the lean *aligns* with the community's lean.

H4. Closed-minded comments are more likely to receive a reply.

H5. Among comments receiving a reply, the closed-minded ones are more likely to receive a closed-minded reply.

A test of these hypotheses requires a rich dataset with information about the context where online discussions take place. To meet this need, we used data from Reddit (Baumgartner et al., 2020), a platform where discussions are hosted in communities (subreddits), with users making posts that can receive comments from other users. Comments can themselves receive comments, resulting in multi-threaded discussions. There are thousands of subreddits, with their topics spanning the spectrum of activities that concern people's everyday lives (Waller & Anderson, 2021). Therefore, using Reddit data

allows us to examine the prevalence of closed-mindedness in political subreddits by comparing them to nonpolitical ones and to identify political subreddits with a known lean.

In addition to information about the context where comments are posted, a study of our hypotheses requires classifying comments as to whether they display closed-mindedness. To our knowledge, no such text classifier exists, and thus, our first goal was to develop one. To this end, we fine-tuned a pre-trained large language model (Jiang et al., 2023), re-training it with validated Reddit comments labeled by human annotators who were extensively trained on the psychological construct of closed-mindedness. This classifier is publicly available at: https://huggingface.co/mamakos/CMClassifier. We outline this methodological contribution in the next section and then present the results related to our hypotheses.

Classifying closed-mindedness

To classify Reddit comments as to whether they display closed-mindedness, we fine-tuned Mistral 7B—a 7-billion-parameter, open-source language model. This model is free to use, time-efficient due to its relatively small size, and performs comparably to larger models (Jiang et al., 2023). Here, we provide an overview of the development of this closed-mindedness classifier, with extensive details available in the Supplementary Material.

To create fine-tuning and validation datasets of comments with binary labels about closed-mindedness, we recruited seven annotators, all undergraduate students. The annotators received instructions and intensive training to classify Reddit comments as closed-minded or not. The fine-tuning dataset consisted of 2,400 comments sampled from left-wing (e.g., r/democrats, r/Liberals), right-wing (e.g., r/Republican, r/Conservative), and nonpolitical (e.g., r/Music, r/movies) subreddits. Binary labels of closed-mindedness were assigned to these comments based on the consensus rating of the annotators, who demonstrated high consistency (ICC = .84). Approximately the same percentage of comments from

both left-wing and right-wing subreddits were classified as closed-minded by the annotators (14% for each subreddit type, with annotators blind to the comments' subreddit), whereas this percentage was lower (4%) for comments from nonpolitical subreddits. The statistics for the validation dataset (which consisted of 3,600 comments) were similar.

Table 1. Examples of comments with a very high and very low estimated probability of displaying closed-mindedness: an unwillingness to have one's knowledge confronted by alternative opinions or inconsistent evidence.

Probability higher than 99%
Don't annoy me with the facts, my mind is already made up.
No one would vote for Trump unless they're uninformed. There's lots of uninformed people out there, though.
It's common sense that abortion is wrong.
That is nothing but fear mongering nonsense.
I'm confused with your post. Did you comment without watching the video? How do you argue against the obvious?
Yeah, people like you don't understand how it works.
Probability lower than 1%
I think the debts are supposed to be paid off by a small tax on stock trades.
He is South Carolina's Democratic candidate for U.S. Senate.
I think you are forgetting Glen Allen Walken (John Goodman). He was a good Republican.
What happened?
If I remember correctly, in economics, we just call it profit maximization or cost minimization.
Well, there goes any free time I might have had.

After the model was fine-tuned, it was validated, with the classification metrics indicating the success of the fine-tuned model in classifying closed-mindedness (Accuracy = .94, F1 = .70). These metrics were similar for comments from left-wing, right-wing, and nonpolitical subreddits. We also assessed divergent validity against other text classifiers, including the PerspectiveAPI classifier of

toxicity, which is widely employed in social media studies (Wulczyn et al., 2017; Falkenberg et al., 2024). Table 1 presents some comments to which our classifier assigned a probability higher than 99% to display closed-mindedness and some comments to which it assigned a probability lower than 1%.

Results

Closed-mindedness is prevalent in political communities

We first examined the hypothesis that closed-mindedness is more prevalent in political than nonpolitical subreddits. We sampled 252,011,417 comments from 10,006 subreddits considered previously in the literature (Waller & Anderson, 2021)—these are the subreddits that capture most of the activity on Reddit. Comments were classified as closed-minded if the probability estimated by our classifier exceeded 0.5. For each subreddit, we computed its proportion of closed-minded comments. We categorized subreddits as political if they appeared in the list of political subreddits devised by Hofmann et al. (2022).

In Figure 1, we observed that political subreddits were vastly more closed-minded than nonpolitical subreddits ($d = 4.68$, $p < .001$). On the average political subreddit, 13% of the comments were closed-minded. On the average nonpolitical subreddit, this proportion was only 2%. We conducted three robustness checks on this difference.
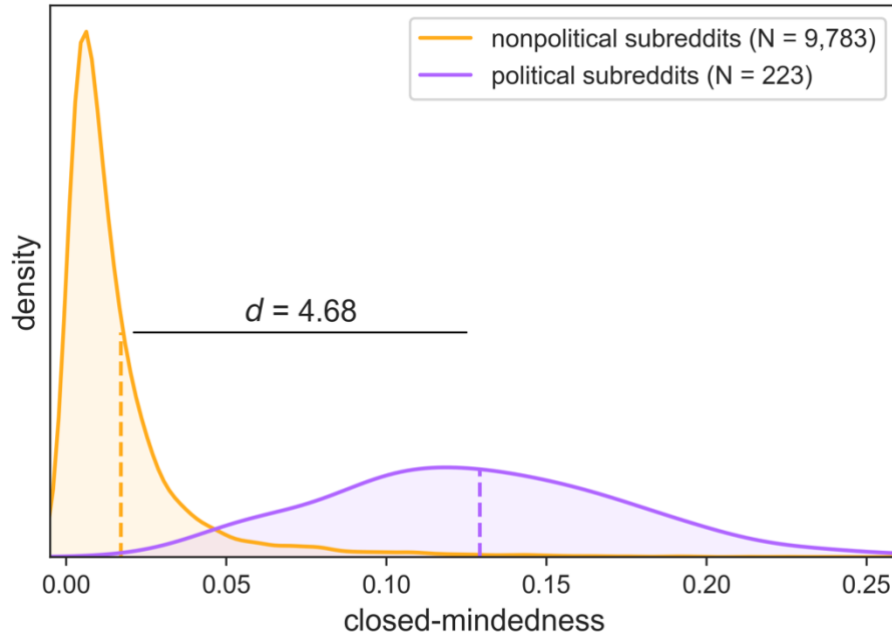
Figure 1. Density plot of the closed-mindedness of nonpolitical and political subreddits. The dashed lines indicate the means.

First, to rule out the possibility that our classifier measures political content rather than closed-mindedness, we repeated this analysis after excluding comments with words about politics (Simchon et al., 2022). Even after this exclusion, political subreddits remained much more closed-minded than nonpolitical ones ($d = 3.91$, $p < .001$). Second, when we compared the political subreddits against 30 subreddit categories (including one about sports) individually, the smallest difference in closed-mindedness was still very large ($d = 2.38$, $p < .001$). Third, we sought to rule out the possibility that our findings reflected differences in toxicity rather than closed-mindedness. In line with previous research (Sun et al., 2021; Mamakos & Finkel, 2023), we found that political subreddits were more toxic than nonpolitical ones, although this difference was much smaller ($d = 1.17$, $p < .001$). Logistic regression results revealed that the subreddits' closed-mindedness was about twice as strong as their toxicity as a predictor of whether they were political. Overall, these results supported our Hypothesis 1, showing that closed-mindedness was much more prevalent in political than nonpolitical subreddits. Further details about these three robustness checks can be found in the Supplementary Material.

In line with previous research suggesting that conservatism is associated with closed-mindedness (Jost et al., 2003; Acosta & Kemmelmeier, 2022), we found that among political subreddits, closed-mindedness was positively correlated with Waller and Anderson's sociometric measure of conservatism ($r = .27$, $p < .001$). However, political subreddits across the whole spectrum of partisanship were much more closed-minded than the nonpolitical ones (see SM Section 2).

Partisan comments are more likely to be closed-minded

Next, we sought to examine Hypotheses 2-5. We sampled 1 million comments at random from political subreddits with a known lean. To test Hypothesis 2, that comments are more likely to be closed-minded when they exhibit a partisan lean, we used GPT-4o to classify comments as to whether they exhibited this lean, i.e., whether they expressed views aligning or misaligning with Democratic or Republican views (vs. not expressing any party-relevant views). We validated with human annotators the ability of GPT-4o to provide these ratings of partisan lean (see SM Section 3).

The results presented in Figure 2 supported our Hypothesis 2 regarding the prevalence of closed-mindedness in comments displaying a partisan lean: Such comments were much more likely to be closed-minded than those not displaying a partisan lean (24% vs. 9%, $p < .001$).
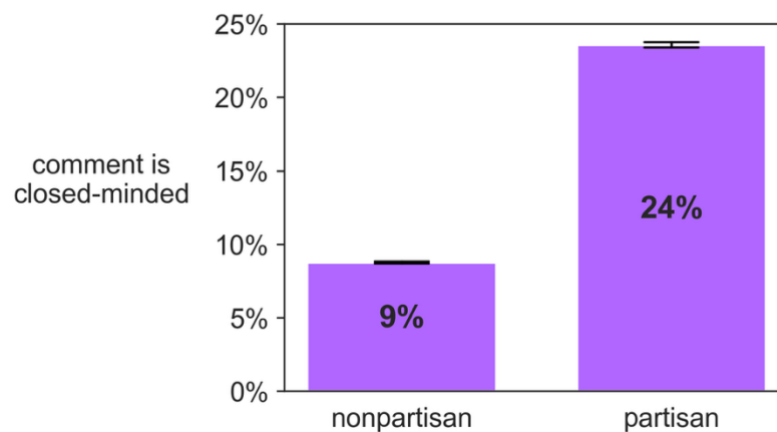

Figure 2. The proportion of comments that displayed closed-mindedness in partisan political subreddits, based on whether they displayed a partisan lean (95% CI).

Closed-minded comments with misaligned partisanship are especially disapproved of

Next, we examined Hypothesis 3, positing that for partisan comments, closed-mindedness predicts differences in social approval based on whether a comment's lean aligns or misaligns with the subreddit's lean. To this end, we considered the comments' approval rating: the number of upvotes minus downvotes. Note that Reddit makes available only this metric rather than the number of upvotes and downvotes separately. Because the distribution of this approval rating was extremely skewed ($M =$ 10, $SD = 58$, *Median* = 3) and remained so even after outlier analyses (see SM Section 4), we dichotomized comments based on whether their rating was negative (more downvotes than upvotes), indicating general disapproval by the subreddit's members.

In Figure 3, we observed that comments with a partisan lean misaligned with the subreddit's lean were more likely to be disapproved when they were closed-minded compared to when they were not closed-minded (24% vs. 15%, $p < .001$). This result supported the part of our Hypothesis 3 about the lean-misaligned comments, that when they display closed-mindedness, they are more likely to receive social sanction. However, contrary to the part of our Hypothesis 3 concerning lean-aligned comments, we observed that also these comments were more likely to be disapproved when they were closed-minded compared to when they were not closed-minded (4% vs. 3%, $p < .001$).

To assess whether closed-mindedness was particularly predictive of social disapproval for misaligned comments, we performed a logistic regression whose dependent variable modeled whether the approval rating was negative. The results of this logistic regression supported the interaction ($OR =$ 1.33, $p < .001$) between closed-mindedness ($OR = 1.36$, $p < .001$) and lean misalignment ($OR = 6.29$, $p < .001$). Therefore, our Hypothesis 3 about closed-mindedness predicting social sanction based on the alignment between a comment's and a subreddit's lean was supported for misaligned comments, which were especially likely to be disapproved when displaying closed-mindedness. Nevertheless, this hypothesis was not supported for comments whose lean aligned with the subreddit's lean. These

comments, too, received higher disapproval when they were closed-minded, although they were rarely

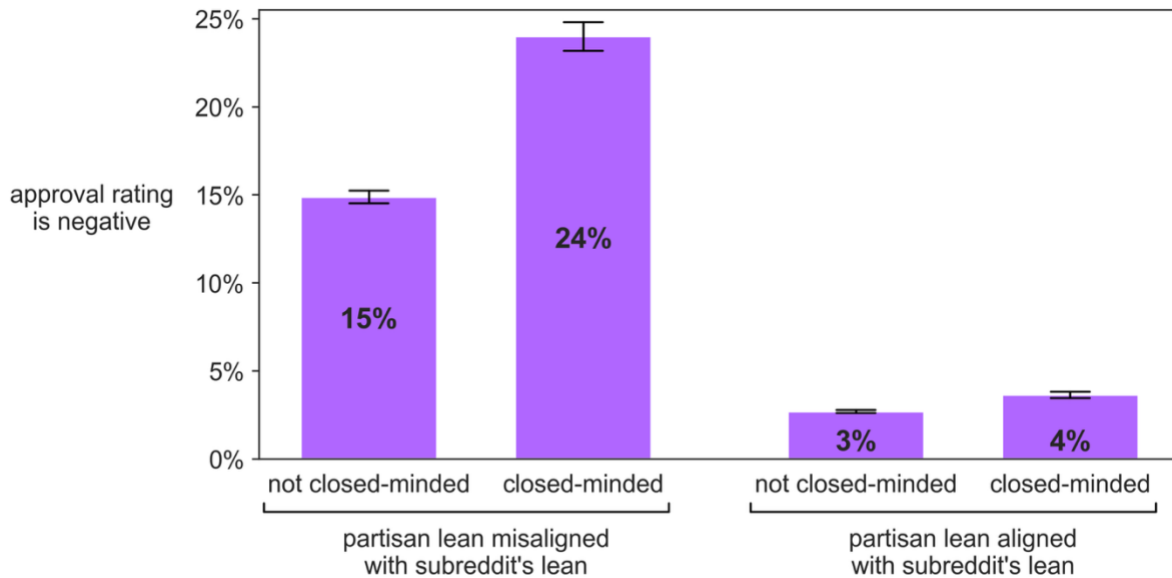disapproved even when displaying closed-mindedness.

Figure 3. The proportion of comments that received a negative approval rating in partisan political subreddits, based on their closed-mindedness and subreddit lean alignment (95% CI).

Closed-minded comments are more likely to receive a (closed-minded) reply

Next, we examined Hypotheses 4 and 5, positing that closed-minded comments are more likely to

receive replies and that these replies are more likely to be closed-minded, respectively. Support for both

these hypotheses would suggest a cascading phenomenon, whereby closed-minded comments amplify a

self-perpetuating cycle of closed-mindedness.

In Figure 4a, we observed among the 1 million comments posted in political subreddits with a

known lean, the likelihood of receiving a reply was higher for comments that displayed closed-

mindedness than those that did not (42% vs. 36%, $p < .001$), thus supporting our Hypothesis 4. To

examine the likelihood that a reply was closed-minded, and because 68% of the comments that received

a reply had exactly one reply (only 3% received five replies or more; see SM Section 5), we simplified

our analysis by considering only the first reply for the remaining 32% of the comments. The results

presented in Figure 4b supported our Hypothesis 5: Replies to comments were more likely to be closed-

minded when the replied-to comments were also closed-minded than when they were not (17% vs. 7%,
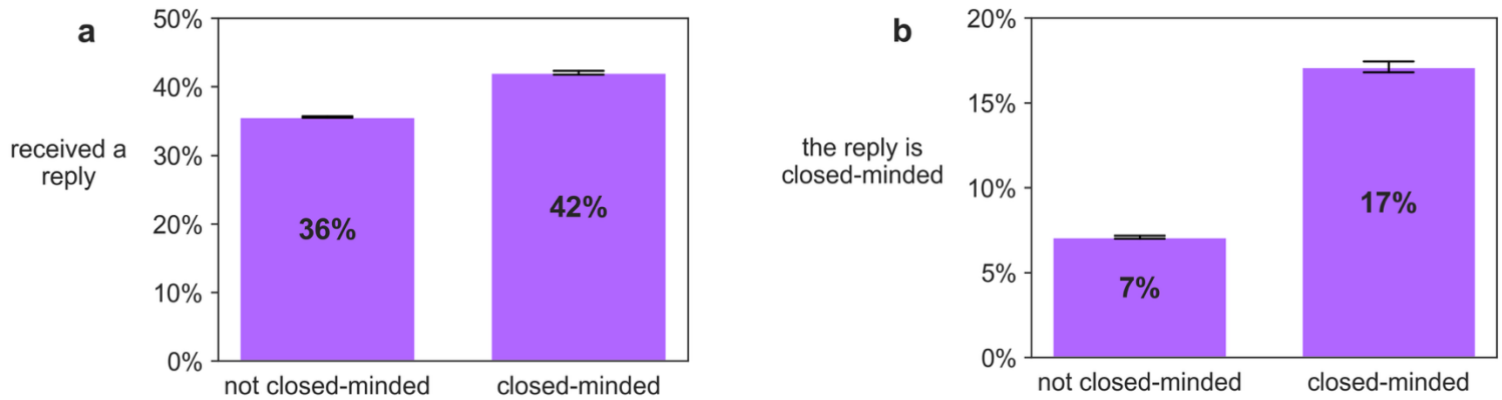
p < .001).

Figure 4. The proportion of comments in partisan political subreddits that (a) received a reply, based on their closed-mindedness, and (b) the proportion of the replies that were closed-minded (95% CI).

To explore the functions that closed-minded replies serve in political communication, we used

GPT-4o to classify the replies as to whether they displayed agreement, disagreement, or neither with the

replied-to comments (see SM Section 7 for validation with human annotators). As shown in Figure 5,

the proportion of replies agreeing with the replied-to comment was nearly the same whether the replies

were closed-minded or not (23% vs. 24%). In contrast, replies were much more likely to disagree with

the replied-to comment when they were closed-minded than when they were not (44% vs. 18%, *p* <

.001). These results suggest that closed-mindedness is a means of highlighting disagreement, without a

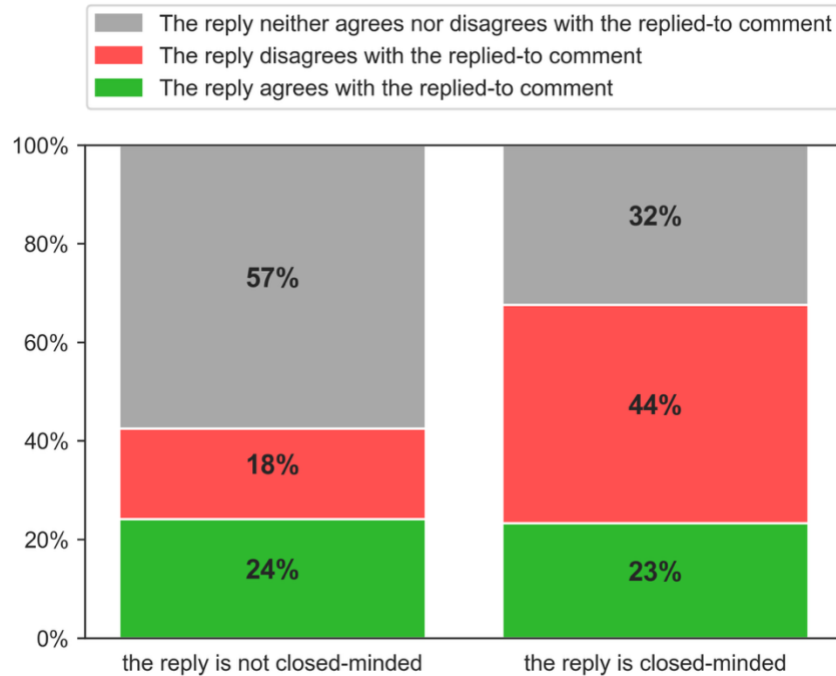discernible difference in the tendency to agree.

Figure 5. The distribution of the replies based on whether they were closed-minded and whether they displayed agreement, disagreement, or neither with the replied-to comment.

Discussion

This research examined closed-mindedness in online political spaces. To achieve this, we first developed and validated a text classifier of closed-mindedness. Then, we used it to assess the closed-mindedness of Reddit comments. Our results suggest that closed-mindedness is prevalent in online political spaces. Political communities are much more closed-minded than nonpolitical ones, and this difference cannot be attributed to toxicity. In these political communities, comments are especially likely to display closed-mindedness when they also express a partisan lean.

We also examined the social approval received by the comments expressing a partisan lean. We found that among those with a lean misaligned to the subreddit's lean, those that were also closed-minded were more likely to face social sanction. If closed-mindedness is indeed a means of highlighting deviant behavior, this finding aligns with previous literature suggesting that groups enforce their norms by punishing deviants (Jetten & Hornsey, 2014). Future work can investigate the consequences of such

social sanctioning. For instance, do targets of social sanctions adapt their behavior to comply with group norms? Is such compliance displayed only when among specific groups or does it generalize across social contexts (Mamakos et al., 2025)? Contrary to our expectations, we did not find that partisan comments aligning with a subreddit's lean were less likely to face social sanction when they displayed closed-mindedness (although social sanction was rare for all lean-aligned comments, closed-minded or not). A limitation of this finding is that we cannot be sure that this social disapproval has come from members of the subreddits' ingroup. While partisan political subreddits are, by definition, predominantly populated by ingroup members, they can also be joined by outgroup members who form a minority in these subreddits and downvote comments based on their preferences.

Our findings also suggest that closed-minded comments result in more closed-minded replies, thereby creating a perpetual cycle that fosters unconstructive discourse. Under these conditions, it is unlikely that open-mindedness will emerge to pave the way toward reducing partisan antipathy (Dolbier et al., 2024), a contribution that is much needed in today's polarized society (Iyengar et al., 2012; Finkel et al., 2020). Our results suggest that steering discussions toward receptivity to different perspectives can be challenging when closed-mindedness has prevailed from the outset. Future research can investigate the extent to which this closed-mindedness is due to individual differences (Costello et al., 2023) rather than situational factors.

The ability to consider differing views is a fundamental component of democratic functioning. The best chance for discovering effective policies arises from the free exchange and open-minded evaluation of a wide range of possible ideas. Social media platforms are especially accessible for citizens to contribute to the marketplace of political ideas. However, the present research suggests that closed-mindedness is widespread in online political spaces, where it often prevails from the outset, leaving little

room for constructive deliberation. These findings highlight the need to create new online spaces that promote deliberative engagement and willingness to entertain alternative viewpoints.

Methods

The dataset used to validate the classifier of closed-mindedness consisted of 3,600 comments from left-wing, right-wing, and nonpolitical subreddits (1,200 randomly sampled comments from each of these three subreddit types). To examine the generality of our classifier regarding assessing comments from subreddits not included in the fine-tuning phase, this validation dataset comprised comments both from subreddits that contributed comments to the fine-tuning dataset and from subreddits that did not contribute to the fine-tuning dataset (see SM Table S1). The human annotators demonstrated high consistency in this validation dataset (ICC = .82). The validation metrics were satisfactory and similar for comments from left-wing, right-wing, and nonpolitical subreddits (see SM Table S9), as well as for comments from subreddits that did vs. did not contribute to the fine-tuning dataset (see SM Table S10).

The 1 million comments sampled to test Hypotheses 2-5 satisfied both criteria of (i) being drawn from political subreddits with a known lean and (ii) being direct responses to posts and not to other comments (i.e., "level-1" comments). Criterion (i) was needed because a test of Hypothesis 3 requires knowledge of the subreddits' lean. Using the partisanship measure of subreddits derived by Waller and Anderson (2021), we considered only political subreddits with partisanship at least either 1 SD below the mean (left-wing subreddits) or 1 SD above the mean (right-wing subreddits). These partisan political subreddits comprised 140 of the 223 total political subreddits (86 left-wing and 54 right-wing subreddits). Criterion (ii) was needed because, when testing Hypothesis 4, the position of a comment in a multi-threaded discussion can confound the probability of receiving a reply. To avoid this incidental confounder, and since a multi-threaded discussion cannot start without a level-1 comment, we considered only those comments. This implies that the replies (which were themselves comments) in the

tests of Hypotheses 4 and 5 were level-2 comments. The results in Figure 3 involved only the comments that displayed a partisan lean (22% of the 1 million comments). All comments in our analyses were sampled from 2011 to 2022 (see SM Section 2).

References

Acosta, J., & Kemmelmeier, M. (2022). The changing association between political ideology and closed-mindedness: Left and right have become more alike. *Journal of Social and Political Psychology*, *10*(2), 657-675.

Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review*, *110*(3), 629-676.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., Lee, J, Mann, M., Merhout F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, *115*(37), 9216-9221.

Barabas, J. (2004). How deliberation affects policy opinions. *American Political Science Review*, *98*(4), 687-701.

Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., & Blackburn, J. (2020). The pushshift reddit dataset. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 14, pp. 830-839).

Bohman, J. (2006). Deliberative democracy and the epistemic benefits of diversity. *Episteme*, *3*(3), 175-191.

Bolsen, T., Druckman, J. N., & Cook, F. L. (2014). The influence of partisan motivated reasoning on public opinion. *Political Behavior*, *36*, 235-262.

Costello, T. H., Bowes, S. M., Baldwin, M. W., Malka, A., & Tasimi, A. (2023). Revisiting the rigidity-of-the-right hypothesis: A meta-analytic review. *Journal of Personality and Social Psychology*, 124(5), 1025.

Dolbier, S. Y., Dieffenbach, M. C., & Lieberman, M. D. (2024). Open-mindedness: An integrative review of interventions. *Psychological Review*.

Falkenberg, M., Zollo, F., Quattrociocchi, W., Pfeffer, J., & Baronchelli, A. (2024). Patterns of partisan toxicity and engagement reveal the common structure of online political communication across countries. *Nature Communications*, 15(1), 9560.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.

Finkel, E. J., Bail, C. A., Cikara, M., Ditto, P. H., Iyengar, S., Klar, S., ... & Druckman, J. N. (2020). Political sectarianism in America. *Science*, *370*(6516), 533-536.

Gutmann, A., & Thompson, D. F. (1996). *Democracy and disagreement*. Harvard University Press.

Habermas, J. (1985). *The theory of communicative action: Volume 1: Reason and the rationalization of society*. Beacon Press.

Heltzel, G., & Laurin, K. (2024). Why Twitter sometimes rewards what most people disapprove of: The case of cross-party political relations. *Psychological Science*, *35*(9), 976-994.

Hofmann, V., Schütze, H., & Pierrehumbert, J. B. (2022). The reddit politosphere: a large-scale text and network resource of online political discourse. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 16, pp. 1259-1267).

Huddy, L., Mason, L., & Aarøe, L. (2015). Expressive partisanship: Campaign involvement, political emotion, and partisan identity. *American Political Science Review*, 109(1), 1-17.

Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideology: A social identity perspective on polarization. *Public Opinion Quarterly*, *76*(3), 405-431.

Jetten, J., & Hornsey, M. J. (2014). Deviance and dissent in groups. *Annual Review of Psychology*, *65*(1), 461-485.

Jiang, A. Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D. S., de las Casas, D., Bressand, F., Lengyel, G., Lample, G., Saulnier, L., Lavaud, L. R., Lachaux, M.-A., Stock, P., Scao, T. L., Lavril, T., Wang, T., Lacroix, T., & Sayed, W. E. (2023). Mistral 7B. https://arxiv.org/abs/2310.06825

Jost, J. T., Glaser, J., Sulloway, F., & Kruglanski, A. W. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin*, 129, 339 –375.

Kalmoe, N. P., & Mason, L. (2022). *Radical American partisanship: Mapping violent hostility, its causes, and the consequences for democracy*. University of Chicago Press.

Kruglanski, A. W., & Webster, D. M. (1991). Group members' reactions to opinion deviates and conformists at varying degrees of proximity to decision deadline and of environmental noise. *Journal of Personality and Social Psychology*, *61*(2), 212.

Kruglanski, A. W., Pierro, A., Mannetti, L., & De Grada, E. (2006). Groups as epistemic providers: need for closure and the unfolding of group-centrism. *Psychological Review*, *113*(1), 84.

Lorenz-Spreen, P., Oswald, L., Lewandowsky, S., & Hertwig, R. (2023). A systematic review of worldwide causal and correlational evidence on digital media and democracy. *Nature Human Behaviour*, *7*(1), 74-101.

Mamakos, M., & Finkel, E. J. (2023). The social media discourse of engaged partisans is toxic even when politics are irrelevant. *PNAS Nexus*, *2*(10), pgad325.

Mamakos, M., Charlesworth T. E. S., & Finkel, E. J. (2025). Moralizing partisanship when surrounded by copartisans versus in mixed company. *PNAS Nexus*, *4*(4), pgaf105.

Mercier, H., & Landemore, H. (2012). Reasoning is for arguing: Understanding the successes and failures of deliberation. *Political Psychology*, *33*(2), 243-258.

Price, V., Cappella, J. N., & Nir, L. (2002). Does disagreement contribute to more deliberative opinion? *Political Communication*, *19*(1), 95-112.

Rathje, S., Van Bavel, J. J., & Van Der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, *118*(26), e2024292118.

Rawls, J. (1951). Outline of a decision procedure for ethics. *The Philosophical Review*, *60*(2), 177-197.

Schachter, S. (1951). Deviation, rejection, and communication. *The Journal of Abnormal and Social Psychology, 46*(2), 190–207.

Simchon, A., Brady, W. J., & Van Bavel, J. J. (2022). Troll and divide: the language of online polarization. *PNAS Nexus*, 1(1), pgac019.

Sun, Q., Wojcieszak, M., & Davidson, S. (2021). Over-time trends in incivility on social media: evidence from political, non-political, and mixed sub-reddits over eleven years. *Frontiers in Political Science*, *3*, 741605.

Sunstein, C. R. (2018). *Republic: Divided democracy in the age of social media*. Princeton University Press.

Waller, I., & Anderson, A. (2021). Quantifying social organization and political polarization in online platforms. *Nature*, *600*(7888), 264-268.

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology*, *67*(6), 1049.

Wulczyn, E., Thain, N., & Dixon, L. (2017, April). Ex machina: Personal attacks seen at scale. In *Proceedings of the 26th International Conference on World Wide Web* (pp. 1391-1399).

Yarchi, M., Baden, C., & Kligler-Vilenchik, N. (2021). Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. *Political Communication*, *38*(1-2), 98-139.

Young, I. M. (1996). *Communication and the Other: Beyond Deliberative Democracy. Democracy and Difference: Contesting the Boundaries of the Political.* Princeton University Press.