Whether groups value agreement or dissent depends on the strength of consensus

James C. Mellody

Center for Information Technology Policy

Princeton University

Abstract:

I investigate the conditions under which groups value agreement versus dissent in collective decision making processes. I argue that which kind of contribution a group values more will depend on the strength of the consensus. As a consensus evolves from weak to moderate to strong, I predict that groups will prefer agreement, dissent, and then agreement again. These predictions are in line with a multi-phase decision-making process in which groups pursue sequential goals reflected in the evolving consensus: establishing an initial consensus, exploring alternative perspectives, and settling on a final decision. I find support for these predictions using data from the Reddit community r/AmItheAsshole, in which people make normative judgments of social situations.

Keywords: collective intelligence, decision making, consensus, social influence

Through collective intelligence, the aggregation of inputs from many people can produce solutions and decisions that outperform individual efforts [1–4]. Whereas collective intelligence once emerged only locally, through co-located groups and teams [5], the continued proliferation of the internet has extended the reach of collective intelligence to a wide variety of tasks—including evaluation [6], prediction [7], and innovation [8]—that now benefit from global networks of individuals offering thoughts, ideas, opinions, and judgments.

A large body of work has explored the role of social influence in collective intelligence—mostly investigating how, and under what conditions, social influence either inhibits the diversity of individual judgments and contributes to premature or suboptimal convergence [9, 10] or alternatively offers a means of learning that facilitates convergence on optimal judgments from the crowd [11–15].

In addition to the process of making and updating judgments, another key process in collective intelligence is how the collective weights, or values, different kinds of judgments [16]. Much of the research on this process compares different weighting strategies to see which might be the most optimal [17–21]. Given the importance of social influence in the processes of making and updating judgments, it is likely that social influence also plays an important role in the way that collectives value different kinds of judgments in practice.

In this article, I investigate the role of social influence in the process of valuing contributions to the collective. To do so, I examine how new judgments relate to existing judgments collected by the group and whether this relationship matters for the valuation of these new judgments. More specifically, I ask the following research question: how does the value that a group places on an individual judgment depend on whether the judgment agrees with, and thus bolsters, or disagrees with, and thus questions, the existing group consensus?

There is evidence that groups value both agreement and dissent—as both can facilitate decision-making in different ways [22]. Groups value agreement because it fuels convergence on a decision, which is the ultimate goal of many collective intelligence tasks [23–27]. Additionally, groups may value agreement as a means not only of reaching a decision, but also of fostering cohesion, which can veer into groupthink when the pursuit of cohesion incentivizes conformity and crowds out alternative perspectives [28–31]. On the other hand, groups may value dissent—especially when they approach decision-making as problem solving—because it represents diversity of thought, which can bolster decision quality [32,33]. Whether a particular group values agreement or dissent may depend on the nature of the task (e.g., where the task falls on the speed-accuracy tradeoff [34]) as well as group norms around consensus and criticality [35].

While particular groups may have a tendency to value agreement or dissent more, it is also possible that, within the course of a single decision-making process, conditions evolve such that groups alternate between valuing agreement more and valuing dissent more. Research on group decision-making finds that decision-making occurs in phases, across which the motivation of the group can shift. Much of this work suggests a multi-phase process, in which the group's motivation shifts from general concerns around orienting to a problem, to exchanging ideas, and finally to settling on a solution [36–38]. Further research suggests a link between a group's motivation and how the group values different kinds of contributions, finding that an approaching deadline can induce a "need for closure," in turn leading groups to reject dissenting opinions in favor of opinion uniformity [39–41]. Building from these ideas, I explore how groups value dissent versus agreement as they move through a decision-making process. I argue that the strength of the group consensus offers a social reference point enabling individuals in the collective to track the phase of the decision-making process, such that whether a group places greater value on agreement or dissent will depend on the strength of the consensus.

More specifically—and drawing from research on the benefits of agreement and dissent as well as the phases of decision-making—I predict a curvilinear relationship between the strength of consensus and the relative value placed on agreement versus dissent. I predict that when the consensus is weak, groups will value agreement more, in line with a goal of strengthening an initial consensus [23, 24]; when the consensus is moderate, groups will value dissent more, in line with a desire to explore diverse perspectives and pressure test the emerging consensus [32, 33, 36]; and when a consensus is strong, groups will once again value agreement more, in line with a need for closure [36, 39–41].

Setting and Data

I test these predictions in the domain of online evaluations—an increasingly prominent context of collective intelligence [6]. I leverage data from the Reddit community r/AmItheAsshole—a forum that crowd-sources normative evaluations of social situations. In each thread on the forum, the original poster (OP) will post a contentious social situation from their life, and then other people make comments judging who was the "asshole" in the situation. Commenters make one of the following judgments: "you're the asshole (YTA)" denoting the poster as the asshole, "not the asshole (NTA)" denoting the other party as the asshole, "everyone sucks here (ESH)" denoting all parties as assholes, "no assholes here (NAH)" denoting no parties as assholes, and finally "not enough info (INFO)" to convey that more information is needed to make a judgment. Fig. 1 displays descriptive information about the community—including number of posts and users over time, and the distribution of judgments across posts over time.

I focused the analysis on comments made on posts that ultimately received a judgment of either YTA or NTA, because these two judgments were unambiguous. Posts judged as YTA and NTA made up the bulk of posts, as seen in Fig. 1. I analyzed only comments that contributed a judgment of YTA or NTA, because these comments had an unambiguous relationship of dissent or agreement relative to the post judgments. I only analyzed top-level comments with judgments, as commenters are supposed to make judgments using top-level comments, according to the subreddit rules.

The rules of the subreddit assign a final post judgment based on the judgment of the comment that received the highest score 18 hours after the post was made. For the purposes of the analyses here, I measured the final judgment in two ways, and then limited the sample to posts for which these two measures agreed (97% of all posts). I did this to ensure that posts in the analyses had an unambiguous judgment, and to capture a more complete picture of the judgment based on the contributions of all in the community, rather than just a single comment. First, I captured the judgment of the single comment that received the highest score. Second, I added up the scores received by comments of each judgment type ("NTA", "YTA", "NAH", "ESH", or "INFO") to see which judgment received the highest score.

I then tracked the evolving strength of the consensus judgment of the post—a measure I call *Consensus Strength*. I measured this as a function of both the score-weighted proportion of comments at time t that expressed the majority judgment (I tested the robustness with an unweighted measure—i.e., proportion of comment counts—and found similar results; see *SI Appendix*), and the logged number of comments with judgments at time t to capture whether the proportion variable reflects many or few comments. In addition to data limitations which exclude the possibility of measuring the evolving score of a single comment, I measured the consensus strength across all comments with judgments in order to capture the evolving consensus of the full group, rather than focusing on the single comment that received the most attention. Doing so captures a more complete picture of the evolving consensus of the group.

In the analysis, I modeled the logged number of upvotes that each comment received—a measure of how the group valued the comment—as a function of the strength of the consensus at the time the comment was made as well as a binary variable indicating whether the focal comment dissented or agreed with the current consensus, along with a number of other control variables. See *Materials and Methods* for more information on the data, measures, and analyses.

Results

Main Result

The results of the model offer evidence in support of the predicted patterns. The main result is a curvilinear relationship between the strength of the consensus and the relative value placed on dissent versus agreement. Fig. 2 displays the average marginal effect of dissenting with the current consensus on the predicted score (log) of a comment, across different consensus strengths. When the consensus strength was very weak (i.e., two standard deviations below the mean), dissenting comments were slightly discounted, receiving an average score that was 9.09% lower ($p < 0.01$) than the average score of agreeing comments. As the consensus got a bit stronger (i.e., one SD below the mean) the relationship reversed: dissenting comments received an average score that was 21.40% higher ($p < 0.001$) than the average score of agreeing comments. At the mean, dissent was favored even more: dissenting comments received an average score that was 30.37% higher ($p < 0.001$) than the average score of agreeing comments As the strength of the consensus further increased (i.e., one SD above the mean), dissent was still slightly favored: dissenting comments received an average score that was 12.58% higher ($p < 0.001$) than the average score of agreeing comments. Finally, once the consensus was very strong (i.e., two SD above the mean), agreement was once again favored: dissenting comments received an average score that was 21.82% lower ($p < 0.001$) than the average score of agreeing comments.

These results offer evidence in support of both the prediction that the relative value placed on dissent versus agreement would vary with the strength of the consensus, and the more specific predictions for how this relationship would vary as the consensus evolved. This main pattern was robust across a variety of criteria for filtering the sample (see *SI Appendix*).

Regression Discontinuity in Time

To investigate the causal relationship between the strength of consensus and how agreement and dissent are valued, I conducted a regression discontinuity in time (RDiT) analysis [42]. This is a regression discontinuity design in which time operates as the running variable. The goal of this analysis was to see if the kind of contributions the group values would shift after a shock to the strength of the consensus. I used instances when the original poster (OP) of a post made a comment on their own post as a shock to the strength of the consensus. The logic is that the OP commenting on their own post provides new information, painting a more complete picture of the situation and thus lending strength to the consensus. See the *SI Appendix* for graphical evidence of the jump in the strength of consensus when OP posts a comment. The shift from pre- to post-OP making a comment is not a transition into a fully established consensus, but rather a transition from a first phase of building an initial consensus to a second phase of beginning to explore alternative perspectives. The expectation is that there should be a change in the kind of contribution valued by the group: away from valuing agreement and toward valuing dissent.

I used the time difference between the focal comment and the first comment made by the original poster (OP) as the running variable, and I limited the analysis to comments made 150 minutes before and after the OP's first comment, as RDiT estimates a local treatment effect around the event [42]. Fig. 3 shows the discontinuity around the first OP comment: the y-axis displays the mean difference in log score between dissenting and agreeing comments. Before OP made a comment, agreement was valued more, whereas after, there was a shift toward valuing dissent, in line with a move toward exploration. Predictive margins from the RDiT regression offer statistical support in line with these graphical results. Before OP made a comment, dissenting comments received an average log score of 0.74, which was 51% lower ($p < 0.001$) than the average log score of agreeing comments (1.51). After the OP made a comment, dissenting comments received an average log score of 1.48, which was 3.10% higher ($p > 0.001$) than the average log score of agreeing comments (1.44). These results offer evidence that a shock to the strength of the consensus can shift the kind of contribution valued by the group, and more specifically, that an increase in the strength of the consensus can shift the group away from valuing agreement and toward valuing dissent.

Exploring Heterogeneity across Topics

I explored potential heterogeneity in the main pattern by breaking down the analysis by post topic. I ran these models with data from the full set of years (i.e., 2013 - 2022), in order to increase the sample size for each topic. The goal here was to test the generalizability of the main result across different substantive topic areas. To identify the topics of posts, I ran a machine learning model, Top2Vec [43], which makes use of joint document and word embeddings to identify topic vectors. See *Materials and Methods* for more information. The 10 primary topics identified cover a wide range of social situations: health, neighbors, money, events, chores, dating, bigotry, social media, school, and work. Fig. 4 A displays a two-dimensional UMAP representation of the post embeddings, colored by primary topic.

Fig. 4 B displays the average marginal effects of dissenting, from the models with the data sub-setted to each of the 10 identified topics. The graphs show a relatively stable pattern across these 10 topics. Variation comes mainly at the lower end of the consensus strength spectrum: for some topics (i.e., neighbors, dating, social media, school, and work), agreeing comments were valued more—in line with the main pattern, whereas for other topics (i.e., health, money, events, chores, and bigotry), there was not a statistically significant difference between the scores received by dissenting and agreeing comments. This suggests that the initial phase of the decision-making process may differ depending on the topic—sometimes the focus of this phase may be valuing agreement to strengthen the consensus, and other times the focus may be on gathering all kinds of contributions.

Discussion

The present study investigated how groups value agreement and dissent throughout the process of making a decision. Leveraging data from the Reddit community r/AmItheAsshole, I found evidence supporting the prediction that the relative value placed on agreement versus dissent would vary with the strength of the consensus. Specifically, I found evidence of a curvilinear relationship between the strength of the consensus and the relative value placed on dissent versus agreement. Groups valued agreement over dissent when a consensus was very weak or very strong, but placed greater value on dissent when the consensus was moderate.

This study offers two main contributions. First, this study contributes to our understanding of social influence in collective intelligence. Whereas past work has focused on the role of social influence in how individuals make and update their judgments [9–15], here I focus on the role that social influence plays in how groups value different kinds of contributions. Specifically, I find that the group's existing consensus operates as a reference point, such that individuals in the group value new contributions based on how these contributions relate to the existing consensus. Together with prior research on social influence in collective intelligence, this helps to paint a more complete picture of exploration and convergence in collective intelligence. Whether, and how, groups explore diverse perspectives and/or converge on particular judgments depends both on how judgments are made and updated, and also on how the group values different kinds of judgments. These processes are interrelated in collective intelligence, as individuals are motivated to take actions that they see receiving recognition [44].

Second, this study contributes to research on the phases of group decision-making [36–38]. The results offer evidence in support of a three-phase decision-making process, in which groups move from orienting to the problem and building an initial opinion, to evaluating and considering different opinions, and finally to settling on a decision [36]. The present study shows that these different phases have implications for how groups value dissent versus agreement. This matters because decision quality relies on a group's ability to consider diverse perspectives and avoid converging on suboptimal solutions or ideas.

It is important to consider the limitations of this study. This study leveraged data from one online context. It is likely that the general pattern identified in this study—that whether groups value agreement or dissent more varies with the strength of the consensus—will generalize to other settings. This is because all decision-making processes likely share the characteristic that the salient goal will shift over time as the strength of consensus changes. The specific evolution (i.e., when dissent or agreement is valued more) may be more context dependent, relying on factors such as the group's normative orientation toward consensus and criticality [35], the group's specific process for making decisions [45], and the nature of the task itself—for example, the difficulty and whether speed or accuracy is prioritized [46]. While the specific patterns of valuing agreement and dissent identified in this study may generalize to other online evaluation contexts in which these conditions are similar, it is an open question how far-reaching these specific patterns are to other contexts. Another area for future research is examining whether different patterns of alternating between valuing agreement and dissent are more or less optimal for achieving accurate or otherwise effective decisions.

Materials and Methods

Data and Sample

Data was gathered through data dumps provided by Pushshift.io [47]—available from the start of the subreddit in 2013 through the end of 2022. For the main analyses, I used matched comments on posts made in 2022—consisting of 6,799,071 comments—because some rules of the subreddit have changed over time and I wanted to capture a period of subreddit stability and maturity, and this is the last full year of Reddit data available. I conducted a robustness check leveraging comments on posts from all years (i.e., 2013-2022), finding consistent results (see *SI Appendix* for this and other robustness checks). Because the goal of the study was to look at the valuation of contributions throughout the course of a decision-making process, I limited the sample to posts which received a minimum of 15 comments with judgments so that each post had a critical mass of judgments. Findings were robust to a different choice of minimum comment threshold. In the main model, I included comments from posts whose main text was deleted by the author or removed by the moderator team [1], and found the results to be consistent if I excluded these comments. In the main analyses, I also removed comments made when the existing consensus was fully balanced (i.e., the proportion

[1] It was a relatively common occurrence for posters to remove their post eventually, but the Reddit AutoModerator bot would repost the original post's content so that people could keep commenting.

variable was equal to 0.5 exactly). This is because the relationship between these comments and the existing consensus was ambiguous. I ran robustness checks including these comments (demarcating them as being in agreement with the existing consensus because a comment made when the consensus is fully balanced effectively establishes a new consensus), and found largely consistent results. [2]

I also limited the analyses to only include comments that received scores equal to or greater than 1. A comment score can take on any integer value. Comments start out with a score of 1, and then can move to a score of 0 and into negative scores if the comment receives more downvotes than upvotes. I limited the sample in this way because the rules of r/AmItheAsshole state that downvoting is reserved not for expressing disapproval of a comment's judgment, but instead for signaling that a comment is off topic. This means that comments receiving more downvotes than upvotes (i.e., those with a score of 0 or lower) were likely to be irrelevant or off-topic comments, which I wanted to exclude from the analyses.

Measures

In measuring the strength of the consensus, I created a composite measure taking two elements into account. First, in line with measurements of consensus in discrete choice scenarios [48], I measured the strength of the current majority consensus. To measure this, I captured a running tally of the number of comments with the judgment "YTA" and a second running tally of the number of comments with the judgment "NTA." I then captured the proportion of comments belonging to the majority judgment—meaning a score from (0.5, 1]. This was an unweighted measure of consensus. I also constructed a weighted measure of consensus, as users can upvote a comment if they agree with it, and so the strength of consensus can reflect not just the count of different judgments, but also the number of upvotes that the comments expressing these judgments have received. To construct this weighted measure of consensus, I multiplied each comment by its score, using the sum of these weights for all "YTA" and "NTA" comments to represent the total weighted opinion for each.[3] I then calculated the proportion of weighted comments contributing to the majority—again a score from (0.5, 1]. I used the weighted measure in the main analysis, and conducted a robustness check with the unweighted measure, finding consistent results (see *SI Appendix*). Second, I then multiplied the weighted or unweighted proportion variable by the logged number of other comments made at the time of the focal

[2] Although note that at low consensus strength, there was no difference in valuing dissent vs. agreement, likely because these newly included comments occurred when the consensus was weak—because a balanced consensus often occurs early on with only a few comments, meaning the consensus strength is weak. The unclear nature of how these comments relate to the existing consensus likely adds noise at this spectrum of consensus strength, thus why I exclude these comments from the main analysis.

[3] A data-based limitation of this approach is that the weight is based on the score at the time the data was collected, rather than exactly around the time the comment was made. While the weight thus does not capture the exact score of each comment at the time the comment was made, it does capture how the comment was ultimately received. Additionally, there is always a question of timing when capturing a live measure of weighted influence of a comment. In other words, there is always a question of when the comment's score should be captured, as a comment will likely not accumulate upvotes and downvotes until some amount of time has elapsed. Finally, the weighted and unweighted measures yield similar results, suggesting that the finding is robust to different ways of capturing the consensus strength.

comment. The goal here was to contextualize the proportion variable with information on the volume of comments contributing to that proportion. See the *SI Appendix* for more information on these measures.

I also measured whether a comment dissented or agreed with the current consensus: *Consensus Dissent*. This is a binary variable—0 if the comment agreed with the consensus, and 1 if it dissented. To capture whether a comment dissented or agreed with the consensus, I measured the judgment of the comment and whether that judgment agreed with the judgment that was currently in the majority under the current consensus. So, for example, if "YTA" currently had a 65% majority, and the focal comment expressed "YTA," then it was assigned a 0 for this variable, denoting agreement. If the focal comment expressed "NTA," then it was assigned a 1, denoting dissent.

I included several control variables that likely impacted the score received by a comment. I included the length of the comment based on the number of characters: *Comment Length (ln)*, the time in minutes since the original post: *Minutes Since Post (ln)*, the number of other comments in the thread at the time of the focal comment (as a measure of competition for attention)—I added 1 to this variable before taking the log: *Comment Competition (ln)*, the score of the comment author's other comments made previously in the subreddit (as a measure of author skill at expressing a judgment)—I used STATA's lnskew0 command to take the log of this variable which ranged from negative to positive values: *Author Score (ln)*, and finally, indicators to capture time-invariant characteristics of the post in which the comment was made, the month of the year in which the comment was made, the hour of the day in which the comment was made, and the day of the week on which the comment was made. See the *SI Appendix* for descriptive characteristics and correlations for all variables.

Analyses

I ran linear regressions predicting the score (log) received by each comment. I estimated the following regression equation:

$$y_i = \beta_0 + \beta_1 D_i + \beta_2 CS_i + \beta_3 CS_i^2 + \beta_4 D_i CS_i + \beta_5 D_i CS_i^2 + \delta X_i + \alpha_i + \alpha_m + \alpha_d + \alpha_h + \epsilon_i$$

Where y_i is the logged score of the focal comment, D_i is an indicator for whether the comment dissents with the existing consensus, CS_i is the current strength of the consensus. β_5 is the main quantity of interest: the interaction effect between the dissent variable and the squared consensus strength term, to account for possible non-linearity. X_i is a vector of covariates and δ a vector of accompanying coefficients, α_i is a fixed effect for the post in which the focal comment appears, α_m is a month fixed effect, α_d is a day-of-week fixed effect, α_h is an hour-of-day fixed effect, and ϵ_i is the error term. To simplify the interpretation of the interaction effect of interest, I estimated the average marginal effect of dissenting with the consensus, at different consensus strengths, using the STATA margins command.

In order to move toward a causal interpretation of the findings, I took two additional steps. First, I used a coarsened exact matching procedure [49, 50] to generate weights for each observation, which I then included in the main regression. This has the dual benefit of improving covariate balance between comparison units (here comments that dissent versus agree with the consensus) and also of operating as a regression pre-processing step that reduces model dependence [51]. I coarsened four variables (i.e., the strength of the consensus, the time since the post in minutes, the score of the author's other comments in the subreddit, and the length of the comment) into 12 equally sized quantile bins, and a fifth variable (i.e., the number of other comments) into 25 equally sized quantile bins. This process generated weights for each observation, based on how many times each control observation (i.e., agreeing comment) was used as a match for a treated observation (i.e., dissenting comment). I included these weights in the main regression. Table SI5 shows the pre- and post-matching covariate balance, showing a significant improvement from matching.

Second, I conducted a regression discontinuity in time analysis [42]. I estimated the following equation:

$$y_i = \beta_0 + \beta_1 D_i + \beta_2 OP_i + \beta_3 D_i OP_i + \gamma f(TOP_i) + \delta X_i + \alpha_i + \alpha_y + \alpha_m + \alpha_d + \alpha_h + \epsilon_i$$

Where y_i is the logged score of the comment, D_i is an indicator for whether the comment dissents with the existing consensus, OP_i is an indicator for whether the original poster (OP) has posted a comment yet or not, and TOP_i is the running variable (i.e., minutes from first OP comment). Identification comes from the assumption that the potentially endogenous relationship between ϵ_i and the running variable TOP_i is captured by the flexible function f(.) [52]. To estimate this, I included in the model a three-way interaction between TOP_i, D_i, and OP_i, as well as the two-way interactions between TOP_i and each of these other two variables. X_i is a vector of covariates and δ a vector of accompanying coefficients, α_i is a fixed effect for the post in which the focal comment appears, α_y is a year fixed effect, α_m is a month fixed effect, α_d is a day-of-week fixed effect, α_h is an hour-of-day fixed effect, and ϵ_i is the error term. To simplify the interpretation of the interaction effect of interest, I estimated the marginal effect of dissenting with the consensus, before and after the OP's first comment. I used data from all years for this analysis because data in the small time threshold (150 minutes before/after OP commented) is limited in any given year.

For the heterogeneity in topic analyses, I identified post topics using the Top2Vec algorithm [43]. Top2Vec calculates topic vectors using the following steps: 1) generate joint documents and word embeddings—I used the Doc2Vec algorithm here, 2) create lower dimensional embeddings with UMAP, 3) identify dense areas—topics—using HDBSCAN, 4) then calculate the centroid of document vectors in each dense area—this is the topic vector for each dense area, 5) locate the n-closest word vectors which comprise the resulting topic words. I then ran a hierarchical topic reduction algorithm, to reduce the number of topics to ten, and assigned a topic to each post based on the best fit topic. I labeled the topics by inspecting representative topic words from each. See the *SI Appendix* for example representative topic words from each of the 10 topics. I then ran ten regressions, subsetting to each of the topics in turn.

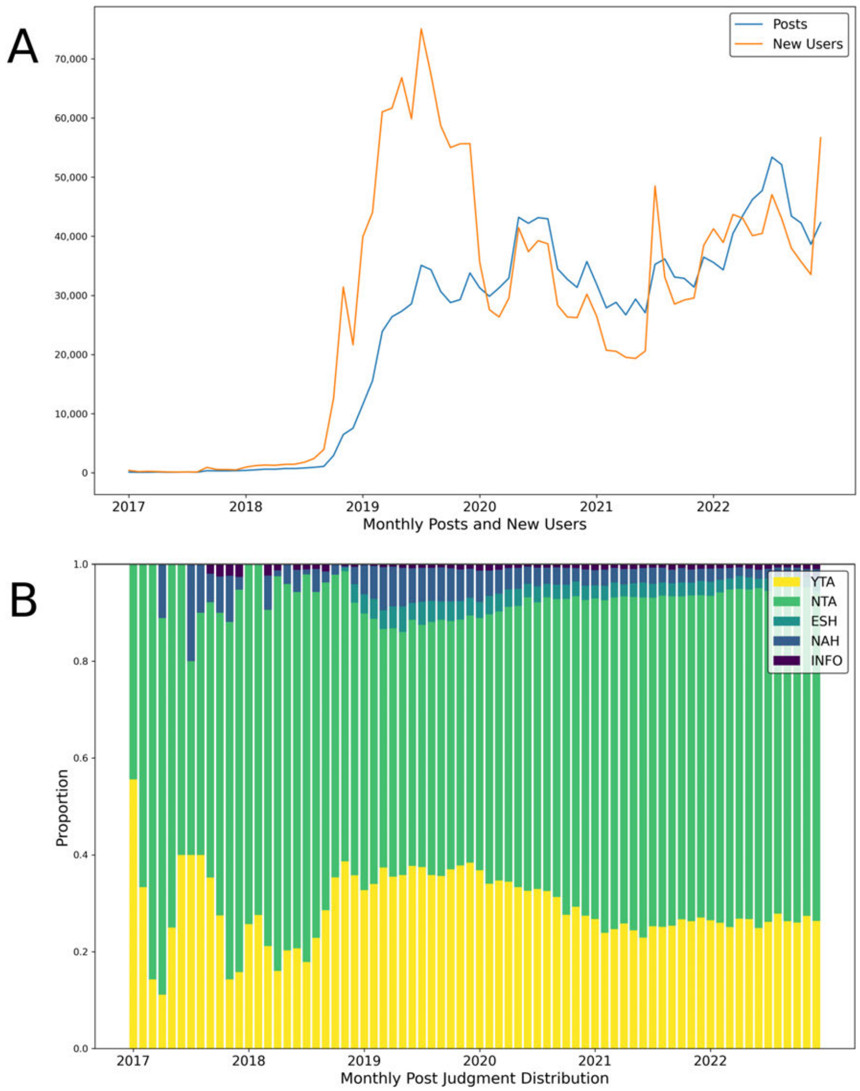

Figure 1: Information about r/AmItheAsshole. Data between 2017, before which subreddit activity was limited, and 2022. (A) Number of posts and users each month. The subreddit rapidly increased in popularity in late 2018. (B) Proportion of posts with judgments YTA, NTA, ESH, NAH, and INFO each month (limited to posts with > 14 comments).

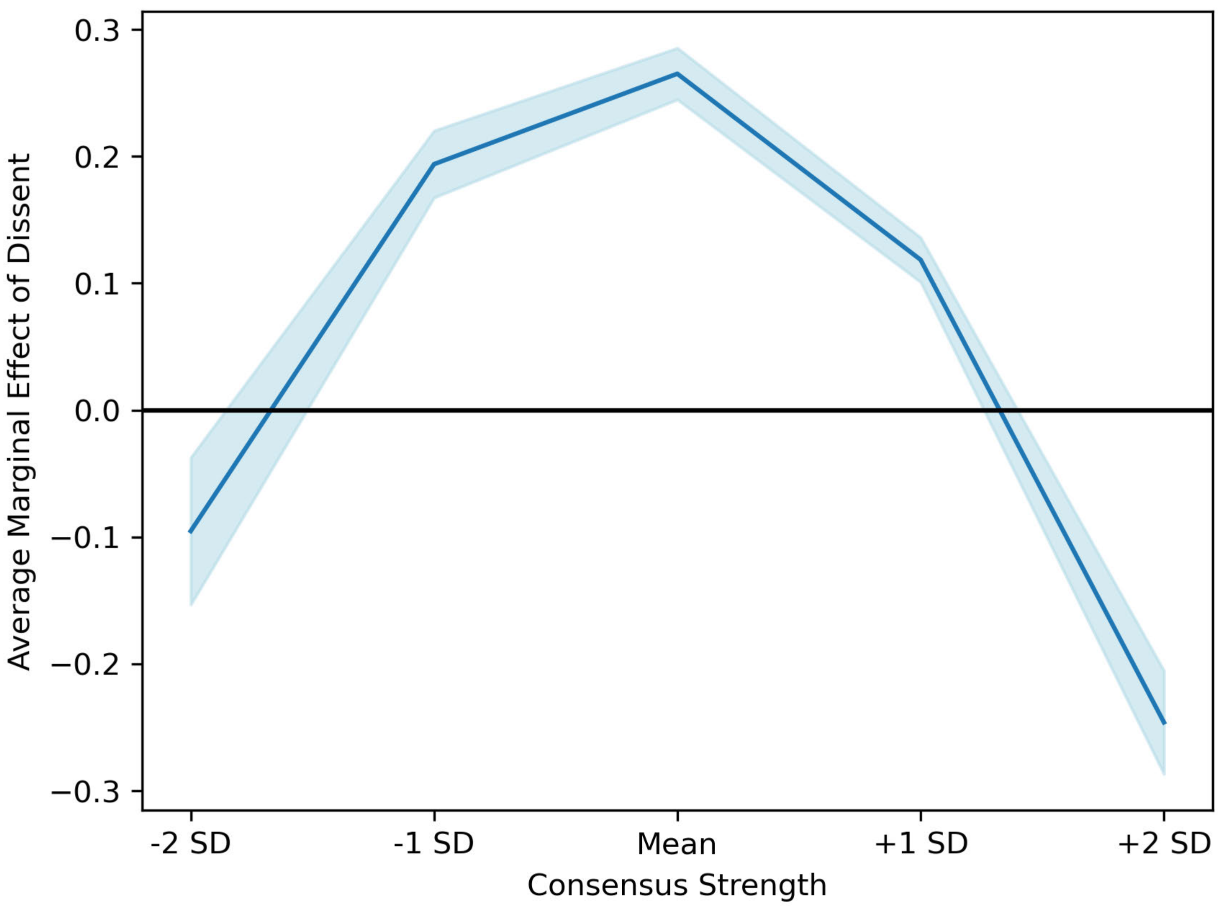

Figure 2: The main result, showing that, when the consensus strength is weak, agreement is valued more than dissent. When the consensus strength is moderate, dissent is valued more. When the consensus strength is high, agreement is once again valued over dissent.

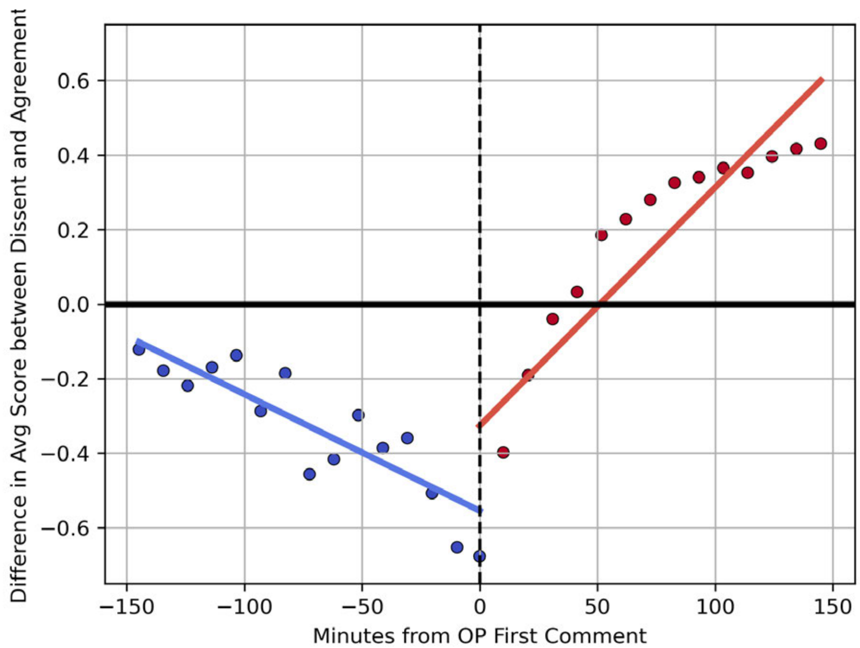

Figure 3: Local discontinuity in the outcome of interest (difference in average log score between comments that dissented versus agreed with the consensus). This graph shows the local effect based on a small bandwidth around the time of treatment: 150 minutes before and after the original poster (OP) first commented on the post. Outcomes are binned for each 10-minute interval. This shows that, before the OP's first comment, agreement was valued more, whereas after, there was a shift toward valuing dissent.

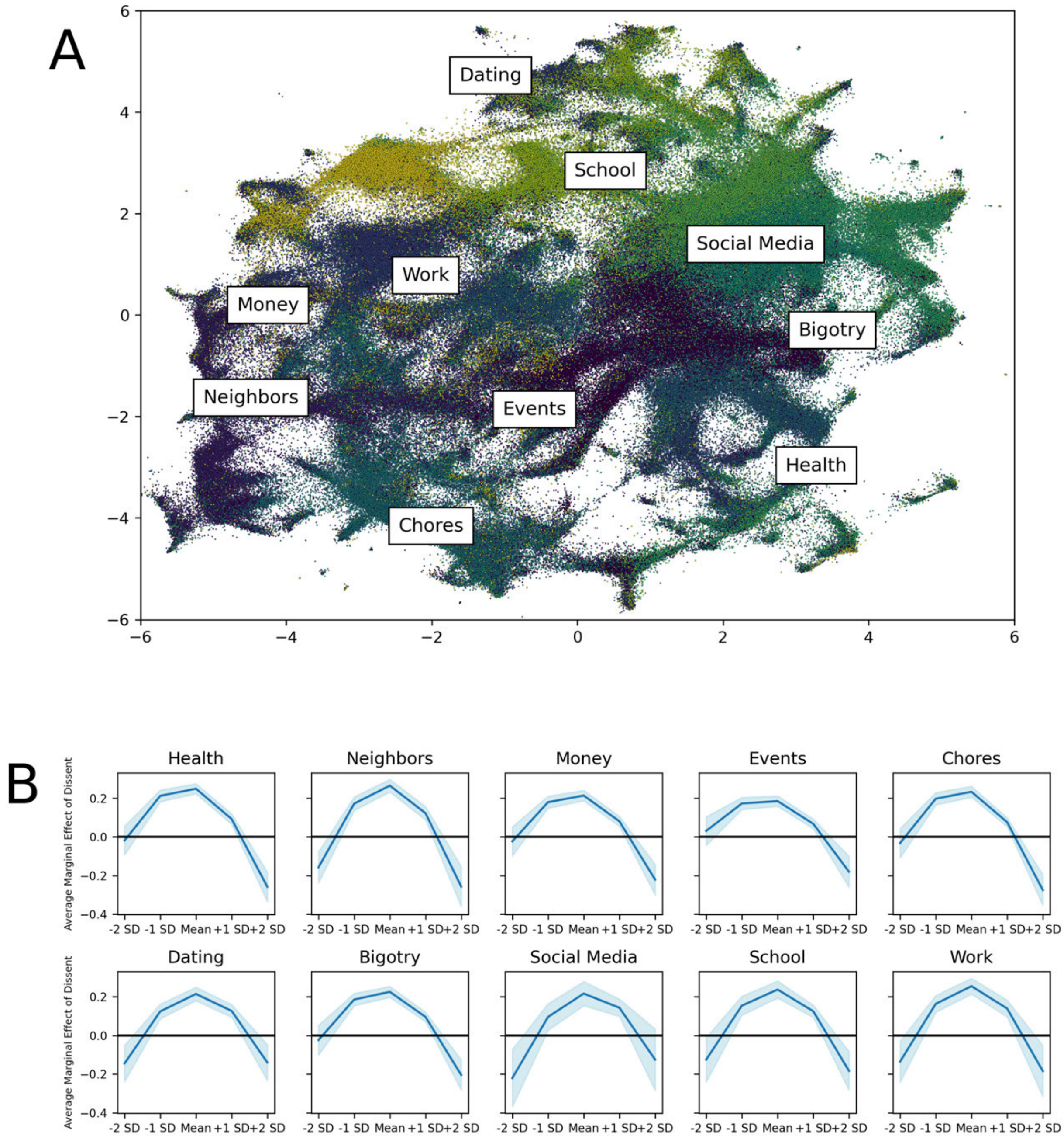

Figure 4: (A) 2-dimensional UMAP projection of post embeddings generated through Top2Vec model. (B) Average marginal effects of dissent across topics.

References

[1] F Galton, One vote, one value. *Nature* **75**, 414–414 (1907).

[2] I Lorge, D Fox, J Davitz, M Brenner, A survey of studies contrasting the quality of group performance and individual performance, 1920-1957. *Psychological bulletin* **55**, 337 (1958).

[3] J Surowiecki, The wisdom of crowds. *Surowiecki, J* (2005).

[4] RP Larrick, JB Soll, Intuitions about combining opinions: Misappreciation of the averaging principle. *Management science* **52**, 111–127 (2006).

[5] AW Woolley, CF Chabris, A Pentland, N Hashmi, TW Malone, Evidence for a collective intelligence factor in the performance of human groups. *science* **330**, 686–688 (2010).

[6] C Dellarocas, The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management science* **49**, 1407–1424 (2003).

[7] KJ Arrow, et al., The promise of prediction markets (2008).

[8] KJ Boudreau, N Lacetera, KR Lakhani, Incentives and problem uncertainty in innovation contests: An empirical analysis. *Management science* **57**, 843–863 (2011).

[9] J Lorenz, H Rauhut, F Schweitzer, D Helbing, How social influence can undermine the wisdom of crowd effect. *Proceedings of the national academy of sciences* **108**, 9020–9025 (2011).

[10] L Muchnik, S Aral, SJ Taylor, Social influence bias: A randomized experiment. *Science* **341**, 647–651 (2013).

[11] J Becker, D Brackbill, D Centola, Network dynamics of social influence in the wisdom of crowds. *Proceedings of the national academy of sciences* **114**, E5070–E5076 (2017).

[12] B Jayles, et al., How social information can improve estimation accuracy in human groups. *Proceedings of the National Academy of Sciences* **114**, 12620–12625 (2017).

[13] E Bernstein, J Shore, D Lazer, How intermittent breaks in interaction improve collective intelligence. *Proceedings of the National Academy of Sciences* **115**, 8734–8739 (2018).

[14] J Becker, E Porter, D Centola, The wisdom of partisan crowds. *Proceedings of the National Academy of Sciences* **116**, 10717–10722 (2019).

[15] A Almaatouq, et al., Adaptive social networks promote the wisdom of crowds. *Proceedings of the National Academy of Sciences* **117**, 11379–11386 (2020).

[16] F Galton, The ballot-box. *Nature* **75**, 509–510 (1907).

[17] B Bahrami, et al., Optimally interacting minds. *Science* **329**, 1081–1085 (2010).

[18] CP Davis-Stober, DV Budescu, J Dana, SB Broomell, When is a crowd wise? *Decision* **1**, 79 (2014).

[19] DV Budescu, E Chen, Identifying expertise to extract the wisdom of crowds. *Management science* **61**, 267–280 (2015).

[20] A Mahmoodi, et al., Equality bias impairs collective decision-making across cultures. *Proceedings of the National Academy of Sciences* **112**, 3835–3840 (2015).

[21] AB Palley, VA Satopää, Boosting the wisdom of crowds within a single judgment problem: Weighted averaging based on peer predictions. *Management Science* **69**, 5128–5146 (2023).

[22] J Jetten, MJ Hornsey, Deviance and dissent in groups. *Annual review of psychology* **65**, 461–485 (2014).

[23] L Festinger, Informal social communication. *Psychological review* **57**, 271 (1950).

[24] S Schachter, Deviation, rejection, and communication. *The Journal of Abnormal and Social Psychology* **46**, 190 (1951).

[25] S Moscovici, G Mugny, E Van Avermaet, *Perspectives on minority influence.* (Cambridge University Press) Vol. 9, (1985).

[26] CJ Nemeth, Differential contributions of majority and minority influence. *Psychological review* **93**, 23 (1986).

[27] G Whyte, Groupthink reconsidered. *Academy of Management Review* **14**, 40–56 (1989).

[28] SE Asch, Opinions and social pressure. *Scientific American* **193**, 31–35 (1955).

[29] A Gero, Conflict avoidance in consensual decision processes. *Small Group Behavior* **16**, 487–499 (1985).

[30] JM Marques, D Paez, The 'black sheep effect': Social categorization, rejection of ingroup deviates, and perception of group variability. *European review of social psychology* **5**, 37–68 (1994).

[31] IL Janis, Groupthink. *IEEE Engineering Management Review* **36**, 36 (2008).

[32] G Stasser, D Stewart, Discovery of hidden profiles by decision-making groups: Solving a problem versus making a judgment. *Journal of personality and social psychology* **63**, 426 (1992).

[33] S Schulz-Hardt, FC Brodbeck, A Mojzisch, R Kerschreiter, D Frey, Group decision making in hidden profile situations: dissent as a facilitator for decision quality. *Journal of personality and social psychology* **91**, 1080 (2006).

[34] WA Wickelgren, Speed-accuracy tradeoff and information processing dynamics. *Acta psychologica* **41**, 67–85 (1977).

[35] T Postmes, R Spears, S Cihangir, Quality of decision making and group norms. *Journal of personality and social psychology* **80**, 918 (2001).

[36] RF Bales, FL Strodtbeck, Phases in group problem-solving. *The Journal of Abnormal and Social Psychology* **46**, 485 (1951).

[37] BW Tuckman, Developmental sequence in small groups. *Psychological bulletin* **63**, 384 (1965).

[38] BA Fisher, Decision emergence: Phases in group decision-making. *Communications Monographs* **37**, 53–66 (1970).

[39] AW Kruglanski, DM Webster, 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**, 212 (1991).

[40] AW Kruglanski, DM Webster, Motivated closing of the mind:" seizing" and" freezing.". *Psychological review* **103**, 263 (1996).

[41] AW Kruglanski, A Pierro, L Mannetti, E De Grada, Groups as epistemic providers: need for closure and the unfolding of group-centrism. *Psychological review* **113**, 84 (2006).

[42] C Hausman, DS Rapson, Regression discontinuity in time: Considerations for empirical applications. *Annual Review of Resource Economics* **10**, 533–552 (2018).

[43] D Angelov, Top2vec: Distributed representations of topics. *arXiv preprint arXiv:2008.09470* (2020).

[44] TM Dumas, M Maxwell-Smith, JP Davis, PA Giulietti, Lying or longing for likes? narcissism, peer belonging, loneliness and normative versus deceptive like-seeking on instagram in emerging adulthood. *Computers in human behavior* **71**, 1–10 (2017).

[45] T Kameda, W Toyokawa, RS Tindale, Information aggregation and collective intelligence beyond the wisdom of crowds. *Nature Reviews Psychology* **1**, 345–357 (2022).

[46] T Sasaki, B Granovskiy, RP Mann, DJ Sumpter, SC Pratt, Ant colonies outperform individuals when a sensory discrimination task is difficult but not when it is easy. *Proceedings of the National Academy of Sciences* **110**, 13769–13773 (2013).

[47] J Baumgartner, S Zannettou, B Keegan, M Squire, J Blackburn, The pushshift reddit dataset in *Proceedings of the international AAAI conference on web and social media*. Vol. 14, pp. 830–839 (2020).

[48] MW Riley, JW Riley, ML Toby, The measurement of consensus. *Social Forces* pp. 97–106 (1952).

[49] M Blackwell, S Iacus, G King, G Porro, cem: Coarsened exact matching in stata. *The Stata Journal* **9**, 524–546 (2009).

[50] SM Iacus, G King, G Porro, Causal inference without balance checking: Coarsened exact matching. *Political analysis* **20**, 1–24 (2012).

[51] DE Ho, K Imai, G King, EA Stuart, Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis* **15**, 199–236 (2007).

[52] ML Anderson, Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion. *American Economic Review* **104**, 2763–2796 (2014).

Supplemental Information

Capturing Judgments Expressed in Comments

In order to capture the judgment expressed in a comment, I searched the comment text for both the set of abbreviations: YTA, NTA, ESH, NAH, and INFO, as well as alternative abbreviations including YWBTA ("You Would Be the Asshole") and YWNBTA ("You Would Not Be the Asshole"). I also searched for full-text equivalents including the following variations: "youre the asshole," "u r the asshole," "ur the asshole," "you would be the asshole," "you'd be the asshole," "not the asshole," "you would not be the asshole," "you'd not be the asshole", "everyone sucks here," "no assholes here," and "not enough info."

Measuring Consensus Strength

Here I breakdown the consensus strength measure into its two components. I reran the main analysis with each component in turn, and compared the results to the results from the model using the composite measure (i.e., the main analysis). Average marginal effects of dissenting with the consensus from these three models are shown in Fig. SI1. First, I ran the analysis using just the proportion variable, representing the score-weighted proportion of comments expressing the majority judgment, ranging from (0.5, 1]. The first graph in Fig. SI1 shows that dissent is valued when the proportion is low, and then agreement valued when the proportion is high. Next, I ran the analysis using just the variable capturing the number of other comments existing at the time of the focal comment. The results here, shown in the second graph of Fig. SI1 show a similar, but more exaggerated pattern, relative to the composite results, which are shown in the third graph of Fig. SI1.

Together, these results point to the benefits of using a composite measure. On its own, the proportion variable does not provide enough information. The results from the proportion variable show that dissent is valued when the proportion is low, and agreement is valued when the proportion is high. The issue with accepting this pattern on its own is that a proportion variable at one extreme or the other could occur with only a small number of comments (representing a consensus that is likely to change as more comments accumulate) or when there are many comments (representing a well-established consensus that is much more robust). It is more informative to contextualize the proportion variable with the number of comments that contribute to the proportion, as this means the consensus strength reflects the evolving number of comments and thus the stage of the decision making. In other words, a "strong" or "weak" consensus depends on both of these components: the first component—the proportion variable—reflects how strong the group majority is, and the second component—the number of comments making up this proportion—reflects how robust (i.e., likely to change) the majority is.

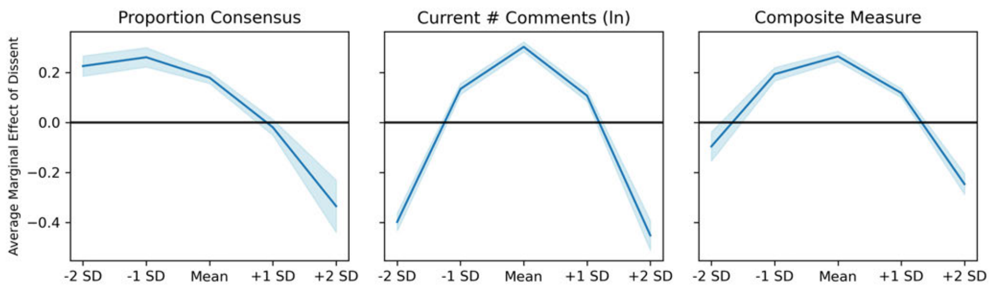

Figure SI1: Breaking down the consensus strength measure. The first graph shows the results using just the proportion variable. The second graph shows the results using just the number of comments (ln) contributing to the proportion. The third graph shows the results using the composite measure.

Descriptive Statistics and Correlation Table for Key Variables

Table SI1: Descriptive Statistics

	Mean	SD	Min	Max
Comment Competition	377.5	591.8	1	8346
Comment Length	285.8	295.1	3	9892
Author Score	15673.0	72886.2	-43816	2846005
Minutes Since Post	782.4	6423.6	0.817	259196
Consensus Strength	4.798	1.675	0.555	9.023
Consensus Dissent	0.0370	0.189	0	1
Observations	6799071			

Table SI2: Correlation Table

	1	2	3	4	5	6
(1) Comment Competition (ln)	1					
(2) Comment Length (ln)	0.0142	1				
(3) Author Score (ln)	-0.322	0.0434	1			
(4) Mintes Since Post (ln)	0.697	-0.00718	-0.297	1		
(5) Consensus Strength (ln)	0.993	0.0107	-0.318	0.687	1	
(6) Consensus Dissent	0.00395	0.0351	-0.0246	0.0464	-0.0403	1

Regression Table for Main Result

Table SI3: Main Regression Results

	(1)
Consensus Strength	-1.725***
	(0.061)
Consensus Strength^2	0.101***
	(0.002)
Dissent	-0.525***
	(0.056)
Consensus Strength × Dissent	0.374***
	(0.024)
Consensus Strength^2× Dissent	-0.044***
	(0.002)
Comment Competition (ln)	0.226***
	(0.057)
Min Since Post (ln)	-0.697***
	(0.017)
Min Since Post (ln)^2	0.045***
	(0.001)
Author Score (ln)	0.150***
	(0.009)
Comment Length (ln)	0.065***
	(0.001)
Constant	5.388***
	(0.101)
Post FE	Yes
Hour FE	Yes
Day of Week FE	Yes
Month FE	Yes
Observations	6,799,071

Note: Standard errors in parentheses are clustered at the post level. Dependent variable is logged comment score. Estimates are from a regression using 2022 data with weights from coarsened exact matching. * p<0.05, ** p<0.01, *** p<0.001 (two-tailed tests).

Robustness Checks

I ran a series of robustness checks to test the robustness of the main result to different sample creation strategies. Fig. SI2 displays the average marginal effects of dissenting with the consensus, across these different samples. These results show that the main pattern is robust across these different samples. I conducted all robustness checks on the non-matched sample, because some of the robustness checks included observations which I excluded when doing the matching procedure. Robustness checks included the following: 1) The main sample without matching. 2) The sample including only comments made before the "official" judgment is posted at 18 hours. 3) The sample including posts that meet a minimum comment threshold of 10 comments rather than 15, which was the threshold used in the main analysis. 4) The sample excluding posts that were deleted or removed by either the original author or the moderators. 5) The sample with the unweighted rather than weighted consensus score. 6) The sample including comments that were made when the consensus was balanced. 7) The full sample of years from 2013 - 2022. Table SI4 shows the regression results for these various robustness checks.

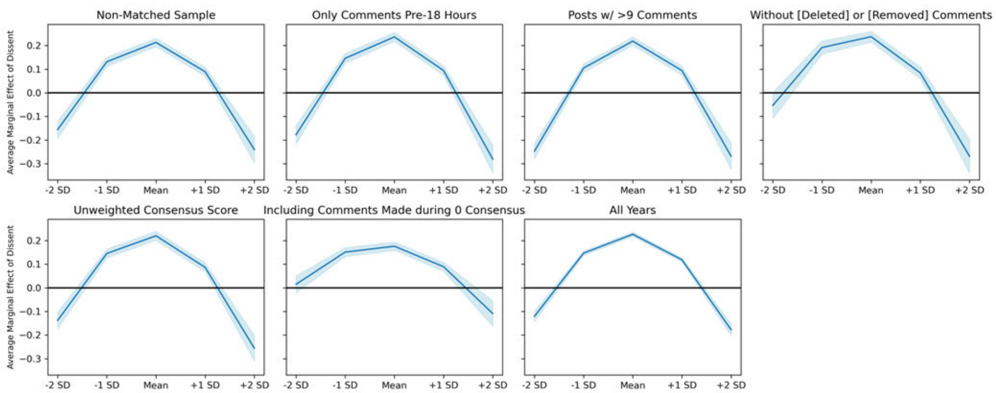

Figure SI2: Robustness checks.

Table SI4: Regression Results from Robustness Checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Non-Matched	Pre-18 Hours	>9 Comments	w/o Deleted/Removed Posts	Unweighted Consensus	w/ Balanced Consensus Comments	All Years
Consensus Strength	-1.045***	-0.758***	-1.020***	-1.298***		-0.754***	-1.401***
	(0.042)	(0.046)	(0.040)	(0.055)		(0.039)	(0.023)
Consensus Strength^2	0.068***	0.045***	0.062***	0.094***		0.036***	0.075***
	(0.002)	(0.003)	(0.002)	(0.003)		(0.002)	(0.001)
Dissent	-0.599***	-0.652***	-0.647***	-0.512***	-0.561***	-0.165***	-0.485***
	(0.041)	(0.042)	(0.031)	(0.063)	(0.040)	(0.036)	(0.022)
Consensus Strength × Dissent	0.348***	0.389***	0.370***	0.333***		0.161***	0.319***
	(0.019)	(0.020)	(0.016)	(0.027)		(0.017)	(0.011)
Consensus Strength^2× Dissent	-0.037***	-0.043***	-0.040***	-0.037***		-0.019***	-0.036***
	(0.002)	(0.002)	(0.002)	(0.003)		(0.002)	(0.001)
Comment Competition (ln)	-0.216***	-0.318***	-0.188***	-0.222***	-0.446***	-0.129***	0.185***
	(0.036)	(0.039)	(0.035)	(0.046)	(0.020)	(0.036)	(0.021)
Min Since Post (ln)	-0.813***	-1.137***	-0.782***	-0.741***	-0.850***	-0.937***	-0.816***
	(0.012)	(0.020)	(0.011)	(0.016)	(0.012)	(0.010)	(0.007)
Min Since Post (ln)^2	0.057***	0.093***	0.055***	0.052***	0.060***	0.064***	0.055***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Author Score (ln)	0.138***	0.138***	0.136***	0.149***	0.139***	0.146***	0.141***
	(0.003)	(0.003)	(0.003)	(0.005)	(0.003)	(0.003)	(0.002)
Comment Length (ln)	0.085***	0.091***	0.087***	0.082***	0.085***	0.092***	0.087***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Consensus Strength Raw					-0.770***		
					(0.027)		
Consensus Strength Raw^2					0.065***		
					(0.002)		
Consensus Strength Raw × Dissent					0.346***		
					(0.020)		
Consensus Strength Raw^2× Dissent					-0.038***		
					(0.002)		
Constant	5.579***	5.835***	5.302***	5.878***	5.555***	4.892***	20.553***
	(0.061)	(0.128)	(0.057)	(0.081)	(0.062)	(0.058)	(0.545)
Post FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hour FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE							Yes
Observations	7,741,416	7,209,646	8,069,059	4,748,336	7,729,811	7,873,728	24404227

Note: Standard errors in parentheses are clustered at the post level. Dependent variable is logged comment score. Estimates are from regressions with each of the following data semples: (1) non-matched sample from 2022, (2) with only comments made before 18 hours after the post, (3) including posts with a minimum comment threshold of 9, (4) without posts that were deleted or removed, (5) using an unweighted consensus strength, (6) including comments made when the consensus was 0, (7) and including all years. * p<0.05, ** p<0.01, *** p<0.001 (two-tailed tests).

Pre and Post Match Balance Table

Table SI5: Pre- and Post-Match Balance Comparison

	Pre-Match Agree		Pre-Match Dissent		Post-Match Agree		Post-Match Dissent	
	mean	sd	mean	sd	mean	sd	mean	sd
Comment Competition (ln)	4.838484	1.674438	4.87339	1.56953	4.867974	1.565502	4.87339	1.56953
Comment Length (ln)	5.140801	1.177336	5.360411	1.25418	5.366316	1.217175	5.360411	1.25418
Author Score (ln)	10.86339	.3823207	10.81385	.3166767	10.81659	.3327121	10.81385	.3166767
Minutes Since Post (ln)	5.351071	1.461141	5.710464	1.414893	5.696405	1.415529	5.710464	1.414893
Consensus Strength	4.811491	1.676084	4.454273	1.598241	4.509558	1.569249	4.454273	1.598241
Observations	6547345		251726		6547345		251726	

Regression Discontinuity in Time

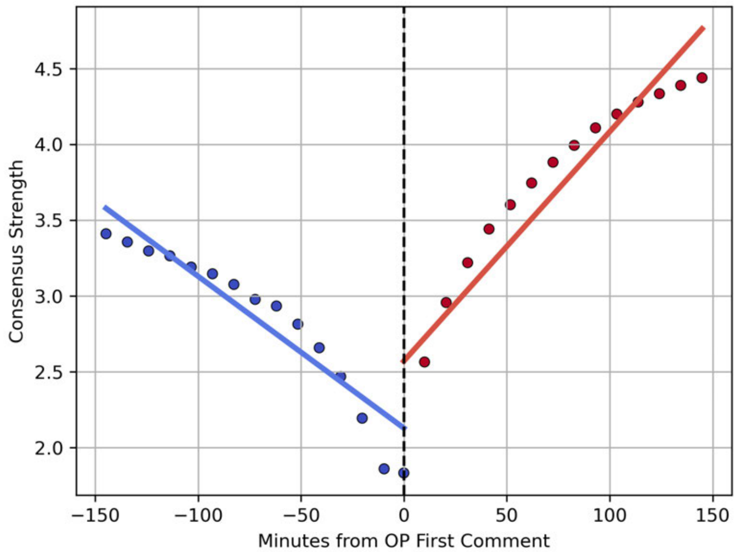

Figure SI3: Shift in the strength of the consensus from before to after the original poster makes a comment. This shows an increase in the strength of the consensus after the OP posted.

Table SI6: Regression Results from RDiT Analysis

	(1)
OP Commented	-0.048***
	(0.007)
Dissent	-0.575***
	(0.020)
OP Commented × Dissent	0.403***
	(0.022)
Min Since OP Com	0.007***
	(0.000)
Min Since OP Com × OP Commented	-0.001***
	(0.000)
Min Since OP Com × Dissent	-0.005***
	(0.000)
Min Since OP Com × OP Commented × Dissent	0.011***
	(0.000)
Min Since Post (ln)	0.640***
	(0.024)
Min Since Post (ln)^2	-0.218***
	(0.003)
Comment Competition (ln)	-0.578***
	(0.014)
Author Score (ln)	0.149***
	(0.008)
Comment Length (ln)	0.200***
	(0.001)
Constant	2.145*
	(0.887)
Post FE	Yes
Hour FE	Yes
Day of Week FE	Yes
Month FE	Yes
Year FE	Yes
Observations	3,564,390

Note: Standard errors in parentheses are clustered at the post level. Dependent variable is logged comment score. Estimates are from regression discontinuity in time analysis. * p<0.05, ** p<0.01, *** p<0.001 (two-tailed tests).

Topic Heterogeneity Analyses

Table SI7: Representative Words from Top2Vec Model

Health	**Neighbors**	**Money**	**Events**	**Chores**	**Dating**	**Bigotry**	**Social Media**	**School**	**Work**
doctors	barking	payment	destination	clean	platonic	lgbtq	instagram	grades	employees
overdose	meowing	money	celebration	sink	flirted	racist	snapchat	class	employee
hospitalized	bark	savings	rsvp	moldy	romantically	conservative	insta	scores	manager
alcoholism	dog	purchase	wedding	cleaning	platonically	lgbt	ig	grade	managers
hospital	neighbor	purchases	festivities	cleanliness	flirting	liberal	chats	graded	boss
father	dogs	debt	venue	unwashed	dating	african	deleted	teacher	staffed
complications	neighbors	cash	rsvpd	cleanest	flirty	homophobic	socials	classmates	supervisor
overdosed	leash	debts	plans	slob	situationship	racism	nudes	physics	bosses
opioids	barks	bank	planned	cleaned	dated	ethnically	dms	professor	supervisors
medications	startles	payments	weddings	cleans	romantic	jewish	memes	assignments	understaffed
alcoholic	banging	investments	inviting	dishes	situationships	homosexuality	deleting	math	hires
congestive	yelping	loan	attend	kitchen	flirt	stereotypes	irl	algebra	corporate
opioid	door	mortgage	rsvped	wash	breakup	christians	delete	geometry	employer
pancreatic	noise	frugal	invitation	washed	hooking	bisexual	discord	classes	customers
rehab	growling	loaned	invited	germaphobe	relationship	transgender	whatsapp	classmate	salaried
induced	leashed	monthly	event	dirty	crush	queer	emojis	students	departments
relapsed	roam	earnings	invitations	scrubbing	fwbs	slang	posted	grading	promoted
liver	knocking	investment	celebrations	washing	fwb	gay	twitter	scored	payroll
drug	fenced	funds	rsvping	tidy	friendzoned	trans	profile	assignment	staff
ectopic	woken	cent	invite	laundry	relationships	atheists	posting	academic	department

Table SI8: Regression Results from Topic Heterogeneity Analyses

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Health	Neighbors	Money	Events	Chores	Dating	Bigotry	Social Media	School	Work
Consensus Strength	-1.669***	-1.770***	-1.538***	-1.467***	-1.699***	-1.564***	-1.543***	-1.820***	-1.690***	-1.589***
	(0.084)	(0.077)	(0.071)	(0.077)	(0.073)	(0.103)	(0.068)	(0.201)	(0.097)	(0.090)
Consensus Strength^2	0.090***	0.104***	0.091***	0.079***	0.098***	0.079***	0.091***	0.073***	0.102***	0.084***
	(0.004)	(0.003)	(0.003)	(0.005)	(0.005)	(0.006)	(0.004)	(0.009)	(0.005)	(0.005)
Dissent	-0.490***	-0.634***	-0.361***	-0.233**	-0.461***	-0.512***	-0.381***	-0.532***	-0.507***	-0.516***
	(0.089)	(0.087)	(0.082)	(0.084)	(0.089)	(0.097)	(0.086)	(0.135)	(0.116)	(0.109)
Consensus Strength × Dissent	0.348***	0.408***	0.277***	0.200***	0.329***	0.316***	0.287***	0.324***	0.334***	0.353***
	(0.039)	(0.042)	(0.037)	(0.036)	(0.040)	(0.045)	(0.039)	(0.068)	(0.054)	(0.054)
Consensus Strength^2× Dissent	-0.040***	-0.046***	-0.033***	-0.023***	-0.039***	-0.034***	-0.034***	-0.035***	-0.037***	-0.040***
	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.008)	(0.006)	(0.006)
Comment Competition (ln)	0.327***	0.292***	0.216***	0.204**	0.301***	0.353***	0.254***	0.639**	0.253**	0.308***
	(0.075)	(0.074)	(0.065)	(0.063)	(0.060)	(0.087)	(0.060)	(0.200)	(0.088)	(0.081)
Min Since Post (ln)	-0.878***	-0.689***	-0.825***	-0.895***	-0.821***	-0.782***	-0.793***	-0.740***	-0.684***	-0.777***
	(0.024)	(0.020)	(0.022)	(0.030)	(0.030)	(0.036)	(0.024)	(0.044)	(0.029)	(0.031)
Min Since Post (ln)^2	0.062***	0.045***	0.057***	0.065***	0.056***	0.051***	0.052***	0.049***	0.044***	0.052***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.004)	(0.002)	(0.003)
Author Score (ln)	0.148***	0.138***	0.117***	0.132***	0.144***	0.118***	0.138***	0.137***	0.124***	0.143***
	(0.003)	(0.004)	(0.017)	(0.003)	(0.003)	(0.005)	(0.004)	(0.007)	(0.005)	(0.005)
Comment Length (ln)	0.079***	0.091***	0.080***	0.071***	0.083***	0.080***	0.080***	0.091***	0.086***	0.093***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
Constant	1.002***	17.163***	19.522***	32.664***	18.288***	16.567***	2.201***	14.212***	12.052***	15.198***
	(0.233)	(1.240)	(0.912)	(1.494)	(1.194)	(1.597)	(0.300)	(1.974)	(1.244)	(1.355)
Post FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hour FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,850,243	1,739,014	2,149,580	2,372,073	2,071,527	1,055,309	1,852,144	484,813	861,396	975,650

Note: Standard errors in parentheses are clustered at the post level. Dependent variable is logged comment score. Estimates are from regressions with data subsetted to each of the following topics: health, neighbors, money, events, chores, dating, bigotry, social media, school, and work. * $p<0.05$, ** $p<0.01$, *** $p<0.001$ (two-tailed tests).