Are Disagreements Just Differences in Beliefs?

Kerem Oktar* (oktar@princeton.edu)

J. Branson Byers* (jbbyers@princeton.edu)

Tania Lombrozo (lombrozo@princeton.edu)

Department of Psychology, Peretsman Scully Hall
Princeton, NJ 08540 USA
*Equal contribution.

Abstract

Decades of research have examined the consequences of disagreement, both negative (harm to relationships) and positive (fostering learning opportunities). Yet the psychological mechanisms underlying disagreement judgments themselves are poorly understood. Much research assumes that disagreement tracks *divergence*: the difference between two individuals' beliefs with respect to a proposition. We test divergence as a theory of interpersonal disagreement through two experiments (N = 60, N = 60) and predictive models. Our data and modeling show that judgments of disagreement track divergence, but also the direction and extremity of beliefs. Critically, disagreement judgments track key social judgments (e.g., inferences of warmth, competence, and bias) above and beyond divergence, with notable variation across domains.

Keywords: disagreement; persistence; social cognition

Introduction

Disagreement is ubiquitous—from petty arguments about where to order dinner to heated debates about fiscal policy, we frequently find ourselves at odds with one another. Though much research has studied the consequences of disagreement, we know surprisingly little about what *constitutes* disagreement. This is because most relevant research across decades and disciplines has implicitly assumed that perceived disagreement is simply a function of divergence in belief, either in binary terms (e.g., if Alex believes that climate change is real, and Sam thinks it is not, they disagree) or as a continuous measure (e.g., if Alex thinks it is 70% likely that climate change is real, and Sam thinks it is 30% likely, their disagreement is 40%). Yet, to our knowledge, no research has empirically investigated what underlies judgments of disagreement themselves—are judgments of disagreement really just about differences in beliefs?

In this paper, we present two experiments that are among the first to investigate the psychological mechanisms that underlie disagreement judgments. Our experiments lead to the following novel contributions. First, we find that perceived disagreement broadly tracks divergence across diverse domains, controversies, and measures. Second, we show that disagreement is nevertheless irreducible to divergence, as divergence misses out on key nuances—such as the extremity and direction of beliefs—that explain substantial variation in disagreement judgments. Third, we find that the nuances lost by treating disagreement as divergence are quite important, in that they hold unique predictive potential over key social judgments, from warmth to competence, bias to conflict. In the General Discussion, we consider the implications of these findings for belief revision, persistence, and polarization.

The Study of Disagreement

Psychologists have studied the consequences of encountering dissent on individual behavior for decades (Asch, 1951). Social psychologists have primarily focused on the negative consequences of dissent, from discomfort (Matz & Wood, 2005) to lower self-esteem (Pool et al., 1998), and from failures of communication (Ziembowicz et al., 2023) to escalating conflict (Kennedy & Pronin, 2008). Cognitive and developmental psychologists have instead often focused on the positive consequences of encountering diverse, conflicting opinions, from the benefits of integrating multiple perspectives in perception tasks (Bahrami et al., 2010), to the importance of transient diversity for problem solving (Smaldino et al., 2023); from advice-taking (Soll & Larrick, 2009) to testimonial learning (Harris, 2012). Beyond the psychological literature, the study of disagreement is of central interest to philosophers (Frances & Matheson, 2019), political scientists (Iyengar & Westwood, 2015), and beyond.

Disagreement as Divergence

Across much of this research, a shared assumption is that disagreement corresponds to a *divergence* of belief with regards to a proposition, as illustrated in the prior example of Alex and Sam having differing beliefs about whether climate change is real. How can we formalize these judgments? In formal epistemology, two dominant approaches to the nature of belief lead to two familiar notions of divergence.

Those who view belief as categorical (e.g., "I believe in climate change") can appeal to *binary divergence*, whereby we disagree if we do not share the same outright view, as in our example of Alex and Sam (Jackson, 2020). Bayesian epistemologists, on the other hand, view belief as a subjective probability judgment (Bovens & Hartmann, 2003). For

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

example, if I believe in climate change, that means I assign a high credence to it, say 80% probability of climate change being real. Now consider someone who does not share this estimate (for instance, they may believe that climate change is 30% likely to be real). *Continuous divergence* corresponds to the difference in our estimates (80% - 30% = 50%).

These intuitive formalizations are often leveraged to operationalize disagreement in studies of its consequences. For example, disagreement over art has been operationalized through binary divergence in two people's stated endorsement of the value of particular pieces (Cheek et al., 2021), whereas disagreements with others over propositions such as whether Brexit will be good for the British economy have been analyzed through continuous divergence in people's probabilistic estimates (Pothos et al., 2021).

What Lies Beyond Divergence

Recent research in political science highlights that "common measurement practices have emerged without sufficient attention having been given to defining disagreement"; moreover, they call for research to examine "the impact that alternative measurements have on models used to evaluate behavioral consequences" (Klofstad et al., 2013). This is in part due to differences in how pivotal research has calculated divergence, with some relying on issue-level continuous divergence as we discussed above (Huckfeldt et al., 2004), and others computing binary and continuous divergence across a broader set of measures (Mutz, 2006), resulting in work that has reached different conclusions about disagreement despite using the same data. For instance, Huckfeldt et al. (2004) concluded that disagreement is the modal condition in the American electorate, whereas Mutz (2006) concluded that there are low levels of disagreement.

The debate in political science exposes one dimension along which divergence may miss out on important nuance: Divergence is *univariate*—that is, typically evaluated over single issues, and hence is insensitive to differences across multiple beliefs and representations (Oktar et al., 2023).

Moreover, divergence is a *linear* measure: a one-point difference in belief, independent of the initial extremity of one's own view, corresponds to the same amount of disagreement. Yet research on attitude strength has shown that middling views lead to different judgments than extreme views in many settings (Howe & Krosnick, 2017).

Finally, divergence is a *symmetric* measure: A one-point difference of belief in any direction thus corresponds to the same amount of disagreement. Yet research has shown that social judgments are highly sensitive to the direction of deviation. For instance, people prefer to associate with others who have more extreme political views than their own, rather than more moderate (Goldenberg et al., 2023).

Cumulatively, these points raise the possibility that divergence alone (see Fig 1) may not capture important nuances in judgments of disagreement. This raises many questions: Does divergence track disagreement at all—and if so, how well? Do the nuances that we miss with binary and continuous divergence influence judgments that we care about, such as reactions to disagreeing others?

Overview of Experiments

We present two experiments that answer these questions. In Experiment 1, we measure perceived disagreement with others who vary systematically in their divergence from each participant's own views. Doing so enables us to examine whether factors such as extremity or directionality capture variation in disagreement judgments above and beyond divergence.

In Experiment 2, we use more realistic stimuli to investigate social reactions to disagreement. We examine whether disagreement judgments capture variation in social judgments (such as warmth and competence) above and beyond divergence.

Experiment 1

Participants reported their beliefs (i.e., subjective probabilities) concerning six issues, encountered 18 characters with differing beliefs about the same issues, and rated how much they disagreed with the characters. Our aims were to test whether binary or continuous divergence were sufficient to characterize and predict perceived disagreement, and to identify additional factors that predict disagreement.

In addition to providing insight into the nature of disagreement judgments, identifying these factors would allow researchers to converge on an evidence-based operationalization of disagreement that can be consistently deployed across interventions and experiments. Moreover, developing a quantitative characterization of disagreement would enable researchers to build models that incorporate disagreement judgments to provide precise predictions.

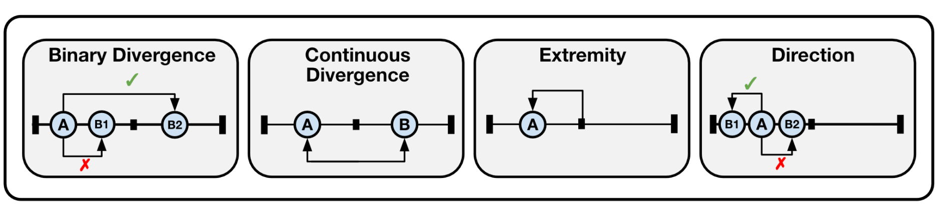

Figure 1. Binary divergence represents whether A and B are on the same side of the midpoint. Continuous divergence captures the absolute value of the difference between A and B. Extremity is formalized as the distance of a belief from the midpoint. Direction indicates whether B is *both* at least as extreme as A and on the same side of the midpoint, or not.

Methods

Participants Participants were 60 adults (27 men, 32 women, 1 other, mean age = 38) recruited on Prolific in exchange for monetary compensation ($1.40 for a 7-minute study). Participation across all studies was restricted to users currently residing in the United States with an approval rating ≥ 98% on at least 100 tasks. Repeat participation within and across studies was restricted using the Prolific platform.

Materials and Procedure Participants were told that they would be presented with the beliefs of randomly selected Americans "collected from 'Attitudes in America,' a project investigating American beliefs across a wide range of issues."

Participants first indicated their beliefs about the six issues on slider scales from 'Definitely False' [0] to 'Definitely True' [100]. The issues were selected to span a variety of domains and degrees of importance, and were sampled from Ransom et al. (2021) and the World Values Survey (2020).

Participants then saw other characters' beliefs indicated on the same scale they used to indicate their own beliefs. These beliefs were systematically generated to span levels of divergence: For each statement, participants saw three characters with beliefs randomly sampled from each tertile (0-33; 34-66; 67-100). Participants indicated "whether and how much [they] agree or disagree with [character]," using a slider from 'Totally Agree' [0] to 'Totally Disagree' [100], with 'Neither Agree Nor Disagree' as a neutral midpoint [50]. They then rated other dimensions of belief using sliders, including importance ("how important to you is your belief?"), confidence ("how confident are you in your beliefs about these issues?"), and belief perseverence ([if given compelling evidence] "would you change your beliefs?"). Participants answered demographics (age, sex, education, religiosity, political affiliation) and were debriefed.

Results

Analytical Strategy: Nested Model Comparisons

To investigate whether divergence (binary or continuous) is sufficient to accurately capture disagreement judgments, or whether additional factors contribute to predictions (see Figure 2), we compared corresponding models.

We considered six models (see Figure 3A for details): *Binary Divergence* and *Continuous Divergence* predicted disagreement from divergence, following much previous literature. Three other models incorporated *Extremity,* capturing non-linearities in disagreement due to attitude strength (Howe & Krosnick, 2017), *Direction,* capturing directional asymmetries in disagreement (Goldenberg et al., 2023), and *Rich Belief,* capturing whether extremity and direction jointly provide additional predictive value. Finally, we investigated whether additional properties of belief (confidence, importance, and robustness) predict additional variance through the *Meta-Belief* model (Tormala, 2016). These models were nested, allowing us to examine the additional contribution of each factor through model comparisons (Figure 3A).

Disagreement judgments spanned the full range and were often at the bounds of the scales (Figure 3B). Given that responses were a complex, continuous mixture of graded and extreme judgments, Zero-One Inflated Beta Regression (ZOIB) was the most appropriate way to analyze judgments. To fit these models, we used the `brms` package in *R* (warmup = 1000, samples = 2000, chains = 4). The same regressors were used for both mean and the Zero-One Inflation term.

Goodness of fit was calculated with Pareto-smoothed-importance-sampling of Leave-One-Out Cross-Validation (PSIS-LOO; Fig 3C). Importantly, using PSIS-LOO allows us to penalize models that account for variance simply by overfitting to the dataset while remaining more robust to weak priors and influential observations than WAIC (Vehtari et al., 2015). This results in the fairest possible comparison of predictive power within a dataset. The output of this calculation is Expected Log Pointwise Predictive Density for a new dataset (ELPD; Less negative indicates better fit).

Is Divergence Sufficient?

The Continuous Divergence Model (ELPD = -225.2 ± 40.2; reported as mean ± standard error across samples; higher is better) predicted disagreement judgements better than the model that included only binary divergence (ELPD = -524.7 ± 30.4).

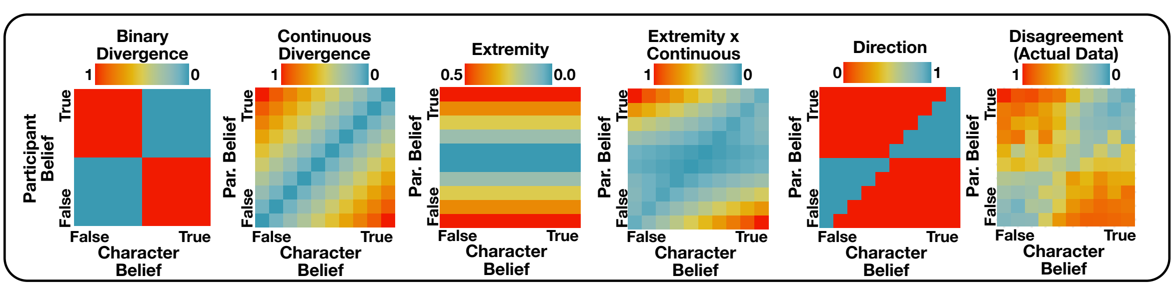

Figure 2. We visualize each regressor across participant and character belief. On the far right, actual disagreement judgements can be compared to each factor. If disagreement were fully predicted by a single regressor (e.g. Binary Divergence), we would expect actual disagreement judgements (far right) to match the predictions of that regressor (e.g. far left). Instead, disagreement appears as a summed combination of multiple regressors, not mapping precisely to any one factor (see Figure 3).

To test whether Extremity and Direction meaningfully predicted disagreement beyond divergence, we compared the Extremity, Direction, and Rich Belief models to a model with both binary and continuous divergence. Extremity outperformed Continuous Divergence (ELPD difference = 77.1 ± 12.9). Even though Direction only marginally outperformed Continuous Divergence (ELPD difference = 3.3 ± 3.3), Rich Belief was the best fitting model overall (ELPD 16.3 ± 6.1 more than Extremity; see Figure 3C).

Beyond asking whether factors beyond divergence help predict disagreement across our entire dataset, we can ask whether *individual* responses are also best characterized by the richer model. To investigate the alignment between individual responses and model predictions, we performed the following analysis. For each judgment of each participant, we used the predictors (e.g., the extent of divergence, extremity) to simulate model predictions (by drawing 20,000 posterior samples, and calculating the posterior mean), with the model parameters coming from the best-fitting population level models. We then computed, for each participant: a) the correlation between their 18 actual disagreement judgments and model predictions across those judgments, and b) the deviation between their judgments and predictions (the mean squared error). We finally computed the proportion of participants that each model best characterized.

The Rich Belief model best described most individual participants (58% of participants' judgments were most strongly correlated with the Rich Belief model, and 71% of participants' judgments deviated the least from the predictions of this model; see Fig. 3D).

We note that Rich Belief marginally outperformed Meta-Belief (ELPD = -1.2 ± 3.3 less than Rich Belief), suggesting that meta-belief measures were redundant with the regressors in Rich Belief. This was confirmed with an ad hoc model comparison by modularly adding the meta-belief responses as regressors to the Rich Belief model, yielding either no significant improvement over the Rich Belief model.

We also note that a logistic regression predicting a binarized agreement vs. disagreement measure replicates the analyses reported above, suggesting that factors beyond divergence do not merely play a role in moderating inferences about the *strength* of disagreement, but also play a role in people's judgments of what *constitutes* disagreement.

Discussion

The results of Experiment 1 establish that a model including both binary and continuous divergence better captures perceived disagreement than binary divergence alone. However, Extremity and Direction capture additional variance as well: above and beyond the magnitude of divergence, participants reported greater disagreement when they held more extreme beliefs themselves, and when others' divergence was towards (vs. away) from the midpoint.

These results indicate that the implicit assumptions behind equating disagreement with divergence do not hold. The relationship between divergence and disagreement is (i) influenced by factors such as extremity, (ii) non-linear, and (iii) asymmetric.

Though Experiment 1 establishes that disagreement is not *statistically* reducible to divergence, it does not speak to whether these statistical differences are *practically* important. Are these differences that make a difference in, say, our ability to predict consequential social judgments?

Experiment 2

In Experiment 2, we aimed to investigate whether richer models of disagreement allow better predictions of important social judgments. To do so, we presented participants with a richer set of 16 controversial statements from science, religion, politics, and morality. Participants provided their own beliefs, and were presented with characters whose views either diverged substantially from their own, or diverged a small amount. Participants then indicated how much they disagreed with these characters and evaluated them on a battery of social judgments (e.g., warmth and competence).

An additional aim of this study was to investigate the relationship between divergence and disagreement using more naturalistic stimuli. In everyday contexts, we do not observe precise estimates of others' views presented on slider scales (as in Experiment 1), but instead infer others' beliefs from verbal statements (e.g., 'It is definitely true that abortion is moral'). We thus presented characters' views verbally.

Finally, we also aimed to address an important question: Does the subject of disagreement influence how disagreement is itself evaluated? Using statements sampled across domains allows us to address this possibility.

Methods

Participants Participants were 60 adults (35 men, 25 women, mean age = 40) recruited on Prolific using the same recruitment criteria used in Experiment 1, but for a longer, 10 minute study.

Materials and Procedure Participants were first assigned to one of four domains (science, religion, politics, morality). Participants read the same "Attitudes in America" cover story used in Experiment 1. They then rated their own beliefs on four issues within the assigned domain on an 8-point truth scale, from 'Definitely False' [1], to 'Definitely True' [8], with no neutral midpoint.

Participants were then shown the beliefs of four other characters, one for each issue in the domain (the items were taken from Oktar & Lombrozo, 2022). Two characters were randomly assigned to express views that diverged substantially (at the opposite end of the scale, or 1 less than the endpoint), and two characters expressed views that diverged little (always one point away from the participant, randomly above or below, but never crossing the midpoint).

Participants first indicated how much they disagreed with the character on a 9-point disagreement scale, from 'Totally Disagree' [1] to 'Totally Agree' [9]. They then provided key judgments from the social psychology literature (see Figure 4) and provided additional belief ratings from Experiment 1.

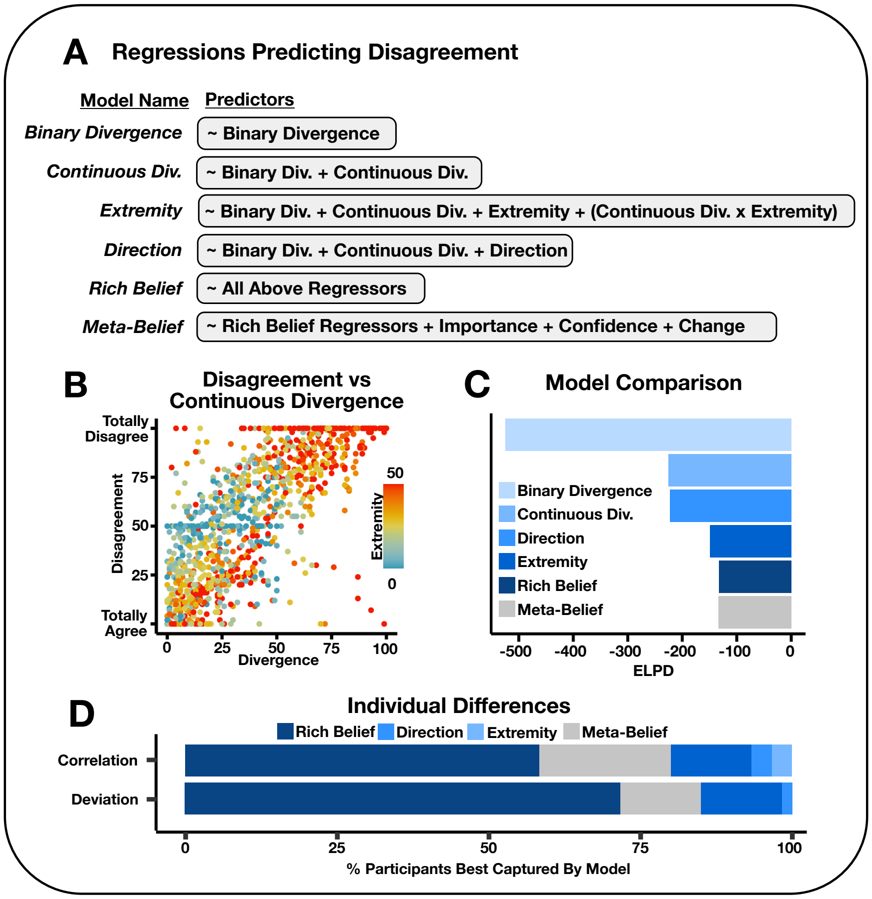

Figure 3. (A) Zero-One Inflated Beta Regression (ZOIB) models progressively include more regressors so that the predictive power of each factor beyond the previous can be assessed through model comparison. For example, inclusion of binary divergence as a regressor in the *Continuous Divergence* model allows us to evaluate how much predictive power continuous divergence adds over binary divergence alone. Note that only the extremity model includes an interaction with divergence—this is because more extreme beliefs cause more agreement for low divergence, but more disagreement for high divergence. The *Rich Belief* model included all of these regressors. The *Meta-Belief* model included importance, confidence, and likelihood to change on top of the Rich Belief model. (B) Disagreement widely tracks continuous divergence, though with additional nuance: There is a clear effect of extremity (red) pushing judgments to the bounds. (C) Larger ELPD values indicate better predictive performance. The Rich Belief model performs the best. We can see that adding regressors incrementally improves model performance, indicating that they add predictive value beyond those already added to the model, though the meta-belief measures do not provide any further predictive power than the other factors of interest. (D) The Rich Belief model best describes a majority of participants. This implies that rich representations of belief better predict individual disagreement judgements – not just judgements in aggregate across individuals. The winning model per individual was selected using the mean squared error (deviation) and correlation of simulations (from models fitted to all subjects) with each individual's data.

Results

Is Divergence Sufficient to Predict Social Judgments?

We tested whether perceived disagreement predicts social judgements above and beyond divergence through model comparisons. We compared linear models predicting each social judgment from either divergence alone or divergence and disagreement. Intuitively, if disagreement is *practically* reducible to divergence, disagreement should not contribute additional predictive power to regressions of social judgments.

We broadly found that disagreement predicts social judgments beyond divergence over social measures across all domains (Figure 4A; we omit exact ELPD values due to space). Furthermore, we found that the extent to which disagreement outperformed divergence alone varied across domains (Figure 4B). For example, divergence better predicted disagreement judgements for warmth for political controversies (ELDP differences = -9.2 ± 4.5), but not for religious ones (ELPD difference = 1.1 ± 0.5; where the simpler model, divergence only, won).

The presence of cross-domain variation addresses a subtle worry: Divergence and disagreement could be noisy proxies of the same latent construct, such that including both improves predictions simply by reducing noise. However, this account would not predict systematic variation across domains. Observing such variation provides further evidence of the complexity of disagreement judgments.

Discussion

The results of Experiment 2 reaffirm our previous conclusion that disagreement is not reducible to divergence. Moreover, the differences between divergence and disagreement are practically consequential: Components of disagreement beyond divergence explain why we see disagreeing others as cold, incompetent, unrelatable, and biased; and how we think conflicts with them can be resolved. Moreover, these relationships show cross-domain variation, providing further evidence for the richness of disagreement judgments.

General Discussion

Are disagreements just differences in beliefs? Prior research across disciplines—from psychology to politics and philosophy—has often implicitly assumed that they are. However, whether we can treat perceived disagreement and divergence in beliefs exchangeably is an empirical question. Our analyses across two experiments show that the answer to this question is 'no.' Disagreement judgments capture substantial, rich properties of beliefs, from how extreme one's own views are, to the direction in which others diverge. Moreover, the differences between divergence and perceived disagreement are not merely statistically significant, but potentially consequential: Disagreement judgments better predict key social outcomes, from whom we judge as cold or incompetent, to which conflicts we take to be resolvable.

Most of these relationships show complex patterns of variation across domains, further revealing the richness of disagreement judgments (and future work should explore the mechanisms driving such cross-domain variation). Beyond demonstrating that disagreement is a richer relation between beliefs than mere difference in subjective probability, our analyses highlight what we miss when we use divergence as a proxy. Divergence is a linear and symmetric function that takes in a single input, whereas our results suggest that disagreement is a non-linear and asymmetric function.

Importantly, our experiments considered both divergence and disagreement with respect to single beliefs, as each character was only evaluated on the basis of a single viewpoint. Yet disagreements unfold over time, and the process of trying to reach mutual understanding typically takes in a much richer representation of the relevant issues—one that involves multiple beliefs. A key question for future research is therefore evaluating how judgments of disagreement respond to the accumulation of additional pieces of evidence about disagreeing others' larger set of beliefs—that is, how do we evaluate and respond to *representational misalignment* across worldviews (Oktar et al., 2023)?

Another important direction for future research is explaining what underlies the complex patterns of cross-domain variation we observed in the relationship between disagreement and other social judgments. Given that domains are abstractions that capture similarities in latent features across issues, the problem here is identifying which issue-level features drive variation in social judgments. Developing models that can generate accurate predictions about how people will respond to novel disagreements requires understanding such variation.

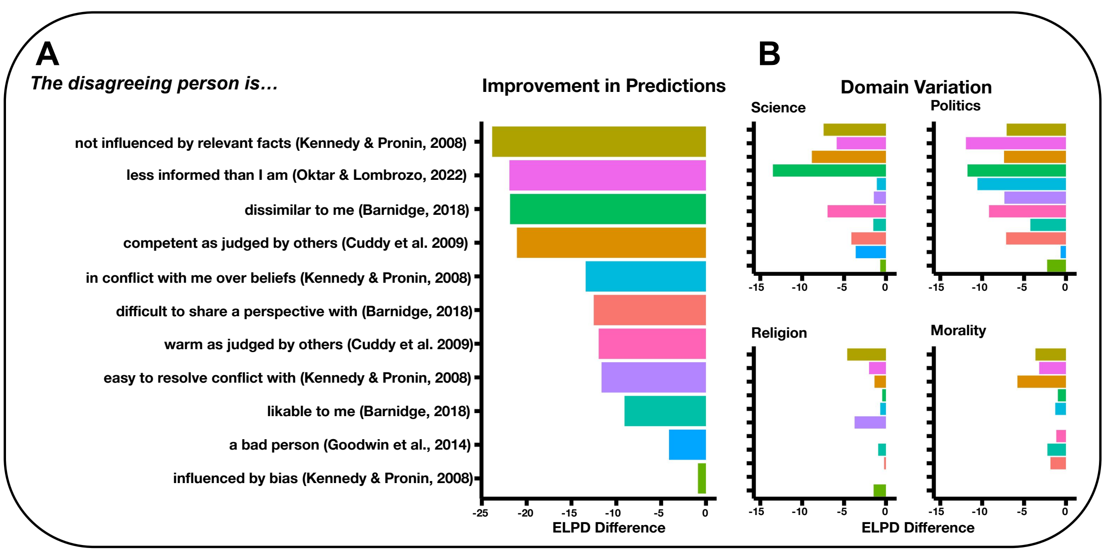

Figure 4. (A) Linear regression of Disagreement + Divergence better predicts social judgments aggregated across domains in comparison to Divergence alone, as reflected in negative ELPD for Disagreement + Divergence. (B) The extent to which disagreement predicts social judgements beyond divergence varies across domains. For instance, disagreement out-predicts divergence on warmth and moral judgement in politics, but not for the religious topics. Overall, the variation in outperformance of disagreement over divergence demonstrates the richness of disagreement judgements.

Acknowledgments

We thank members of the Concepts and Cognition Lab, as well as members of the Niv lab for their valuable feedback.

References

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In *Groups, leadership and men; research in human relations* (pp. 177–190). Carnegie Press.

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally Interacting Minds. *Science*, *329*(5995), 1081–1085. https://doi.org/10.1126/science.1185718

Bovens, L., & Hartmann, S. (2003). *Bayesian Epistemology*. Oxford: Oxford University Press.

Cheek, N. N., Blackman, S. F., & Pronin, E. (2021). Seeing the subjective as objective: People perceive the taste of those they disagree with as biased and wrong. *Journal of Behavioral Decision Making*, *34*(2), 167–182. https://doi.org/10.1002/bdm.2201

Frances, B., & Matheson, J. (2019). Disagreement. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2019). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2019/entries/disagreement/

Goldenberg, A., Abruzzo, J. M., Huang, Z., Schöne, J., Bailey, D., Willer, R., Halperin, E., & Gross, J. J. (2023). Homophily and acrophily as drivers of political segregation. *Nature Human Behaviour*, *7*(2), Article 2. https://doi.org/10.1038/s41562-022-01474-9

Harris, P. L. (2012). *Trusting What You're Told: How Children Learn from Others*. Harvard University Press.

Howe, L. C., & Krosnick, J. A. (2017). Attitude Strength. *Annual Review of Psychology*, *68*(1), 327–351. https://doi.org/10.1146/annurev-psych-122414-033600

Huckfeldt, R., Johnson, P. E., & Sprague, J. (2004). *Political Disagreement: The Survival of Diverse Opinions within Communication Networks*. Cambridge University Press. https://doi.org/10.1017/CBO9780511617102

Iyengar, S., & Westwood, S. J. (2015). Fear and Loathing across Party Lines: New Evidence on Group Polarization. *American Journal of Political Science*, *59*(3), 690–707. https://doi.org/10.1111/ajps.12152

Jackson, E. G. (2020). The relationship between belief and credence. *Philosophy Compass*, *15*(6), e12668. https://doi.org/10.1111/phc3.12668

Kennedy, K. A., & Pronin, E. (2008). When disagreement gets ugly: Perceptions of bias and the escalation of conflict. *Personality and Social Psychology Bulletin*, *34*(6), 833–848.

Klofstad, C. A., Sokhey, A. E., & McClurg, S. D. (2013). Disagreeing about Disagreement: How Conflict in Social Networks Affects Political Behavior. *American Journal of Political Science*, *57*(1), 120–134. https://doi.org/10.1111/j.1540-5907.2012.00620.x

Matz, D. C., & Wood, W. (2005). Cognitive dissonance in groups: The consequences of disagreement. *Journal of Personality and Social Psychology*, *88*(1), 22–37. https://doi.org/10.1037/0022-3514.88.1.22

Mutz, D. C. (2006). *Hearing the Other Side: Deliberative Versus Participatory Democracy*. Cambridge University Press.

Oktar, K., & Lombrozo, T. (2022). Mechanisms of Belief Persistence in the Face of Societal Disagreement. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44). https://escholarship.org/uc/item/3380n01h

Oktar, K., Sucholutsky, I., Lombrozo, T., & Griffiths, T. L. (2023). *Dimensions of Disagreement: Unpacking Divergence and Misalignment in Cognitive Science and Artificial Intelligence* (arXiv:2310.12994). arXiv. https://doi.org/10.48550/arXiv.2310.12994

Pool, G. J., Wood, W., & Leck, K. (1998). The self-esteem motive in social influence: Agreement with valued majorities and disagreement with derogated minorities. *Journal of Personality and Social Psychology*, *75*(4), 967–975. https://doi.org/10.1037/0022-3514.75.4.967

Pothos, E. M., Lewandowsky, S., Basieva, I., Barque-Duran, A., Tapper, K., & Khrennikov, A. (2021). Information overload for (bounded) rational agents. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1944), 20202957. https://doi.org/10.1098/rspb.2020.2957

Smaldino, P. E., Moser, C., Pérez Velilla, A., & Werling, M. (2023). Maintaining Transient Diversity Is a General Principle for Improving Collective Problem Solving. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 17456916231180100. https://doi.org/10.1177/17456916231180100

Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(3), 780–805. https://doi.org/10.1037/a0015145

Tormala, Z. L. (2016). The role of certainty (and uncertainty) in attitudes and persuasion. *Current Opinion in Psychology*, *10*, 6–11. https://doi.org/10.1016/j.copsyc.2015.10.017

Ziembowicz, K., Rychwalska, A., & Nowak, A. (2023). Arguments at Odds—Dyadic Turn-Taking and Conflict Development in Consensus-Making Groups. *Small Group Research*, *54*(4), 551–589. https://doi.org/10.1177/10464964221118674