Diversity and homophily in social networks

Sina Fazelpour (s.fazel-pour@northeastern.edu)
Department of Philosophy and Religion & Khoury College of Computer Sciences, Northeastern University
Boston, MA USA

Hannah Rubin (hrubin@nd.edu)
Department of Philosophy, University of Notre Dame
Notre Dame, IN USA

Preprint, Forthcoming in the *Proceedings of the 44th Annual Conference of the Cognitive Science Society*

Abstract

Diversity of social identities can improve the performance of groups through varied cognitive and communicative pathways. Recently, research efforts have focused on identifying when we should expect to see these potential benefits in real-world settings. While most research to date has studied this topic at individual and interpersonal levels, in this paper, we develop an agent-based model to explore how various aspects of *homophily*, the tendency of individuals to associate with similar others, affects performance at a larger scale. Study 1 examines how two types of homophily—identity-driven and opinion-driven—impact collective performance on a sequential decision-making task via modulating network formation and trust relations. Study 2 considers how the presence of identity-based conformity pressure can affect the findings from the first study. Overall, we find that the effect of homophily on performance is complex, depending on the operative dimensions of similarity, mediating pathways, and the specific outcome of interest. Finally, we discuss the implications of our results for policy interventions aiming to improve group performance.

Keywords: diversity; social networks; homophily; assortativity; trust; conformity; computational modeling

Introduction

Diversity, broadly construed, is important to successful inquiry within groups, ranging from juries and deliberative mini-publics to scientific communities. However, it has recently been noted that, in many cases, there is a gap between the potential benefits of diversity and its observed impacts. Motivated by this "diversity gap," Sulik, Bahrami, and Deroy (2021) argue we should shift our focus from whether, why, and how diversity beneficial to *when* it is beneficial. Previous work has identified a number of personal and interpersonal-level effect modifiers and enabling conditions that shape the presence and extent of diversity's benefits (e.g., the nature and difficulty of the task, diversity mindset, ...) (Phillips, 2017; Page, 2017). Here, we focus on a network-level consideration that though critical remains understudied in current cognitive science literature: homophily—the tendency of individuals to associate with similar others within diverse communities. Because of homophily's impacts, groups with the same demographic composition can have different levels of "local diversity" (Gomez & Lazer, 2019) and behave in radically different ways.

Homophily is a pervasive feature of diverse communities (Jackson, 2010). In particular, scientific communities are homophilic, especially when it comes to co-authorship patterns (Ferber & Teiman, 1980; McDowell & Smith, 1992; Boschini & Sjögren, 2007; del Carmen & Bing, 2000; West, Jacquet, King, Correll, & Bergstrom, 2013; Wang, Lee, West, Bergstrom, & Erosheva, 2019) and citation patterns (Wardle, 1995; Paris, De Leo, Menozzi, & Gatto, 1998; Ghiasi, Mongeon, Sugimoto, & Larivière, 2018). Previous works provide reasons to think that homophily might be beneficial to scientific inquiry: When dealing with difficult or complex tasks, limiting information flow can prevent scientists from erroneously converging on a belief that a less effective treatment is best (Zollman, 2007, 2010) or it can spur a greater number of independent innovations upon which to ultimately build (Derex & Boyd, 2016). And, homophily has been shown to slow the spread of ideas (Golub & Jackson, 2012b), thus hindering information flow in ways that can be beneficial to inquiry.

However, we will show that the relationship between homophily and successful inquiry is more complicated than this reasoning would suggest. This is because homophily can be driven by many different dimensions of similarity (e.g., social identities, attitudes and beliefs, or values) (Monge et al., 2003) and it can manifest in different structural and behavioral effects (e.g., forming connections, trust relations, or desire to conform) (Fazelpour & Steel, 2021). The overall impact of homophily depends on which dimension(s) of similarity are operative, and how their manifestations interact. In this paper, we examine the impacts of two types of homophily—driven by similarity of identities and of opinions. Study 1 examines how these two types of homophily impact collective performance by modulating network formation and trust relations. We consider both the effect of opinion-based trust, for which we develop a novel formalization, and identity-based trust on inquiry. We find that homophilic networks are generally more successful than non-homophilic random networks in this context. Study 2 further considers what happens when we add another potential pathway of homophily's influence, namely, the pressure to conform with others in your social identity group. We find that conformity generally impedes inquiry, and that the effect is more pronounced in homophilic networks and those with identity-based trust.

In addition to exploring when homophily is beneficial to inquiry, we make two further contributions. First, we provide a more fine-grained articulation of the nature of the benefits

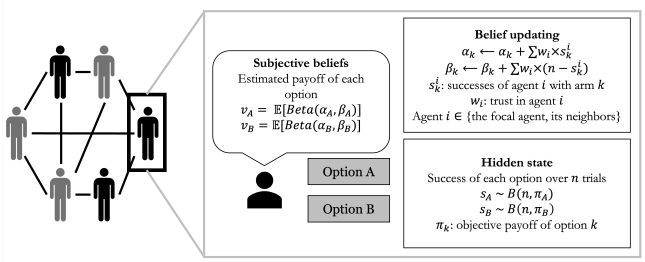


Figure 1: A schematic overview of the agent-based model used in our studies.

themselves. Clearly specifying the outcomes that we value is important not only for determining what we consider beneficial, but also for anticipating and navigating potential tradeoffs between different performance desiderata (e.g. speed of inquiry vs. likelihood of converging on the truth). Second, we complement psychological studies, which give evidence regarding the nature of short term interactions in small groups, by using simulation studies, which give us information regarding outcomes of long term interactions in larger groups.

Study 1

Study 1 examines the impact of two types of homophily—*identity-driven* and *opinion-driven*—on collective performance in sequential decision-making tasks. We investigated this influence along two specific pathways by which homophilic tendencies can impact group performance: (1) *preferential association* with similar others in network formation and (2) *higher trust* in the testimony of similar others. We explored the impacts of both types of homophily—identity and opinion—along the trust pathway, but only considered the network formation effects of identity-induced homophily.

Method

Basic computational model. We constructed an agent-based model to investigate these effects. The agents in our model face a two-armed bandit task, which is a standard formalization of a key type of sequential decision-making task (Sutton & Barto, 2018; Daw, O'doherty, Dayan, Seymour, & Dolan, 2006). At each time point, an agent must decide between one of two options (e.g., a doctor choosing between two choices of treatment). While the agent is unaware of the objective payoffs of the options, it has subjective beliefs about these payoffs. By choosing to experiment with an option, the agent can be thought of as conducting a number of trials and observing the number of successes and failures that ensue. The task is to learn from this feedback at each time point (e.g., number of patients recovered) to find the superior alternative. We model the successes of experimenting with option (or arm), k, at a given time point as a random draw from a binomial distribution, $B(n, \pi_k)$, where n is the number of trials and π_k is k's objective probability of success (Zollman, 2010). The subjective beliefs about the successes of arm, k is modeled as a beta distribution, $Beta(\alpha_k, \beta_k)$.[1] Throughout, we use $n = 1000$ and randomly draw the initial values of α and β for each agent from the uniform distribution $U(0, 4)$.

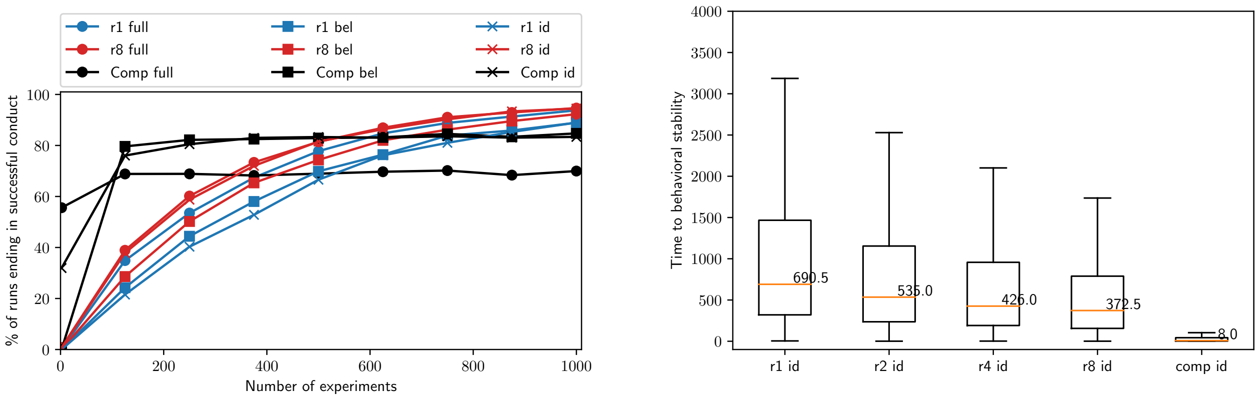
Each agent in our studies can belong to one of two identity groups—a membership that can influence patterns of network formation, trust relation, or both (as described below). To model social relations, the agents are placed on networks of various types (see below). A connection between two agents in the network indicates a direct line of influence between them—e.g., in terms of receiving testimonial evidence from one another or observing each others' behavior. In addition to their direct observations, then, each agent also receives evidence from their neighbors in the social network. In this way, depending on the choice of their neighbors, the agents might also receive evidence about an option they themselves did not choose. At each time point, the agents update their beliefs about the payoff of options by incorporating the *weighted* sum of evidence (i.e., successes and failures observed over n trials) collected by themselves and their neighbors. The weighting on a piece of evidence received from a neighbor depends on the focal agent's trust in that neighbor. Finally, given these belief distributions, agents always choose the option that currently has the highest estimated mean (or, as in the next study, highest overall epistemic and non-epistemic value).[2] Figure 1 provides a schematic overview of the model.

Network formation. We examine collective performance across three general types of network structure: (1) *complete networks*: a fully connected network in which there is a direct link between any two agents, (2) *homophilic networks*: a toplogy where the pattern of connections between agents is shaped by their *identities* (Golub & Jackson, 2012a; Rubin & O'Connor, 2018), and (3) *random* where connections are formed independent of identity. Specifically, in homophilic networks, an agent will be connected to in-groups with a probability p_{in} and to out-groups with a potentially different probability p_{out}. Of course, varying p_{in} and p_{out} changes both (i) the likelihood of identity-based clustering and (ii) a network's overall sparsity. We can disaggregate the impact of these two factors, and focus specifically on (i), by comparing homophilic networks with *different ratios* $\frac{p_{in}}{p_{out}}$, allowing p_{in} and p_{out} to vary such that, in all such networks, regardless of the ratio, the probability of a connection between any two agents, whatever their identities, remains *invariant*. The random networks that we considered correspond to structures where $\frac{p_{in}}{p_{out}} = 1$.

Trust relations between agents. As mentioned above, in this study, we considered two determinants of trust based on (1) shared group *identity* and (2) similarity of

[1] Beta distribution is the conjugate prior for binomial distribution (used here to model observed successes), which makes belief updating easier (Blitzstein & Hwang, 2015).

[2] In other words, the agents are greedy and never explore seemingly inferior options. Exploration thus depends on the evidence from neighboring agents.



(a) (b)

Figure 2: The impact of homophily on (a) efficiency and (b) speed via network formation and trust. *rk* refers to homophilic networks where $p_{in} = k \times p_{out}$ while *keeping the overall connectivity fixed*. "bel" refers to opinion-based trust, "id" to identity-based trust with $w = 0.1$, and "full" to full trust. The complete network is included for contrast purposes. All networks of 40 agents with parity of representation. The first data point in (a) corresponds to 2 experiments.

opinions or like-mindedness. In general, we model trust as a weighting factor, w, in integrating information—successes and failures—arriving to an agent from its neighbors (see Figure 1). In the case of identity-based trust w values remains fixed, insofar as we assume that group identities remain stable in our model. In contrast, since opinions can change as agents gather more observations and update their beliefs, w values for opinion-based trust are dynamic. In both cases, we assume agents fully trust themselves, i.e., adopt a $w = 1$ in weighting their own observations. As a baseline comparison to identity-based and belief-based trust relations, we also consider "full trust" cases where the agents uniformly trust all their neighbors, i.e., $w = 1$ for all dyadic relations.

Identity-based trust. We follow Fazelpour and Steel (2021) in modeling the impact of identity-based trust on information integration. According to this model, while the evidence from in-group neighbors is treated as if it was directly observed (i.e., $w = 1$ for in-groups), agents give relatively less weight to evidence arriving from out-groups. That is, the successes and failures reported about an option by an out-group neighbor is weighted by a fixed factor, $0 \leq w \leq 1$.

Opinion-based trust. In order to formalize dynamic trust, we need to specify how agents can estimate like-mindedness given the information available to them. A natural way of doing so is for agents to simply track behavioral similarities with others. Specifically, let $A_n^i = \{a_1^i, a_2^i, ..., a_n^i\}$ be agent i's action sequence up to and including experiment n. Agent i can estimate a neighbor j's like-mindedness by comparing their action sequences, such that i's trust in j at n is given by

$$w_n^{ij} = \frac{\sum_{t=1}^{n} \mathbb{1}[a_t^i = a_t^j]}{n}$$

where $\mathbb{1}[.]$ is an indicator function that takes the value of 1 when the agents perform the same action and is 0 otherwise. Accordingly, agents' trust in others ranges between 0 and 1: an agent fully trusts a neighbor when they have taken exactly the same action at each time point. Conversely, agents will have no trust in a neighbor, if they have always chosen different options.

Experimental design and procedures. We explored the impact of the two types of homophily along network formation and trust pathways across a wide range of parameter settings. In terms of network topologies, in addition to complete network structures, we examined homophily networks with ratios $\{1, 2, 4, 8\}$ (constructed as described above with $r2$ corresponding to a network with $[p_{in}, p_{out}] = [0.4, 0.2]$). While a network with ratio 1 amounts to a random network with no homophilic tendencies, a ratio of 8 indicates that agents are 8 times more likely to connect to in-group others. For identity-based trust, we varied w between 0.05 and 1 (increments of 0.05). Note, however, that below we often refer to the case of $w = 0.1$ as identity-based trust. Finally, throughout, we kept the objective probability of payoff for the two options fixed with $\pi_A = 0.499$ and $\pi_B = 0.5$.

In examining the impact of homophily, we consider three aspects of group performance characterized in terms of three outcomes:

- *Reliability:* The percentage of simulation runs ending in correct, unanimous consensus.
- *Speed:* The number of experiments until the last change in choice behavior.
- *Efficiency:* Evaluated by comparing reliability at different time horizons (or number of experiments).

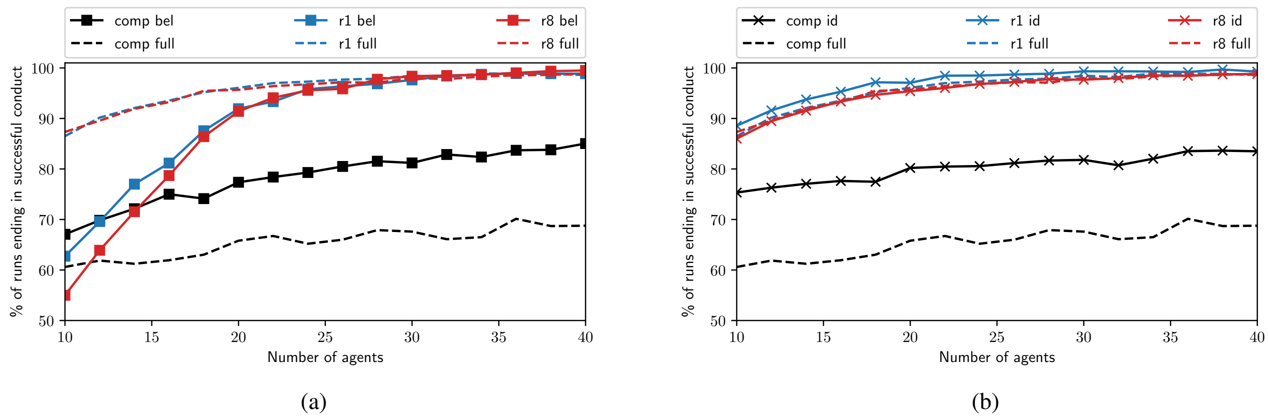


Figure 3: The impact of homophily on reliability across networks of different size. Homophily's impact on performance (a) via opinion-based trust ("bel") and (b) via identity-based trust with $w = 0.1$ ("id"). The dashed lines, used as comparative baselines, depict the impact in identity-driven network formation in the case of full trust. The complete network is included for contrast purposes.

Results and discussion

We find sparser networks (random as well as various degrees of homophily) to be more conducive to successful inquiry at longer time horizons when compared to the complete network. This is in line with previous studies (Zollman, 2007; O'Connor & Weatherall, 2019) and can be explained by the fact that in complete networks the transmission of misleading results is swift and widespread. As a result, agents in complete networks are particularly susceptible to reaching premature consensus on the wrong option. This is made less likely by sparsity in network connection, as can be seen by comparing the longer term performances of networks with full inter-agent trust in Figure 2a ("comp full" vs. "r1 full" and "r8 full"). Importantly, the transmission of (mis)information can also be slowed down by lowering levels of trust. Hence, as the figure shows, the performance of complete networks improves when agents' trust behavior is governed by either identity-based or opinion-based considerations (as opposed to full trust). This is similar to findings by Fazelpour and Steel (2021), though they only examine the identity-driven case. As depicted in Figure 2a, however, the performance of sparser networks and networks with lower than full trust comes at a cost to speed. Accordingly, whether homophilic networks (or homophilic trust relations) are superior to complete networks (or fully trusting relations) depends on the outcome of interest (speed vs. success).

We did not find any appreciable differences in reliability, or success at longer time horizon between networks with varying degrees of homophily (when keeping the type of trust fixed). We did, however, find that across all types of trust, increased identity-driven associations *increased* the *speed and efficiency of learning*, which was particularly salient in the case of identity-driven trust (see Figure 2b).

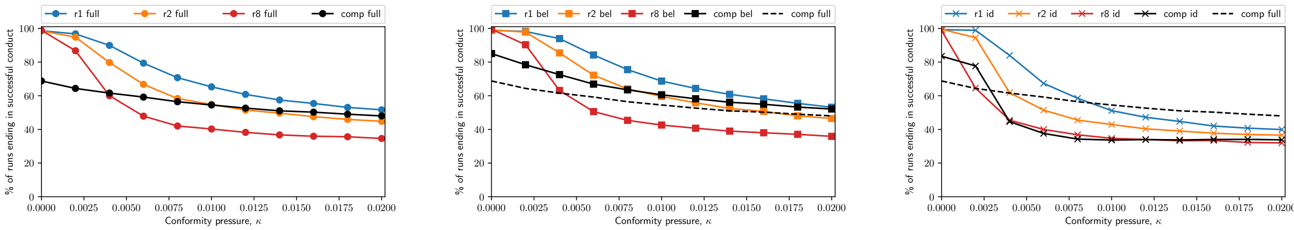
While increasing the size of the network improved performance (both in terms of reliability and speed) across all network and trust types, the pattern of growth in sparser networks with opinion-driven trust stands out ("$r1$ bel" and "$r8$ bel" in Figure 3a). Specifically, while identity-driven (as opposed to full) trust had no appreciable impact on successful performance in these networks, opinion-driven trust was highly detrimental in smaller group sizes. A possible explanation is that the reduction of trust in neighbors who are not seen as like-minded is particularly problematic when the number of neighbors is small to begin with. In such cases agents can quickly end up receiving only evidence that confirms their beliefs. A closer look at our findings supports this explanation: a substantial portion of simulation runs in these networks end up with general polarization (i.e., cases where the collective ends with clusters of opposing, stable opinions that do not fall along identity lines). In "$r1$ bel" networks, for example, such outcomes constitute 32% of runs in groups of size 10, 26% of runs in groups of size 12 (compared to 4% and 3% respectively, in "$r1$ id", where trust is driven by identity).

Study 2

In study 2, we consider how the presence of identity-based conformity can impact the results from the previous section. Such (normative) conformity pressures are prevalent in real-world settings (Cialdini & Goldstein, 2004; Deutsch & Gerard, 1955), and they are critical to incorporate, because in such settings we often do not have the capability to pick and choose which of homophily's effects we'd like to be present.

Method

When conformity pressures are present, the behavior of agents is no longer a faithful reflection of their beliefs. As a result, in addition to formalizing how conformity pressures



(a) (b) (c)

Figure 4: The impact of identity-induced conformity on reliability in (a) full trust, (b) opinion-based trust ("bel"), and (c) identity-based trust with $w = 0.1$ ("id"). All networks of 40 agents with parity of representation.

shape agent behavior, we also need to consider the impact of this effect on other aspects of behavior (e.g., opinion-based trust) and relevant outcomes (e.g., polarization).

Conformity in diverse networks. While conformity in general is a much studied topic in social psychology, the study of conformity's impact in identity diverse groups is relatively recent (Phillips, Mannix, Neale, & Gruenfeld, 2004; Phillips & Loyd, 2006; Gaither, Apfelbaum, Birnbaum, Babbitt, & Sommers, 2018). Here we use the formalization of conformity's impact on individual decision-making in diverse groups in Fazelpour and Steel (2021):

$$u_j^i = (1-\kappa)\times v_j^i + \kappa \times \frac{\mathcal{N}_{in}^i(j)}{\mathcal{N}^i}$$

Where u_j^i represents the total perceived value of pursuing option j for agent i. v_j^i is agent i's perceived expected payoff of option j (see Figure 1). $\mathcal{N}^i$ is the total number of i's neighbors and $\mathcal{N}_{in}^i(j)$ are the subset of neighbors who share the same group identity with i (i.e., are considered in-group by i) that pursued option j in the previous time point. Finally, κ represents i's conformist tendency. When $\kappa = 0$, agents simply follow their personal beliefs, but when $\kappa = 1$, agents just follow the majority decision from in-group majority.[3]

This formulation is a modification of the "Other-Total Ratio" (Stasser & Davis, 1981) that is meant to capture two key findings about conformity's impact in identity diverse groups: (1) individuals primarily feel the (normative) pressure to conform to in-groups (reflected in the numerator of the fraction) (Antonio et al., 2004); and (2) the *mere presence* of out-group individuals, regardless of their views, reduces conformity pressure (reflected in the fixed denominator of the fraction) (Phillips, 2017; Gaither et al., 2018).

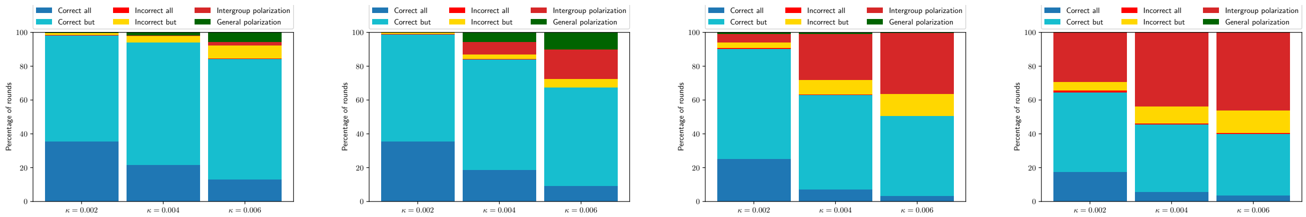
Opinion-based trust in the presence of conformity. The presence of conformity complicates our formulation of opinion-based trust as perceived "like-mindedness", since agents might act contrary to their beliefs because of in-group conformity pressure. In this case, agent i observing neighbor j acting in the same way will not necessarily convey j's "like-mindedness" to i. In fact, one could imagine this surface agreement to *increase* i's distrust in j. To deal with this type of scenario, we assume that, instead of considering what it actually did, the focal agent i compares what it would have done had there been no conformity pressure with j's actual actions. The agents thus adopt an asymmetric attitude towards their own versus others' conduct, downplaying the influence of situational factors (i.e., conformity pressure) in the case of others, but not in their own case. While this is clearly a simplification, as the literature on fundamental attribution error in social psychology shows, in many circumstances people do seem to act in similar ways (Ross, 1977).

Experimental design and procedures. We explored how the presence of identity-based conformity might influence the outcomes of previous section by varying the extent of conformity pressure between 0 and 0.02 (with increments of 0.002).[4] The presence of conformity requires that we adopt a more fine-grained lens on dependent outcomes. We introduce six new categories of dependent outcomes:

- *Correct all:* Simulation runs that end with all agents pursuing the superior option and believing in their choice.
- *Correct but:* Simulation runs that end with all agents pursuing the superior option, despite the fact that some agents do so as a result of conformity and against their beliefs.
- *Incorrect all:* Simulation runs that end with all agents pursuing the inferior option and believing in their choice.
- *Incorrect but:* Simulation runs that end with all agents pursuing the inferior option, despite the fact that some agents do so as a result of conformity and against their beliefs.
- *Inter-group polarization:* Simulation runs that end with (belief) consensus within identity groups and opposing views between groups.
- *General polarization:* Simulations runs that end with no consensus (in general or within groups).

[3] Throughout, we use the same κ for all agents.

[4] Given the small different between the objective payoff of the two options, anything outside this range simply amounts to purely conformist behavior.



(a) r1 opinion-based trust | (b) r1 identity-based trust | (c) r8 opinion-based trust | (d) r8 identity-based trust

Figure 5: The impact of identity-induced conformity on performance disaggregate by outcome types in random networks (r1) with (a) opinion-based, (b) identity-based trust; and homophilic networks (r8) with (c) opinion-based and (d) identity-based trust. All networks of 40 agents with parity of representation.

Results and discussion

As shown in Figure 4 and consistent with previous studies (O'Connor & Weatherall, 2018; Fazelpour & Steel, 2021), we find normative conformity to be detrimental to successful performance across all network and trust types. Importantly, our results go beyond previous findings that were mainly focused on the impact of conformity in *complete* networks. In particular, we find that the extent of conformity's detrimental impact critically depends on network structure (e.g., random vs. homophilic) and trust type (e.g., full vs. identity- vs. opinion-driven). Specifically, conformity is particularly detrimental with increased structural homophily (i.e., $r8$ vs. $r2$) and identity-driven trust (and worse still when these are combined). This is to be expected, since homophilic networks in effect decrease diversity in the neighborhood of an agent, thus increasing the conformity pressure on that agent. Identity-based trust exacerbates this situation by preventing agents to learn about the potential superiority of alternative courses of action from the testimony of out-groups.

Perhaps surprisingly, opinion-based trust *curtails* the negative influence of conformity even in homophilic network structures. A possible explanation is that while opinion-driven trust cannot decrease the normative influence of in-group conformity, it can decrease the *epistemic* impact of in-groups when the agent disagrees with them. At the same time, it can lead agents to trust like-minded out-group members. Accordingly, agents are more likely to form correct beliefs or at least beliefs that are in line with certain out-group members. This can in turn result in agents pursuing the superior option and an overall decrease in inter-group polarization. A comparison of the extent of inter-group polarization at lower levels of conformist tendency κ provides support this explanation (see Figure 5).

General Discussion

We find that the relationship between homophily and collective performance is complicated; whether it is beneficial depends both on its particular manifestation and on how we characterize the community's aims. Additionally, whether a certain factor is beneficial depends on the presence of other mediating factors. For example, interestingly, we find that opinion-driven trust impedes convergence to truth in Study 1, but the effect flips in the presence of identity-driven conformity where opinion-driven trust is beneficial. Taking these factors into account has implications for how we think about implementing policy proposals aimed at increasing diversity.

Diversity according to social identity has been shown to be important to inquiry, resulting in many arguments that we ought to promote demographic diversity because of the ensuing gains in effective inquiry or performance of groups. Arguments of this sort – referred to variously as "the business case for diversity" (Steel & Bolduc, 2020) or "private sins as public goods" arguments (Schneider, Rubin, & O'Connor, 2021) – presume that promoting equity of a certain kind will go hand-in-hand with receiving the benefits of diversity (Steel & Bolduc, 2020). As such, it may convince those only interested in epistemic gains to incidentally promote socially beneficial policies (Schneider et al., 2021).

There are complications with employing this sort of argument; many of the benefits of demographic diversity may rest on problematic aspects of social interactions, e.g. lack of trust or devaluation of testimony from marginalized social identity groups (Fazelpour & Steel, 2021; Wu, 2021), and attempts to promote diversity by intervening on the structure of communities may backfire and further entrench inequity (Schneider et al., 2021). Here, we have shown that there is a further complication: whether a certain feature of socially diverse communities is beneficial may depend on its particular manifestation. These challenges support the calls for caution about framing diversity's value in this instrumental way, as opposed to other, perhaps more fundamental social and ethical rationales (Fazelpour & De-Arteaga, 2022; Starck, Sinclair, & Shelton, 2021; Phillips, 2017)

We find that certain manifestations of homophily may need to be managed or mitigated in diverse communities, depending on what benefits one hopes to see. Since different dimensions and impacts of homophily can be co-present, disentangling the sources and consequences of homophily in real-world communities is key. While restricted by modeling simplifications and abstractions, the simulations presented here offer an important first step in this direction, providing theoretical insights that can guide complementary empirical research on which aspects of homophily impede or promote successful inquiry and under what circumstances.

References

Antonio, A. L., Chang, M. J., Hakuta, K., Kenny, D. A., Levin, S., & Milem, J. F. (2004). Effects of racial diversity on complex thinking in college students. *Psychological Science*, *15*(8), 507–510. doi: 10.1111/j.0956-7976.2004.00710.x

Blitzstein, J. K., & Hwang, J. (2015). *Introduction to probability*. Boca Raton, Florida: CRC Press.

Boschini, A., & Sjögren, A. (2007). Is team formation gender neutral? evidence from coauthorship patterns. *Journal of Labor Economics*, *25*(2), 325–365.

Cialdini, R. B., & Goldstein, N. J. (2004). Social Influence: Compliance and Conformity. *Annual Review of Psychology*, *55*(1), 591–621. doi: 10.1146/annurev.psych.55.090902.142015

Daw, N. D., O'doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876–879.

del Carmen, A., & Bing, R. L. (2000). Academic productivity of African Americans in criminology and criminal justice. *Journal of Criminal Justice Education*, *11*(2), 237–249.

Derex, M., & Boyd, R. (2016). Partial connectivity increases cultural accumulation within groups. *Proceedings of the National Academy of Sciences*, *113*(11), 2982–2987.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, *51*(3), 629–636. doi: 10.1037/h0046408

Fazelpour, S., & De-Arteaga, M. (2022). Diversity in sociotechnical machine learning systems. *Big Data & Society*, *9*(1), 20539517221082027.

Fazelpour, S., & Steel, D. (2021). Diversity, trust and conformity: a simulation study. *Philosophy of Science*.

Ferber, M. A., & Teiman, M. (1980). Are women economists at a disadvantage in publishing journal articles? *Eastern Economic Journal*, *6*(3/4), 189–193.

Gaither, S. E., Apfelbaum, E. P., Birnbaum, H. J., Babbitt, L. G., & Sommers, S. R. (2018). Mere Membership in Racially Diverse Groups Reduces Conformity. *Social Psychological and Personality Science*, *9*(4), 402–410. doi: 10.1177/1948550617708013

Ghiasi, G., Mongeon, P., Sugimoto, C., & Larivière, V. (2018). Gender homophily in citations. In *23rd international conference on science and technology indicators (sti 2018)(september 2018)* (pp. 1519–1525).

Golub, B., & Jackson, M. O. (2012a). How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics*, *127*(3), 1287–1338.

Golub, B., & Jackson, M. O. (2012b). Network structure and the speed of learning measuring homophily based on its consequences. *Annals of Economics and Statistics/ANNALES D'ÉCONOMIE ET DE STATISTIQUE*, 33–48.

Gomez, C. J., & Lazer, D. M. (2019). Clustering knowledge and dispersing abilities enhances collective problem solving in a network. *Nature communications*, *10*(1), 1–11.

Jackson, M. O. (2010). *Social and economic networks*. Princeton university press.

McDowell, J. M., & Smith, J. K. (1992). The effect of gender-sorting on propensity to coauthor: Implications for academic promotion. *Economic Inquiry*, *30*(1), 68–82.

Monge, P. R., Contractor, N. S., Contractor, P. S., Peter, R., Noshir, S., et al. (2003). *Theories of communication networks*. Oxford University Press, USA.

O'Connor, C., & Weatherall, J. O. (2019). *The misinformation age*. Yale University Press.

O'Connor, C., & Weatherall, J. O. (2018). Scientific polarization. *European Journal for Philosophy of Science*, *8*(3), 855–875.

Page, S. E. (2017). *The diversity bonus*. Princeton University Press.

Paris, G., De Leo, G., Menozzi, P., & Gatto, M. (1998). Region-based citation bias in science. *Nature*, *396*(6708), 210–210.

Phillips, K. W. (2017). Commentary. what is the real value of diversity in organizations? questioning our assumptions. In *The diversity bonus* (pp. 223–246). Princeton University Press.

Phillips, K. W., & Loyd, D. L. (2006). When surface and deep-level diversity collide: The effects on dissenting group members. *Organizational Behavior and Human Decision Processes*, *99*, 143–160. doi: 10.1016/j.obhdp.2005.12.001

Phillips, K. W., Mannix, E. A., Neale, M. A., & Gruenfeld, D. H. (2004). Diverse groups and information sharing: The effects of congruent ties. *Journal of Experimental Social Psychology*, *40*, 497–510. doi: 10.1016/j.jesp.2003.10.003

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In *Advances in experimental social psychology* (Vol. 10, pp. 173–220). Elsevier.

Rubin, H., & O'Connor, C. (2018). Discrimination and collaboration in science. *Philosophy of Science*, *85*(3), 380–402.

Schneider, M. D., Rubin, H., & O'Connor, C. (2021). Promoting diverse collaborations. In G. Ramsey & A. De Block (Eds.), *The dynamics of science: Computational frontiers in history and philosophy of science.* Pittsburgh University Press.

Starck, J. G., Sinclair, S., & Shelton, J. N. (2021). How university diversity rationales inform student preferences and outcomes. *Proceedings of the National Academy of Sciences*, *118*(16).

Stasser, G., & Davis, J. H. (1981). Group decision making and social influence: A social interaction sequence model. *Psychological Review*, *88*(6), 523–551. doi: 10.1037/0033-295X.88.6.523

Steel, D., & Bolduc, N. (2020). A closer look at the business case for diversity: The tangled web of equity and epistemic benefits. *Philosophy of the Social Sciences*, *50*(5), 418–443.

Sulik, J., Bahrami, B., & Deroy, O. (2021). The diversity gap: when diversity matters for knowledge.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Wang, Y. S., Lee, C. J., West, J. D., Bergstrom, C. T., & Erosheva, E. A. (2019). *Gender-based homophily in collaborations across a heterogeneous scholarly landscape.*

Wardle, D. A. (1995). Journal citation impact factors and parochial citation practices. *Bulletin of the Ecological Society of America*, *76*(2), 102–104.

West, J. D., Jacquet, J., King, M. M., Correll, S. J., & Bergstrom, C. T. (2013). The role of gender in scholarly authorship. *PloS one*, *8*(7), e66212.

Wu, J. (2021). Epistemic advantage on the margin.

Zollman, K. J. (2007). The communication structure of epistemic communities. *Philosophy of science*, *74*(5), 574–587.

Zollman, K. J. (2010). The epistemic benefit of transient diversity. *Erkenntnis*, *72*(1), 17.