*Philosophy of Science* (2022), **89**, 209–231
doi:10.1017/psa.2021.25

ARTICLE

# Diversity, Trust, and Conformity: A Simulation Study

Sina Fazelpour[1]* and Daniel Steel[2]

[1]Department of Philosophy and Religion and the Khoury College of Computer Sciences, Northeastern University, US and [2]W. Maurice Young Centre for Applied Ethics, University of British Columbia, Canada
*Corresponding author. Email: s.fazel-pour@northeastern.edu

(Received 13 November 2019; revised 09 August 2020; accepted 27 January 2021; first published online 12 January 2022)

## Abstract

Previous simulation models have found positive effects of cognitive diversity on group performance, but have not explored effects of diversity in demographics (e.g., gender, ethnicity). In this paper, we present an agent-based model that captures two empirically supported hypotheses about how demographic diversity can improve group performance. The results of our simulations suggest that, even when social identities are not associated with distinctive task-related cognitive resources, demographic diversity can, in certain circumstances, benefit collective performance by counteracting two types of conformity that can arise in homogeneous groups: those relating to group-based trust and those connected to normative expectations toward in-groups.

## 1. Introduction

As many societies become increasingly diverse along social and ethnic lines, it is important to study the impacts of diversity on group performance. A prominent line of research, pursued especially by philosophers of science and computational social scientists, has been to employ agent-based models (ABMs) for studying diversity's impact on collective performance in simulation settings (Hong and Page 2004; Weisberg and Muldoon 2009; Zollman 2010b). Despite differences in approach, the notion of diversity is understood in essentially the same way in these models, namely, in terms of the variety of cognitive repertoires—background knowledge, problem-solving heuristics, decision rules, etc.—that group members bring to bear on the common task. Let us use the term *cognitive diversity* in referring to diversity in this sense.

This line of research has been a fruitful one, and in some cases its results have been used to support practical or policy recommendations (Grim et al. 2019). Yet, when turning to simulation-based diversity research, societal interest mainly pertains to the potential effects of increasing underrepresented groups or *demographic diversity*. And, given their focus on diversity in a cognitive sense, current ABMs provide little direct evidence about how demographic diversity might influence group performance. Such an approach might be reasonable if demographic diversity only stood to benefit group performance by increasing cognitive diversity. And indeed, this assumption is common in the fields of psychology, sociology, and organizational research (see Steel et al. 2019 for a review of the relevant literature). Demographic diversity can be epistemically beneficial, according to this view, only when it "correlates with or causes germane cognitive diversity" (Page 2017, 9).

A number of empirical researchers have questioned the dominant view, however. They have argued that, at least in certain settings, demographic diversity can enhance group performance, even when its influence is not mediated by cognitive diversity (Bear and Woolley 2011; Phillips 2017). According to one proposal (Levine et al. 2014; Phillips and Apfelbaum 2012), demographic diversity can be beneficial because it counteracts certain detrimental group influences that may plague homogeneous groups. For example, agents in homogeneous groups tend to put too much trust in each other's testimony, resulting in a lack of diligence in processing information from social sources. Similarly, in homogeneous groups, agents may refrain from expressing dissenting perspectives for the fear of disapproval from other group members, leading to an unwillingness to share novel and productive ideas. While presenting intriguing possibilities about the benefits of demographic diversity, a thorough assessment of the hypotheses has been hampered by a variety of factors including relatively small sample sizes, the difficulty of completely decoupling demographic from cognitive diversity, as well as the complications involved in assessing the longitudinal effects of increased diversity in a group of interacting agents, where potential benefits may be overwhelmed by obstacles such as conflict and lack of trust.

In this article, by augmenting a model developed by Zollman (2010b), we construct a model for testing two mechanisms suggested by the hypotheses that demographic diversity can positively impact group performance, even in the absence of any correlation with cognitive diversity. In section 2, we explain how the hypothesized mechanisms are supposed to work and why agent-based modeling is a useful tool for examining these hypotheses. Section 3 introduces the general setup of the simulations. In section 4, we examine the first hypothesis, according to which demographic diversity positively influences group performance by reducing the excessive mutual trust between group members. In section 5, we examine the second hypothesis that diversity epistemically benefits the group by reducing the conformity pressure felt by group members.

The results of the two simulations lend support to the main claims of these hypotheses, but with important qualifications. We also provide a general discussion of the results in section 6. In particular, we (1) highlight some of the limitations of the model; (2) critically examine the uptake of the results of agent-based models that do not represent demographic identity; and (3) provide a suggestion for the use of robustness analysis in evaluating simulation studies that seek to provide empirical insight and practical guidance. Specifically, given that most social phenomena are highly context-sensitive, we suggest that modeling results should not be expected to be robust in general but only robust in ways that correspond to the empirical evidence.

## 2. Benefits of diversity, costs of homogeneity

Understanding the benefits of diversity requires appreciating the potential costs of homogeneity (Phillips and Apfelbaum 2012). Chief among these costs is conformity.

https://doi.org/10.1017/psa.2021.25 Published online by Cambridge University Press

At a general level, conformity refers to a change in belief or behavior in response to real or imagined pressures to resemble others (Cialdini and Goldstein 2004; Gilovich et al. 2018). Seemingly similar conformist behaviors can arise for different reasons, however. A useful initial distinction can be made between *informational* and *normative* social influences (Deutsch and Gerard 1955).

Informational and normative influences arise for distinct reasons, have different effect modifiers, and impact different aspects of an agent's psychology. Informational influences arise in conditions involving ambiguous tasks and uncertain decision environments (see, e.g., Sherif 1936). What drives conformity pressure in response to these influences is the need for others as sources of information about reality. Accordingly, the strength of the resulting conformity pressure can be modulated by factors that impact the need for and reliance on others as sources of information, such as the nature and difficulty of the task, as well as the perceived ability and expertise of others, compared to self (Crano 1970; Hogg and Abrams 1988). By influencing the relative reliability assigned to various informational sources and so shaping the integration of information arriving from these sources, such influences predominantly impact belief formation. Hence, informational influences typically induce conformity in public behavior as well as private belief (Kelman 1958).

Normative social influences, in contrast, depend on an individual's desire to be liked by their group or at least not be punished by them (Asch 1956; Gaither et al. 2018). This type of social influence is typically accompanied by *normative expectations* regarding what other members of the group, whose approval or disapproval one cares about, believe one *ought* to do (Bicchieri 2017). The strength of the resulting conformity pressure is thus a function of the group's real or perceived reward and punishment power as well as the extent to which the agent is moved by this power, possibly as a result of the overall reward structure of the environment (e.g., the benefit and cost of conformist versus nonconformist behavior in the situation) (Raven 2008). By shaping the costs and rewards of pursuing different courses of action, the locus of normative social influences is deliberation and decision making. As a result, with normative social influences, private acceptance and public compliance can come apart (Gilovich et al. 2018); agents may act in agreement with the majority position because of normative expectations and *despite* their private beliefs.

The question of who counts as part of some individual's social "group" depends on which type of influence is under discussion. Insofar as informational influences are concerned, a group member may be anyone whose judgment is deemed by the agent to have a bearing about a task. In the case of normative influences, on the other hand, conformity depends on an individual's sense of belonging to a group whose approval and disapproval they care about. While in principle the sense of "groupiness" is predominantly implicated in normative influences, in practice, matters are more complex. Specifically, when an individual perceives in-group members as more trustworthy sources of information about the world, group identification also determines who gets to informationally influence the individual and to what extent. This is precisely what happens in *referent* informational influence (Turner 1982; Turner, Wetherell, and Hogg 1989). Here, members of the group with which the agent identifies are judged to be more competent and trustworthy and their views are taken to be more worthy of consideration. In this way, judgments of trust and reliability stemming from group affiliation also shape individuals' belief revisions.

https://doi.org/10.1017/psa.2021.25 Published online by Cambridge University Press

As a number of experiments have shown, normative and referent informational conformity are relevant to the question of how demographic diversity might impact group performance (Gaither et al. 2018; Levine et al. 2014; Phillips 2017; Phillips, Liljenquist, and Neale 2009). These experiments are designed to break the link between demographic and cognitive diversity, usually by testing group performance on contrived tasks (e.g., solving a murder mystery or trading real estate in a fictional market) wherein relevant information can be supplied exclusively by the experimenters and for which individual ability can be tested and statistically controlled for. These researchers take their results to suggest that demographic diversity can improve performance even when social identity is not associated with task-related information or other relevant cognitive differences (Levine et al. 2014; Phillips 2017). Two types of hypotheses emerge from this literature:

1. **Counteracting referent informational group influences.** Individuals may put too much trust in socially similar others, uncritically accepting others' views and frequently reconsidering their own view in response to the majority position. By decreasing excessive trust in social sources, demographic diversity can result in a more diligent assessment of information (Gaither et al. 2018; Levine et al. 2014).
2. **Counteracting normative group influences.** Individuals in homogeneous groups may be reluctant to voice dissenting opinions, perhaps as result of a normative expectation that people similar to themselves ought to agree with one another. Demographic diversity can reduce this conformity pressure, allowing groups to successfully elicit dissenting views that are valuable to the task (Phillips, Liljenquist, and Neale 2009; Phillips and Loyd 2006).

Consider two experiments that illustrate these hypotheses. In the first (Levine et al. 2014), experimental subjects traded real estate in a fictional online market in ethnically diverse or homogeneous conditions. Participants were provided with the information needed to assess the market value of the properties they were trading, and their individual ability to accurately price properties was assessed prior to engaging in trade with other participants. Before trading, participants sat in a room with their group-mates so that they could see who they would be trading with. The primary findings were that ethnically diverse groups were more likely to price properties accurately and were less prone to bubbles than ethnically homogeneous groups of traders, results that could not be explained by differences of individual trading ability between groups. "Homogeneity," Levine et al. suggest, "imbues people with false confidence in the judgment of coethnics, discouraging them from scrutinizing behavior. In contrast, traders in diverse markets … are less likely to accept inflated offers and more likely to accept offers that are closer to true value, thereby thwarting bubbles" (Levine et al. 2014).

In the second example we consider (Phillips, Liljenquist, and Neale 2009), experimental subjects were presented with the task of finding the culprit in a murder mystery. Membership in a sorority or fraternity was used as a marker for social identity. Participants read the case and indicated the most likely suspect before being placed in groups. Each group consisted of three individuals from the same fraternity or sorority, dubbed "oldtimers," who would have 5 minutes to discuss who the culprit was prior to the arrival of a newcomer, who could either be an in-group or out-group

https://doi.org/10.1017/psa.2021.25 Published online by Cambridge University Press

individual. Groups were arranged so that the newcomer had 0, 1, or 2 "opinion allies" among the oldtimers who agreed with the newcomer about the most likely suspect. Thus, the experiment was a 2 $\times$ 3 design: two types of newcomer (in-group or out-group) and three conditions (0, 1, or 2 opinion allies). For each condition, groups were more likely to correctly identify the culprit when the newcomer was an out-group individual, although participants of groups with out-group newcomers reported lower confidence in their answers and perceived their groups' interactions as less effective. Moreover, the advantage for groups with out-group newcomers was enhanced significantly when the newcomer had 1 or 2 opinion allies. Phillips et al. suggest that this effect is related to oldtimers paying greater attention to views expressed by out-group than in-group newcomers. Prior work (Phillips and Loyd 2006) found that people are more likely to be annoyed when dissenting views are expressed by members of their own identity group, which could lead to paying greater attention to a dissenting view when it is expressed by an out-group person. That in turn could explain why groups with out-group newcomers would perform better in the 0 opinion ally condition. Phillips et al. suggest that the larger positive effect of out-group newcomers in the 1 and 2 opinion ally conditions is due to the desire to maintain positive relationships with fellow in-group individuals. They hypothesize that while agreement with an in-group newcomer is likely to be taken as merely confirming one's views, finding oneself siding against a member of one's own group by agreeing with an out-group individual is likely to generate anxiety about damaging in-group ties. Phillips et al. suggest that social risks involved in taking sides against one's identity group may result in more careful attention to reasons that underlie the differences of views and, consequently, to better results. In sum, Phillips et al.'s interpretation of their results focuses on normative conformity: people tend to think that members of their own identity group should see eye to eye, which can result in dismissing dissenting opinions when voiced by in-group newcomers or to a heightened concern to probe disagreements when one is unexpectedly allied with an outsider against a member of one's own group.

But a variety of uncertainties confront the experiments such as those just described. The sample sizes involved are typically rather small (rarely more than a few hundred people), and it is also difficult to ensure that the connection between demographic and cognitive diversity has been completely severed. Regarding the latter point, consider providing participants with relevant information and testing for individual ability on the task, as in the Levine et al. experiment. Yet it is conceivable that differing social identities might be statistically associated with distinct problem-solving heuristics and that a diversity of heuristics improves group performance even if, at an individual level, one heuristic does not tend to do better than another (Hong and Page 2004). Furthermore, some of the findings pertain to the conformity-related impact of diversity on *individual* cognition in the sense that, in some studies, other than a focal participant the rest of the group is composed of confederates of the experimenters. This restricted setup means that the studies may neglect the potentially harmful effects on *group* cognition that can arise as a result of dynamic interactions between individual members. This is important because counteracting referent informational and normative conformity is not without potential costs. With respect to referent informational influences, diversity decreases trust in others because of an unjustified view of out-group members as relatively incompetent and

https://doi.org/10.1017/psa.2021.25 Published online by Cambridge University Press

less worthy of trust (Putnam 2007). While excessive trust in others can be epistemically problematic (Langfred 2004), *mutual distrust* between group members can also prove detrimental to the collective performance (Ashleigh and Nandhakumar 2007). Similarly, given conformity's positive influence on group cohesion, a reduction in conformity can also harm the collective. Indeed, disagreements in heterogeneous groups can lead to group polarization along subgroup affiliations (Abrams et al. 1990), resulting in solidarity within subgroups and strife between them. These considerations make it difficult to assess the internal validity and robustness of experiments reporting benefits of demographic diversity that are not mediated by cognitive diversity.

By offering a precise way of testing the implications of different proposed causal dependencies, simulation methods provide a promising way forward that can complement human experiments. In simulation experiments, large sample sizes can be attained at minimal cost, and experimenters enjoy a state of near omniscience regarding the circumstances of the experiment (e.g., ensuring that there are no hidden cognitive differences between diverse and homogeneous groups). Moreover, in simulation experiments, relevant modifying factors can be easily and precisely manipulated, thereby allowing the robustness of results across various conditions to be explored. Of course, simulation experiments also possess a major disadvantage in not having actual humans as their subjects, which requires reliance on simplified and idealized assumptions about cognitive processing and social interaction. Inferences from results of simulation experiments to human social behavior are thus on firmer ground when they reproduce results of human experiments. This is so not only for general effects (e.g., reducing referent informational conformity can improve group performance), but also for the impact of effect modifiers (e.g., the effect is larger when the proportions of demographic groups are closer to parity). Accordingly, we take such agreements to reinforce the results of human experiments. Conversely, a divergence of results between simulations and human experiments indicates a problem of the sort noted by Duhem (1954): it is an indication that an error is present, although it is unclear precisely where.

The simulation results described below are presented in this spirit of the complementary relationship between human experiments and formal models.

## 3. Simulation setup

The model developed here builds on the models discussed in Zollman (2010b; see also Zollman 2007), extending them to include factors such as group affiliation, trust, and conformity. This section provides a general overview of the models.

### 3.1. Two-armed bandit

In our model, agents are faced with a two-armed bandit problem (Berry and Fristedt 1985); each agent is attempting to discover the better of two available options by sequentially choosing an option, experimenting with it, and observing the result. While agents are unaware of the objective payoff of each option, they nonetheless have subjective estimates that inform their choice. Each experimentation involves $n$ trials, whose results are represented by a random draw from a binomial distribution with parameters corresponding to the chosen option's intrinsic payoff and $n$. The agents are thus active inquirers whose choice of which option to pursue shapes their

https://doi.org/10.1017/psa.2021.25 Published online by Cambridge University Press

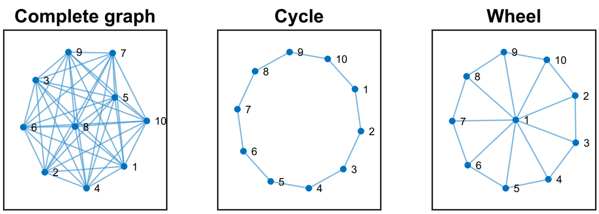


**Figure 1.** Different network topologies representing different types of social interactions.

subsequent observations; if they don't pursue an option, they won't learn anything new about it. Unless, that is, they receive that information from their neighbours.

## 3.2. Social networks

An agent's neighbours are those with whom the agent has direct social ties. The social interaction among agents is represented using graphical networks, where each agent is depicted as a node in the graph, and communicative pathways between agents are represented by undirected edges between the nodes. Figure 1 shows three typical network topologies that will be of interest to us. Whereas in a *complete*, fully connected network every agent is in direct communication with every other, in a *cycle* or ring network, connections are sparser, and each agent is only connected to two neighbours. Despite this difference, in both complete and cycle networks, all agents have an equal number of neighbours. Not so for the *wheel* graph. Here one agent is more central compared to others, and there is variability in the number of neighbours each agent has.

Two points are worth emphasizing: First, thanks to their social ties, agents may get to "virtually" explore options they themselves did not choose to pursue. But, such social learning also opens up the possibility of being *misled* by others' results. Second, we can use different network properties to represent different aspects of group behavior. For instance, some aspects of intergroup dynamics can be captured by modulating the *connection weights* between nodes.

## 3.3. Learning

Having gathered new information through observation and testimony, agents revise their estimates by Bayesian updating. We represent agents' beliefs in terms of continuous random variables with *Beta distributions*, which are specified by two parameters, $\alpha$ and $\beta$, with a mean, $\mu$, that can be calculated as $\frac{\alpha}{\alpha+\beta}$. The Beta distribution is the conjugate prior of the binomial distribution, which is the type of distribution from which our agents' observations are drawn, rendering the distribution amenable to efficient Bayesian updating (Sutton and Barto 1998). Specifically, if the agent's prior is given by $Beta(\alpha, \beta)$, and after $n$ trials, the agent observes that an option was successful $s$ times (according to the evidence gathered by individuals as well as by their neighbours), the agent's posterior is given by $Beta(\alpha + s, \beta + n - s)$ with a mean centred on $\frac{\alpha+s}{\alpha+\beta+n}$.

## 3.4. Decision rule

In experiments that follow, we restrict our investigation to agents who are greedy in their choice behavior, always selecting the option with the highest expected utility. The only source of exploration for agents, therefore, is the information they receive

https://doi.org/10.1017/psa.2021.25 Published online by Cambridge University Press

from neighbours that followed an alternative option. In the simplest case, the only determinant of an option's utility is that option's estimated payoff. To be sure, in real life this way of conducting oneself is reckless; choice behavior is shaped by considerations about a host of issues other than the perceived "efficacy" of actions. To varying extent and depending on the context, our public conduct also reflects norms of etiquette, concerns about the approval and disapproval of others, and so on (Bicchieri 2017; Turner 1991). Such factors become especially important in section 5 when we look into choice behavior of agents with a sense of belonging to groups with reward and punishment powers.

## 4. Simulation 1: Diversity and group-based trust

The first simulation examines what happens when diversity modulates the trust relation between individuals with different group affiliations. Recall that some level of mutual distrust is hypothesized to be beneficial because it reduces the conformity pressure due to referent informational influences, enabling individuals to deal more critically with social information. In this way, diversity is supposed to be advantageous to group performance, even in the absence of any association with cognitive diversity.

### 4.1. Design

Agents in this simulation act on the basis of the simple decision rule discussed in section 3.4; in choosing which option to experiment with, their behavior is fully determined by their beliefs about the payoffs of the two options.[1]

#### *4.1.1. Representing diversity's impact on referent informational influences*

Individuals are most impacted by informational social influences during social learning as they integrate information from various sources in the context of challenging tasks. As a result, informational influences typically result in belief change and internalizing the group's perspective (Deutsch and Gerard 1955; Kelman 1958). In the case of *referent* informational influences, agents give more weight to the testimony of in-groups, possibly because they perceive socially similar others to be more competent and reliable than out-groups (Turner 1982; Turner, Wetherell, and Hogg 1989).

To represent the impact of diversity, agents in the network are divided into two groups, the proportions of which can be varied. To control for the effect of interaction between group membership and beliefs, agents are randomly assigned prior beliefs about the value of the two alternatives.[2] Agents in each group apply a "trust factor" $w$ between 0 and 1 to data received from members of the other group. So, $w = 1$ means complete trust between groups, $w = 0$ complete distrust, and $0 < w < 1$ an intermediate state of distrust. Notice, first, that any $w < 1$ is unjustified distrust of out-group relative to in-group others inasmuch as group membership is irrelevant to an agent's

[1] Each experiment consists of $n = 1000$ trials. The outcome of each agent's experimentation is drawn from the binomial distribution representing the intrinsic payoff of the method chosen by the agent.

[2] $\alpha$ and $\beta$ for each belief were randomly drawn from the range of $[1, 4]$. As discussed in section 3.3 the small values of $\alpha$ and $\beta$ results in high variance, meaning that agents lack confidence in their prior beliefs. This is precisely the situation where informational influences are most effective (Hogg and Abrams 1988).

https://doi.org/10.1017/psa.2021.25 Published online by Cambridge University Press

competence to produce accurate results in the model. Second, given that more weight is given to evidence from in-group neighbours, agents are more likely to reconsider their positions in response to evidence from in-groups.

#### *4.1.2. Dependent measures*

**Probability of success.** Success was defined at the network level and under a time constraint: a population of agents is deemed successful just when it reaches a correct, unanimous consensus after a fixed number of experiments. The probability of successful learning is the proportion of simulation runs in which the population is successful in this sense.[3] We examined how the probability of successful learning varied according to the level of group-based trust, represented by the parameter $w$. We explored this relationship for different group sizes, different proportions of the two group, and different network structures (e.g., the cycle, the wheel, and the complete graph), as well as for different time constraints.

### 4.2. Results

#### *4.2.1. Group-based trust in complete networks*

Figure 2 shows the impact of varying the amount of trust toward out-group individuals on the reliability of the collective performance. As shown in Figure 2a, regardless of their size, networks with high levels of intergroup distrust ($w = 0.005$) were the worst collective performers. Interestingly, however, the second worst collectives were networks wherein individuals placed complete trust in their neighbours, regardless of their group affiliation ($w = 1$). Furthermore, as depicted in Figure 2b, group-based distrust has more positive effects when group proportions are closer to parity. This finding coincides with the empirical finding that effects of increased diversity are more positive when groups are more evenly balanced than when one group is numerically predominant (Bear and Woolley 2011; Reagans and Zuckerman 2001).

#### *4.2.2. Group-based trust and network structure*

The potential benefits of identity-induced distrust with respect to the reliability of collective performance varies from one network structure to another. In particular, the type of mutual distrust induced by diversity is most beneficial in highly connected

[3] Option $B$ had a higher intrinsic payoff (0.501) than option $A$ (0.5). So the network was successful if every agent estimated $p_B$ to be greater than $p_A$ after a fixed number of experiments. The simulation was iterated for 10,000 runs, each involving 3,000 experiments, each consisting of 1000 tries with the chosen option. We follow Zollman (2007, 2010b) in choosing the number of trials and experiments. These numbers provide the accumulation of sufficient evidence whereby *homogeneous* groups with different structures can reach their full potential; further experimentation would not enhance the performance of such groups. These numbers, and assessing performance under the time constraint represented by them, are important for our purposes because they provide a meaningful baseline for assessing how interventions on the composition of homogeneous groups to increase diversity would impact their performance. For example, using such a performance measure allows us to evaluate the worry, discussed in section 2, that a lack of trust between group members undermines group cohesion and hinders identity-diverse groups from achieving their full performance potential that, absent lack of trust, would have been possible given the cognitive repertoires of their members. That said, we also examine the interaction between the impact of distrust and time constraints, specifically in the context of discussing Figure 3a. However, we advise the reader to be mindful of the limitations of the model, discussed in section 6, when interpreting the longer-term performance.

https://doi.org/10.1017/psa.2021.25 Published online by Cambridge University Press

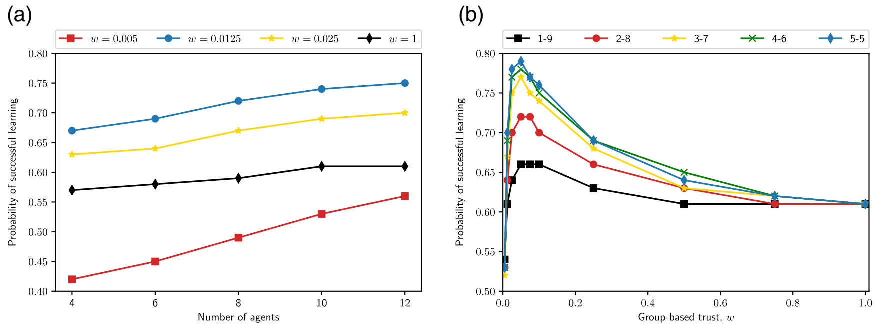


**Figure 2.** Impact of group-based trust, *w*, on collective performance in complete networks. (a) Performance in groups of different size varies as a function of group-based trust. In each case, the network is composed of agents from two equally represented groups. (b) Performance changes as a result of varying the group composition in complete networks of size 10 (1–9 indicates that the network included 1 out-group member).

networks such as the complete graph. For other graph structures, the effect of distrust on performance is marginal before becoming highly negative at lower levels of *w* (see Figure 3a). Reliable convergence to truth is only one aspect of effective performance, however. It also matters how *quickly* groups arrive at a consensus. Viewed from this perspective, group-based distrust is particularly inimical to sparser networks such as the cycle (Figure 3b).[4] In complete networks, in comparison, intermediate levels of distrust produce performance gains while maintaining, to a large extent, the speed of convergence—a feature that makes networks with high levels of connectivity attractive in many circumstances.[5]

## 4.3. Discussion

Simulation 1 supports the view that demographic diversity can improve group performance by reducing the excessive trust individuals place in others.[6] The fact that this reduction of trust is most beneficial in highly connected networks is telling: information travels quickly and widely across highly connected networks. But so does misinformation. So, the misleading result of any one agent can spread quickly across the network, potentially derailing the epistemic pursuit of all the neighbours. This is precisely why sparser networks such as the cycle and the wheel outperform the complete

[4] In some networks the same group composition (e.g., a 50–50 composition) can be instantiated in different patterns; in a cycle network, for example, the members of the two subgroups may be interspersed throughout the network, or each subgroup may form a cluster. With respect to the general trends shown here, the placement did not make much of a difference.

[5] This is not true of high levels of distrust. In complete networks of size 10 and parity of representation, the performance of groups with $w = 0.001$ remains inferior to groups with fully trusting individuals even after 20,000 experiments. In contrast, as shown in Figure 3b, it takes less than 200 experiments for similar networks but with $w = 0.1$ to surpass the reliability of fully trusting networks.

[6] In difficult learning tasks, such as the one studied here, uniform $w = 1$ is *excessive* in the sense that there exists a lower value of *w* that, were it adopted by all individuals, would result in better group performance. The group optimal *w* depends on various contextual factors, including task difficulty and network structure. While in simulations such as ours, the group optimal *w* can be approximated using various search strategies, the problem can quickly become intractable in real-world settings.

https://doi.org/10.1017/psa.2021.25 Published online by Cambridge University Press

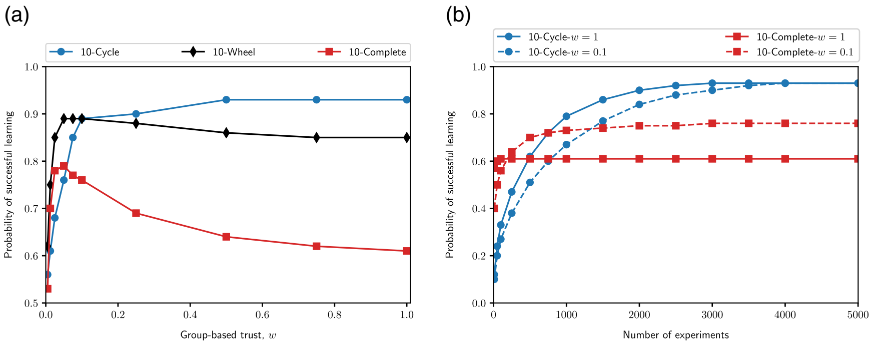


**Figure 3.** Influence of group-based trust, *w*, on the performance of networks of 10 agents with varying graphical structures. (a) The impact of trust on the performance of cycle, complete, and wheel networks. (b) The impact of trust on performance as a function of time constraints—between 10 to 5,000 rounds of experiments, after which performance is probed. The figure thus depicts the influence of trust on the speed of reaching a correct consensus in complete and cycle structures.

graph; because of their sparsity, such networks quarantine misleading results, protecting agents from frequently changing their minds in light of such results. Nonetheless, given the longer time it takes for information to travel across sparser networks, their superior performance comes at the cost of slower convergence to a solution (Zollman 2007, 2010b). Indeed, if what we value is swift convergence to a solution, distrust is particularly harmful to sparser networks.[7]

Angere and Olsson (2017) show that group performance in complete networks improves if individuals agree to only share novel, high-quality evidence. Zollman Zollman (2010b) achieves similar results with "highly confident" agents[8], whose beliefs are not easily changed. By slowing down the speed of convergence to consensus, however, this reduced sensitivity to misinformation comes at a cost. Our simulation shows that similar benefits and tradeoffs can arise when individuals exhibit differential processing of social information. By reducing the weight assigned to information arriving from a subset of social sources, specifically those seen as dissimilar, diversity protects individuals from hastily revising their views in response to social information in ambiguous decision environments.

## 5. Simulation 2: Diversity and normative conformity

Even when individuals manage to remain unswayed by the majority's opinion, they may still refrain from *acting* in nonconformist ways as a result of normative group influences. The second simulation evaluates the hypothesis that diversity can be beneficial to group performance by reducing the cost of expressing dissenting views.

[7] For example, while not shown in Figure 3b, the performance of cycle networks of $w = 1$ and $w = 0.1$ remains unchanged even when probed after 10,000 and 20,000 experiments. This means, at least in the type of setting considered here, distrust only slows such groups from achieving their full performance potential without resulting in tangible performance benefits later.

[8] That is, agents whose beliefs are characterized by high $\alpha$ and $\beta$ values.

https://doi.org/10.1017/psa.2021.25 Published online by Cambridge University Press

### 5.1. Design

The behavior of agents in the previous simulation was a pure reflection of beliefs that were shaped, in part, by informational group influences. Normative group influences, in contrast, put pressure on individuals' *decision making*: individuals are now sensitive to the approval and disapproval of the group, and public conduct is no longer a perfect indicator of private acceptance (Kelley 1952). As discussed in section 3.4, taking such influences into account requires modifying agents' decision rules.

#### *5.1.1. Representing diversity's impact on normative group influences*

There is an extensive psychological literature seeking to mathematically represent the impact of conformity as a function of factors such as majority size (for a review, see Bond 2005). Here, we employ the so-called "Other-Total Ratio," which has been especially useful for capturing conformity pressure due to *normative* influences (Mullen 1983; Stasser and Davis 1981).[9]

Suppose agent $i$ is part of a community of size $\mathcal{N}_i$, consisting of $i$'s direct neighbours as well as $i$ itself. If in the previous round, $\mathcal{N}_i^A$ of the community publicly pursued option $A$, we can think of the pressure on $i$ to choose $A$ in the next round as a function of $\frac{\mathcal{N}_i^A}{\mathcal{N}_i}$: there is minimum incentive to follow the option that wasn't pursued by anyone and maximum incentive to choose the option chosen by all. The overall utility of pursuing $A$ for agent $i$ is now a function of $i$'s belief about $A$'s chance of success, $p_i^A$, as well as the social incentive of doing so:

$$u_i(A) = (1-\kappa) \times p_i^A + \kappa \times \frac{\mathcal{N}_i^A}{\mathcal{N}_i}$$

where $\kappa$ represents agent $i$'s conformist tendency. When $\kappa = 0$, agents act on the basis of their personal beliefs, paying no heed to what others are doing or what is normatively expected of a member of the community. In contrast, when $\kappa = 1$, truth no longer matters; what agents do is a matter of a popularity contest in their group.[10] In the simulations that follow, we assume that all agents have the same conformist tendency.

However, previous applications of the "Other-Total Ratio" have been to homogeneous groups. Some adaptation, then, is required to apply the measure to heterogeneous groups, as we do here. Our approach is to assume, first, that individuals are only concerned to conform with members of their own groups, and second that the presence of out-group individuals, regardless of what views they express, tends to dilute pressure to conform. Regarding the first of these points, the opinions of in-groups are

[9] Among formal models of conformity in philosophy, Mohseni and Williams (2019) use a similar majority ratio. There is, however, no psychological equivalent to the representation of conformity impact used in Zollman (2010a) and Weatherall and O'Connor (2018). These models represent conformity pressure with regards to a given option in terms of the number of individuals endorsing that option. As the psychological literature has shown, however, conformity pressure is not a linear function of the majority size. The impact of informational and normative group influences seems to plateau when the majority size is higher than a certain threshold (Bond 2005).

[10] Notice that whether a given value of $\kappa$ is strong enough to lead to conformist behaviors depends not only on the majority strength but also on the difference in objective payoffs; intuitively, the potential pain of nonconformist behavior is not worth the gain, if the agent can only expect negligible benefits. In this simulation, the objective payoff was 0.5 for option A and 0.51 for option B.

https://doi.org/10.1017/psa.2021.25 Published online by Cambridge University Press

known to have greater impact when it comes to normative social influences deriving from group belonging concerns (Levine et al. 2002; Turner 1982). Likewise, individuals tend to direct their normative expectations to other in-groups (Antonio et al. 2004; Phillips 2003). Regarding the second point, there is empirical evidence to suggest that the presence of out-group individuals has a capacity to reduce conformity that is not tied to what views they express (Gaither et al. 2018). As Phillips et al. put it, "the *mere presence* of social diversity in task groups, even when out-group members are not bringing minority viewpoints to the table, can fundamentally change the behavior of the social majority to enhance group performance" (Phillips, Liljenquist, and Neale 2009). Reduction of pressure to conform to one's identity group is one mechanism through which this enhanced performance can occur, for example, because people tend to experience greater irritation when dissenting views are expressed by in-group rather than out-group individuals (Phillips and Loyd 2006). Thus, in a socially diverse group, a person might expect to provoke less annoyance in others when expressing dissenting views and therefore be less inhibited in doing so.

To reflect the research results sketched above, we restrict $\mathcal{N}_i^A$ to *in-group* individuals in agent *i*'s community, while still interpreting $\mathcal{N}_i$ as the total size of the community, which includes all neighbours regardless of group affiliations;[11] individuals may thus encounter the greatest social pressure in completely homogeneous groups. As the group gets diversified, the maximum possible pressure is reduced to the ratio of in-group community members to total community, reflecting the idea that the mere presence of out-group individuals makes people more willing to express dissenting views. Hence, in our modification of the "Other-Total Ratio," only the expressed opinions of in-group members generate normative conformity pressures, but those pressures can be mitigated by the presence of out-group individuals. Note that our measure of conformity pressure does not attempt to capture all of the effects found in the Phillips et al. experiment (Phillips, Liljenquist, and Neale 2009) discussed in section 2. For example, we do not attempt to model the effect whereby unexpectedly taking sides with an outsider against one's own group causes anxiety about damaging in-group relationships and thereby prompts closer attention to relevant arguments and evidence. Such omissions make our measure *conservative* in the sense of being likely to underestimate the positive effects of demographic diversity associated with counteracting normative conformity.

#### *5.1.2. Dependent measures*

**Compliance, acceptance, and polarization.** The dissociation between *private acceptance* and *public compliance* introduces four possible end results (cf. Weatherall and O'Connor 2018): *correct consensus* indicates that every agent not only publicly pursues the correct option, but also truly believes in its superiority. *Correct but with disagreement* refers to the situation where the collective publicly converges on the correct option, but some agents do so because of social pressure and against their innermost beliefs. Similar definitions apply to *incorrect consensus* and *incorrect but with disagreement*. In addition, our new setup turns *polarization* into a possibility.

[11] Of course, the influence of others need not be all or nothing. For example, we may allow out-groups to exert some influence, just less than in-groups. Similarly, we can modify how individuals react to out-group members; for examples, when one group is more socially powerful.

**Probability of success.** Given the dissociation between private beliefs and public conduct, here we examined how the probability of successful *conduct*, consisting of "correct all" and "correct but" portions of results, varied according to the level of conformist tendency in groups of different composition.[12]

## 5.2. Results

### *5.2.1. Conformity and group composition*

Overall and regardless of group composition, performance was negatively affected by conformism (see Figure 4a).[13] At low levels of $\kappa$, the presence of out-group agents resulted in improved performance. With increased conformity, there was a qualitative shift in this trend, and homogenous groups outperformed heterogenous collectives.

## 5.3. Discussion

Previous simulation studies have found that increased conformity negatively impacts group performance (Weatherall and O'Connor 2018). Simulation 2 shows that the same pattern holds, when agents in the network identify with different groups. Consider, then, the hypothesis that demographic diversity can counteract normative conformity and thereby lessen its negative effects. Our results support this hypothesis, albeit only for low values of $\kappa$. This can be seen in Figure 4a, wherein the homogeneous group (i.e., 0–10) performs worse than demographically mixed groups when $\kappa$ is small but greater than zero.[14] The reasons for the diversity-induced improvements are twofold. First, by reducing the conformity pressure on in-group individuals, the introduction of out-group members promotes *contestation*. As shown in Figure 4b, for a significant portion of runs where homogeneous groups converged on the wrong option, the group included some agents that privately disagreed with the public consensus ("incorrect but" in the graph). Because of conformity pressures, however, these agents did not act in accordance with their personal beliefs, and so their dissenting opinions remained private. With smaller values of $\kappa$, the presence of out-group individuals reduced the conformity pressure on dissenting agents, enabling them to pursue the option they believed best.[15] These publicly observable differences

[12] Insofar as normative influences are driven by observing *in-group behavior*, in this set of simulations there is no conformity pressure on agents in the first round of experimentation. Normative influences impact choice behavior only after agents observe in-group neighbours' conduct. We keep the number of experiments and trials the same. Combined with a difference of 0.1 between the payoff of two options (see Footnote 9), this reflects the setting of empirical studies of normative influence where the influence of conformity pressure on conduct tends to be evaluated in the context of easy tasks.

[13] We do not present the result from groups with 2–8 and 4–6 composition. This is because by making equal votes for the two options possible (e.g., 2–2 and 3–3 in the 4–6 groups) the composition of such groups allows for another way of reducing conformity pressure that does not work via the diversity pathway. While the same could be said of the homogeneous group, we simply accept that case as our baseline. Our results might thus be more conservative in the sense of underestimating the detrimental impact of conformity in homogeneous groups.

[14] Note that the presence of just one out-group individual is sufficient to achieve this effect, a result that corresponds to findings in some human experiments (Phillips, Liljenquist, and Neale 2009).

[15] Though sometimes not reflected in the visualization, the "correct but" and "incorrect but" cases do not completely disappear after the introduction of out-groups.

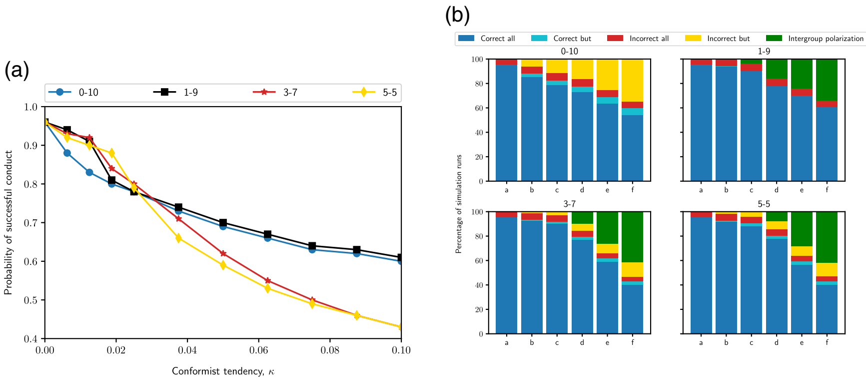


**Figure 4.** Influence of conformity on successful conduct in different complete networks of 10 agents. (a) Performance varies in networks with different proportions of the two groups. (b) A breakdown of the simulation outcomes in terms of the interaction between public consensus, private acceptance, and polarization. Each bar represents a particular value of $\kappa$: a: $\kappa = 0$, b: $\kappa = 0.00625$, c: $\kappa = 0.0125$, d: $\kappa = 0.025$, e: $\kappa = 0.05$, f: $\kappa = 0.1$.

in the option pursued in turn shape the normative influences on neighbours in the subsequent round. As a result, when the group reaches a consensus, it is one that likely reflects agents' private opinions. As $\kappa$ grows, however, agents are less likely to deviate from the majority position. Indeed, any benefit from the presence of out-groups is overwhelmed by between-group polarization.

Second, the introduction of out-groups also promotes exploration. The agents in our simulations are greedy, always following the option with the higher estimated value. Exploration is possible in this setup thanks to observing the result of others. In effect, each agent gets to vicariously explore because the community *allows* the agents to be greedy. By penalizing agents for pursuing less popular options that they nonetheless consider superior, conformity also prevents this mechanism of vicarious exploration. The opportunity to explore via others is restored to some extent by the addition of out-group agents. In this case, even agents whose beliefs were in alignment with the group behavior (i.e., that were not dissenting) may change their belief in light of the new evidence regarding less explored options.

## 6. General discussion

Can diversity benefit group performance in ways that are not mediated by task-relevant attributes of individual members? Going against the commonly presumed perspective, the results of our simulations demonstrate that, under certain conditions, it can. In particular, our findings support the claim that diversity can benefit group performance by counteracting informational and normative group influences that may negatively impact the epistemic endeavours of homogeneous groups (Levine et al. 2014; Phillips 2017). Having discussed the findings separately in previous sections (4.3 and 5.3), here we take a closer look at general limitations of our model and consider broader implications of our approach for simulation-based approaches to studying diversity.

https://doi.org/10.1017/psa.2021.25 Published online by Cambridge University Press

### 6.1. Limitations of the model

Our models are limited in a number of ways. First, they abstract away from and make a number of simplifying assumptions about various factors that can modulate informational and normative group influences in important ways. For example, the agents are assumed to be indistinguishable in terms of their group-based trust and conformist tendencies. This is, of course, a simplifying idealization. In general, it matters who says what to whom. Yet, the models abstract away from considerations such as status, centrality, and (perceived) expertise, even though these factors are important modifiers of informational and normative group influences (Gilovich et al. 2018; Raven 2008).

Even when certain variables are explicitly represented within the model, the mechanisms relating them are highly simplified or even ignored. For instance, we are assuming that there is no relation between group-based trust and the composition of the collective. In many real-world contexts, however, this is clearly an implausible assumption because high levels of group-based distrust may cause minority individuals to leave, making the collective more homogeneous. Similarly, high conformity pressure on the majority group may be caused by factors that also force minority individuals to "assimilate" to "fit in," thereby changing the overall dynamics. More generally, in natural settings, such variables are often related to one another by some mechanism (e.g., a common cause) such that variations in one are associated with variations in others. It is important to be clear about these limitations in interpreting our findings. For example, since mechanisms (e.g., asymmetric power relations) that may potentially connect mutual distrust, group composition, and individual performance are not represented in our model, the findings should not be taken as suggesting that a manager, who is interested in improving her team's performance, should instigate distrust among employees from different demographics.[16]

One could imagine different ways of relaxing some of these assumptions. We may, for instance, let agents differ with respect to status, perhaps as a function of their centrality in the graph, in a way that impacts their interactions with others. For our purposes, however, the assumptions help focus the experiment on the core hypothesis under investigation that, even when unrelated to cognitive diversity, demographic diversity can benefit group performance by counteracting detrimental pressures to conform. Moreover, there are interesting realistic scenarios where the most prominent factor driving behavior is indeed group affiliation; these include settings in which all group members interact with one another for a relatively short time period and where salient identities have limited connection to task-related expertise or ability. Importantly for our purposes, these are precisely the type of settings often used in human experiments investigating the influence of diversity on conformity and group performance (Gaither et al. 2018; Levine et al. 2014; Phillips, Liljenquist, and Neale 2009). Our simulation experiments, then, find that demographic diversity can improve group performance by reducing conformity under circumstances similar to those of human experiments yielding the same result. What is more, as discussed

[16] This is true even when the manager somehow knows the context-specific group optimal level of trust (given the network structure, task difficulty, . . .), which is often infeasible in practice (see Footnote 6). For a discussion of some of the challenges in leveraging these benefits of demographic diversity in practice, see Muldoon (2018).

https://doi.org/10.1017/psa.2021.25 Published online by Cambridge University Press

below, the simulations are suggestive of further background conditions that might also moderate the influence of diversity.

### 6.2. Demographic diversity and the uptake of simulation studies

The work presented here is suggestive for simulation-based studies of diversity in two ways. First, in relation to the point about the expressive limits of a given model, the findings highlight the need for caution in the interpretation, presentation, and uptake of simulation results. This point can be best illustrated with respect to Hong and Page's agent-based model developed to support what they call the "diversity trumps ability" theorem (2004). In this model, agents face the task of finding the maximum of a function, which can be visualized as finding the peak of a circular landscape. Each agent is uniquely represented by a set of heuristics that enables the agent to check different positions ahead of its current location on the landscape. Individual problem solving consists of the ordered application of heuristics. The agent halts when it can find no further improvements. Each agent's ability or performance is measured in terms of the expected value of its end points, given all possible starting positions. In the serial version of the task, *collective* problem solving is represented in terms of agents interacting sequentially: The first agent begins at a randomly assigned position on the circle and applies its heuristics until it can find no further improvement. The next agent begins where the previous agent left off and applies its own heuristics. This process cycles through the team members until no further improvements can be found. Within this set up, Hong and Page find that under certain conditions, "a team of randomly selected agents outperforms a team comprised of the best-performing agents". Assuming that randomly selected teams are likely to be more diverse in terms of their heuristics, *cognitive* diversity can thus be seen as improving group performance.

Importantly, Hong and Page are circumspect about the interpretation of their results as demonstrating the *beneficial effect of demographic diversity* on group performance. This caution is completely reasonable. Hong and Page's model answers the counterfactual question, "what would happen if, all else equal, we made the teams cognitively more diverse?" But, insofar as demographic diversity is not at all represented in the model, it makes little sense to transport the model's answer to that question to the new question, "what would happen if we made the teams demographically more diverse?" This means, however, that such models cannot support the claim that demographic diversity *cannot* benefit group performance, independently of cognitive diversity. Unfortunately, this is precisely what Page (2017) appears to be asserting in the following passage.

> For cognitive diversity to produce a bonus, it must be germane to the task. That same logic applies to identity diversity. For women, by virtue of being women, to create immediate diversity bonuses, women's repertoires ... would have to produce more accurate predictions, more creative ideas, better solutions to problems, or more comprehensive evaluations of projects. (2017, 169).

This claim assumes that, when it comes to communication dynamics, demographic diversity can only lead to "more conflict, more problems with communication, and less mutual respect and trust among members" (Hong and Page 2004, 16385–16386). So, one

https://doi.org/10.1017/psa.2021.25 Published online by Cambridge University Press

can study demographic diversity's *positive* impacts by focusing on its correlation with cognitive diversity and abstracting away from its other consequences, much like the way one studies mechanics assuming frictionless planes or electrical networks assuming negligible wire resistance—factors that in any case only hinder velocity or current. The problem with this inference is that it is based on a model that is not rich enough for actually exploring the complex dynamics instigated by these intergroup frictions. For instead of simply hindering performance, these frictions could enhance performance by *productively* unsettling detrimental interaction dynamics.

The achievement of any task by a group requires eliciting, examining, and integrating the knowledge possessed by individual group members. These interactions are subject to a host of group influences that could perturb the collective's epistemic performance. As suggested by the empirical hypotheses discussed in section 2 and demonstrated by our simulations, it is precisely by explicitly incorporating these types of *epistemic interactions* into our model that we can get a fuller assessment of the dependencies between demographic diversity and group performance. In Hong and Page's model, on the other hand, group interaction merely involves one agent's resumption of the collective task after another agent has been incapable of furthering this task. Group influences such as conformity pressure are thus not represented in the model.

What makes the situation unfortunate is not so much the restricted understanding afforded by Hong and Page's model. That is a feature of most, if not all, simulation-based (and indeed experimental) approaches, including our own. The more serious concern is the frequent use of Hong and Page's results to identify the core rationale for making a "business case" for *demographic* diversity. Aside from the ethical question of whether demographic diversity should require a business case, it is crucial that such discussions consider the consequences of demographic diversity for group interaction. By explicitly incorporating these considerations, our results offer a further rationale for the epistemic benefits of demographic diversity: diversity can benefit group performance because it promotes mechanisms that support *critical* and *contestable* collective inquiry—desirable properties that can be lost even in cognitively diverse but demographically homogeneous groups. This rationale is arguably more fundamental than the one based on *cognitive* diversity because the quality of group performance ultimately depends on the interaction dynamics of information sharing and processing (cf. Anderson 2006; Medina 2013); differences that *could be* cognitively beneficial remain inert, if not actively pursued and voiced.

### 6.3. Empirically sensitive robustness in simulation studies

A second suggestion for simulation-based research on diversity has to do with the manner in which robustness is pursued. In practice, the process of modeling involves constructing rather idiosyncratic representations of the target system. Modelers abstract away from certain aspects of the target system and make a variety of simplifying assumptions about certain other aspects (Weisberg 2007; Woodward 2006). Perhaps inadvertently, they may even neglect important features of the system of interest. When a feature of the phenomenon is selected, one needs to make value judgments about its appropriate formalization. What is more, in studying the target thus represented, modelers investigate what is often only a small subset of the possible parameter space.

https://doi.org/10.1017/psa.2021.25 Published online by Cambridge University Press

Still, one hopes that these idiosyncrasies do not undermine the intended purpose of the model and that the patterns of interest generated by the model would arise even if the target was represented in some other, equally plausible way. Accordingly, one may wish to test the *robustness* of the generated patterns under changes to the model: would we observe the same kind of patterns, had we explored other regions of the parameter space or altered some choices of formalization or relaxed some of the idealizing assumptions? Understandably, therefore, many simulation-based works on diversity and epistemic division of labour are robustness studies (e.g., Grim et al. 2019; Rosenstock, Bruner, and O'Connor 2017; Thoma 2015). Such works delineate the range of parameter settings and assumptions wherein some modeling results hold and so help guard against illicit inferences. Moreover, they further our understanding by highlighting novel ways of representing the target.

Nonetheless, it is important to keep in mind that robustness should be sought as a means of testing whether the particular way in which the target is represented threatens the model's intended aim. Suppose, for example, that a model seeks to provide insight regarding a class of real-world systems. In that case, what matters is not that the findings are robust *per se*, but that they are robust *in an empirically sensitive way*: the model should yield the patterns of interest precisely under those conditions where one expects similar types of effects in the real-world target system. Given the context sensitivity of the vast majority of effects in complex systems, it would be highly surprising if a model of one such system generated patterns that were robust across the board, regardless of the particular context and parameter setting. As a concrete example, consider Zollman's (2010b) model. The model suggests that, under certain conditions, sparser networks are conducive to better epistemic performance. Insofar as sparsity of connections induces a diversity of information, we may interpret the result as suggesting that under the conditions identified cognitive diversity is beneficial to group performance. Nonetheless, as Rosenstock et al. (2017) rightly note, Zollman's findings hold only in limited situations where the task facing the collective is difficult[17]. Sparsity, and so diversity, is no longer beneficial once we explore the regions of the parameter space corresponding to easier tasks.

From a conceptual point of view, however, this lack of robustness is not surprising. Presumably, what makes the diversity of cognitive repertoires beneficial is its contribution to effective exploration of novel alternatives in uncertain and complex task environments. When the answer is obvious (e.g., due to problem simplicity or learned expertise), no exploration is needed and so one should expect the epistemic benefits of diversity to be minimal. Crucially, this is precisely what one finds in empirical research on diversity. In their meta-analysis of empirical literature, for instance, van Dijk et al. (2012) find that the benefits of cognitive diversity are moderated by task complexity and are most pronounced in tasks requiring innovation. Given these conceptual and empirical considerations, we should in fact be suspicious of the empirical plausibility of models according to which cognitive diversity remains beneficial even in easy and routine tasks.

Seeking empirically sensitive robustness results in a natural complementarity between simulations and human experiments. In the context of the current work,

[17] The problem is difficult when the difference in the objective payoffs of different options is small, the amount of information sampled in each experiment is relatively small, or when the population size is small.

the simulations identify certain effects that one would expect given the empirical literature. In our first simulation, for instance, the benefits of demographic diversity are most pronounced when the two groups are evenly balanced, a finding that is in line with empirical results (Bear and Woolley 2011; Joshi and Roh 2009). However, this correspondence is not exact, as the empirical research just cited tends to find *negative* effects when, for instance, a small proportion of women are introduced into a predominantly male field. This may reflect the fact noted above that our model does not represent all mechanisms whereby increased distrust may adversely impact group performance. Moreover, we also identify certain effect modifiers that, to our knowledge, have not been investigated in empirical settings. For example, in our first simulation experiment, demographic diversity made the group more likely to converge on the correct result only in the complete network—a situation approximated in empirical studies when every individual is in contact with every other, either in face-to-face meetings or online. This result suggests that, at least insofar as the referent informational paths are concerned, demographic diversity may be less beneficial in situations where all do not communicate with all. And our second experiment found that demographic diversity lessened adverse effects of normative conformity only when the preference for conformity was relatively slight, suggesting that demographic diversity is more likely to lead to polarization when there are strong within-group pressures to conform. Thus, our simulations suggest hypotheses that could be tested by experiments involving human subjects and online communities.

There have been recent calls to bring agent-based models of social epistemology in closer contact with empirical evidence (Frey and Šešelja 2019, 2018; Harnagel 2019; Martini and Pinto 2017; Steel et al. 2019; Thicke 2020). For example, Thicke (2020) proposes that models be evaluated on the basis of their representational and predictive accuracy: assumptions in models should be plausible idealizations of the phenomena, and predictions derived from the model should match observations. The approach followed in this paper can be interpreted as attempting to attain these criteria but with the twist that the phenomena are primarily derived from experiments involving human subjects. One advantage of this approach is that human experiments are simplified social interactions whose rules and results are better known and are thus more amenable to modeling than messy real-world social life. Moreover, our approach also takes advantage of the complementary strengths and weaknesses of simulation and human experiments. In this way, it suggests a potential aim for simulation experiments that falls between providing "how possibly explanations" and predicting real-life social dynamics in their full complexity. This alternative aim consists of establishing, in conjunction with research involving human subjects, the existence of an experimental phenomenon; for instance, that demographic diversity can improve group performance by counteracting referent informational conformity. Since both simulation and human experiments are a step removed from the social reality that is ultimately of interest, a gap would still remain between an experimental phenomenon and policy recommendations. Our approach, therefore, could be seen as an example of what Harnagel calls "mid-level modeling," wherein empirically calibrated models linked to controlled experiments mediate between theory and real-life phenomena (Harnagel 2019).

Philosophers have developed valuable simulation models for investigating group behavior in a variety of epistemic tasks (see Šešelja, Straußer, and Borg 2020). These

https://doi.org/10.1017/psa.2021.25 Published online by Cambridge University Press

models can be usefully modified to address issues of societal concern regarding the impact of demographic diversity. As diversity researchers investigate the impact of diversity in broader settings and in online communities, we believe that such a philosophical engagement will be mutually beneficial to both empirical and simulation studies of the topic. The work presented here takes a step toward facilitating this complementary and multidisciplinary orientation.

**Acknowledgments.** We would like to thank Kevin Zollman, Sarah Gaither, and three anonymous referees for helpful comments and feedback. This work is based on research supported by funding from a Graduate Fellowship in Applied Ethics from The W. Maurice Young Centre for Applied Ethics, University of British Columbia.

## References

Abrams, Dominic, Margaret Wetherell, Sandra Cochrane, Michael A. Hogg, and John C. Turner. 1990. "Knowing What to Think by Knowing Who You Are: Self-categorization and the Nature of Norm Formation, Conformity and Group Polarization." *British Journal of Social Psychology* 29 (2):97–119.

Anderson, Elizabeth. 2006. "The Epistemology of Democracy." *Episteme* 3 (1–2):8–22.

Angere, Staffan, and Erik J. Olsson. 2017. "Publish Late, Publish Rarely!: Network Density and Group Performance in Scientific Communication." In *Scientific Collaboration and Collective Knowledge: New Essays*, edited by Thomas Boyer-Kassem, Conor Mayo-Wilson, and Michael Weisberg, 34–62. Oxford: Oxford University Press.

Antonio, Anthony Lising, Mitchell J. Chang, Kenji Hakuta, David A. Kenny, Shana Levin, and Jeffrey F. Milem. 2004. "Effects of Racial Diversity on Complex Thinking in College Students." *Psychological Science* 15 (8):507–10.

Asch, Solomon E. 1956. "Studies of Independence and Conformity: I. A Minority of One Against a Unanimous Majority." *Psychological Monographs: General and Applied* 70 (9):1–70.

Ashleigh, Melanie J., and Joe Nandhakumar. 2007. "Trust and Technologies: Implications for Organizational Work Practices." *Decision Support Systems* 43 (2):607–17.

Bear, Julia B., and Anita Williams Woolley. 2011. "The Role of Gender in Team Collaboration and Performance." *Interdisciplinary Science Reviews* 36 (2):146–53.

Berry, Donald A., and Bert Fristedt. 1985. *Bandit Problems: Sequential Allocation of Experiments*, 275. Springer Netherlands.

Bicchieri, Cristina. 2017. *Norms in the Wild*. Oxford: Oxford University Press.

Bond, Rod. 2005. "Group Size and Conformity." *Group Processes & Intergroup Relations* 8 (4):331–54.

Cialdini, Robert B., and Noah J. Goldstein. 2004. "Social Influence: Compliance and Conformity." *Annual Review of Psychology* 55 (1):591–621.

Crano, William D. 1970. "Effects of Sex, Response Order, and Expertise in Conformity: A Dispositional Approach." *Sociometry* 33 (3):239.

Deutsch, Morton, and Harold B. Gerard. 1955. "A Study of Normative and Informational Social Influences upon Individual Judgment." *The Journal of Abnormal and Social Psychology* 51 (3):629–36.

Duhem, Pierre. 1954. *The Aim and Structure of Physical Theory*. Princeton: Princeton University Press.

Frey, Daniel, and Dunja Šešelja. 2019. "Robustness and Idealizations in Agent-based Models of Scientific Interaction." *British Journal for the Philosophy of Science* 71 (4):1411–37.

Frey, Daniel, and Dunja Šešelja. 2018. "What Is the Epistemic Function of Highly Idealized Agent-based Models of Scientific Inquiry?" *Philosophy of the Social Sciences* 48 (4):407–33.

Gaither, Sarah E., Evan P. Apfelbaum, Hannah J. Birnbaum, Laura G. Babbitt, and Samuel R. Sommers. 2018. "Mere Membership in Racially Diverse Groups Reduces Conformity." *Social Psychological and Personality Science* 9 (4):402–10.

Gilovich, Thomas, Dacher Keltner, Serena Chen, and Richard E. Nisbett. 2018. *Social Psychology*. 5th Ed. 561. New York: W. W. Norton & Company.

Grim, Patrick, Daniel J. Singer, Aaron Bramson, Bennett Holman, Sean McGeehan, and William J. Berger. 2019. "Diversity, Ability, and Expertise in Epistemic Communities." *Philosophy of Science* 86 (1): 98–123.

https://doi.org/10.1017/psa.2021.25 Published online by Cambridge University Press

Harnagel, Audrey. 2019. “A Mid-level Approach to Modeling Scientific Communities.” *Studies in History and Philosophy of Science* 76:39–49.

Hogg, Michael A., and Dominic Abrams. 1988. *Social Identifications: A Social Psychology of Intergroup Relations and Group Processes*, 268. Oxfordshire, UK: Routledge.

Hong, Lu, and Scott E. Page. 2004. “Groups of Diverse Problem Solvers Can Outperform Groups of High-ability Problem Solvers.” *Proceedings of the National Academy of Sciences* 101 (46):16385–89.

Joshi, Aparna, and Hyuntak Roh. 2009. “The Role of Context in Work Team Diversity Research: A Meta-analytic Review.” *Academy of Management Journal* 52 (3):599–627.

Kelley, Harold H. 1952. “Two Functions of Reference Groups.” In *Readings in Social Psychology*, edited by Guy E. Swanson, Theodore M. Newcomb, and Eugene L. Hartley, 410–14. New York: Henry Holt/Co.

Kelman, Herbert C. 1958. “Compliance, Identification, and Internalization: Three Processes of Attitude Change.” *Journal of Conflict Resolution* 2 (1):51–60.

Langfred, Claus W. 2004. “Too Much of a Good Thing? Negative Effects of High Trust and Individual Autonomy in Self-Managing Teams.” *Academy of Management Journal* 47 (3):385–99.

Levine, Mark, Clare Cassidy, Gemma Brazier, and Stephen Reicher. 2002. “Self-Categorization and Bystander Non-intervention: Two Experimental Studies.” *Journal of Applied Social Psychology* 32 (7):1452–63.

Levine, Sheen S., Evan P. Apfelbaum, Mark Bernard, Valerie L. Bartelt, Edward J. Zajac, and David Stark. 2014. “Ethnic Diversity Deflates Price Bubbles.” *Proceedings of the National Academy of Sciences* 111 (52):18524–29.

Martini, Carlo, and Manuela Fernández Pinto. 2017. “Modeling the Social Organization of Science: Chasing Complexity through Simulations.” *European Journal for Philosophy of Science* 7 (2):221–38.

Medina, José. 2013. *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and the Social Imagination*. Oxford: Oxford University Press.

Mohseni, Aydin, and Cole Randall Williams. 2019. “Truth and Conformity on Networks.” *Erkenntnis*, 1–22.

Muldoon, Ryan. 2018. “The Paradox of Diversity.” *Georgetown Journal of Law and Public Policy* 16:807.

Mullen, Brian. 1983. “Operationalizing the Effect of the Group on the Individual: A Self-attention Perspective.” *Journal of Experimental Social Psychology* 19 (4):295–322.

Page, Scott E. 2017. *The Diversity Bonus*. Princeton: Princeton University Press.

Phillips, Katherine W. 2003. “The Effects of Categorically based Expectations on Minority Influence: The Importance of Congruence.” *Personality and Social Psychology Bulletin* 29 (1):3–13.

Phillips, Katherine W. 2017. “What is the Real Value of Diversity in Organizations? Questioning Our Assumptions.” In *The Diversity Bonus: How Great Teams Pay off in the Knowledge Economy*, edited by Earl Lewis and Nancy Cantor, 223–45. New Jersey: Princeton University Press.

Phillips, Katherine W., and Evan P. Apfelbaum. 2012. “Delusions of Homogeneity? Reinterpreting the Effects of Group Diversity.” In *Looking Back, Moving Forward: A Review of Group and Team-Based Research: Volume 15*, edited by Margaret A. Neale and Elizabeth A. Mannix, 185–207. Bingley, UK: Emerald Publishing Limited.

Phillips, Katherine W., Katie A. Liljenquist, and Margaret A. Neale. 2009. “Is the Pain Worth the Gain? The Advantages and Liabilities of Agreeing with Socially Distinct Newcomers.” *Personality and Social Psychology Bulletin* 35 (3):336–50.

Phillips, Katherine W., and Denise Lewin Loyd. 2006. “When Surface and Deep-level Diversity Collide: The Effects on Dissenting Group Members.” *Organizational Behavior and Human Decision Processes* 99:143–60.

Putnam, Robert D. 2007. “E Pluribus Unum: Diversity and Community in the Twenty-first Century. The 2006 Johan Skytte Prize Lecture.” *Scandinavian Political Studies* 30 (2):137–74.

Raven, Bertram H. 2008. “The Bases of Power and the Power/Interaction Model of Interpersonal Influence.” *Analyses of Social Issues and Public Policy* 8 (1):1–22.

Reagans, Ray, and Ezra W. Zuckerman. 2001. “Networks, Diversity, and Productivity: The Social Capital of Corporate R&D Teams.” *Organization Science* 12 (4):502–17.

Rosenstock, Sarita, Justin Bruner, and Cailin O’Connor. 2017. “In Epistemic Networks, Is Less Really More?” *Philosophy of Science* 84 (2):234–52.

Sherif, Muzafer. 1936. *The Psychology of Social Norms*. New York: Harper.

Stasser, Garold, and James H. Davis. 1981. “Group Decision Making and Social Influence: A Social Interaction Sequence Model.” *Psychological Review* 88 (6):523–51.

Steel, Daniel, Sina Fazelpour, Bianca Crewe, and Kinley Gillette. 2019. "*Information elaboration and epistemic effects of diversity*." *Synthese*.

Sutton, Richard S., and Andrew G. Barto. 1998. *Reinforcement Learning: An Introduction*. Cambridge: MIT Press.

Thicke, Michael. 2020. "Evaluating Formal Models of Science." *Journal for General Philosophy of Science* 51:315–35.

Thoma, Johanna. 2015. "The Epistemic Division of Labor Revisited." *Philosophy of Science* 82 (3):454–72.

Turner, John C. 1991. *Social Influence*. Milton Keynes: Open University Press.

Turner, John C. 1982. "Towards a Cognitive Redefinition of the Social Group." In *Social Identity and Intergroup Relations*, edited by Henri Tajfel, 15–40. Cambridge: Cambridge University Press.

Turner, John C., Margaret S. Wetherell, and Michael A. Hogg. 1989. "Referent Informational Influence and Group Polarization." *British Journal of Social Psychology* 28 (2):135–47.

Van Dijk, Hans, Marloes L. Van Engen, and Daan Van Knippenberg. 2012. "Defying Conventional Wisdom: A Meta-analytical Examination of the Differences between Demographic and Job-related Diversity Relationships with Performance." *Organizational Behavior and Human Decision Processes* 119 (1):38–53.

Weatherall, James Owen, and Cailin O'Connor. 2018. "Do As I Say, Not As I Do, or, Conformity in Scientific Networks." Preprint, submitted in 2019. http://philsci-archive.pitt.edu/id/eprint/16035.

Weisberg, Michael. 2007. "Three Kinds of Idealization." *Journal of Philosophy* 104 (12):639–59.

Weisberg, Michael, and Ryan Muldoon. 2009. "Epistemic Landscapes and the Division of Cognitive Labor." *Philosophy of Science* 76 (2):225–52.

Woodward, James. 2006. "Some Varieties of Robustness." *Journal of Economic Methodology* 13 (2):219–40.

Zollman, Kevin James Spears. 2010a. "Social Structure and the Effects of Conformity." *Synthese* 172 (3):317–40.

Zollman, Kevin James Spears. 2007. "The Communication Structure of Epistemic Communities." *Philosophy of Science* 74 (5):574–87.

Zollman, Kevin James Spears. 2010b. "The Epistemic Benefit of Transient Diversity." *Erkenntnis* 72 (1): 17–35.

Šešelja, Dunja, Christian Straußer, and AnneMarie Borg. 2020. "Formal Models of Scientific Inquiry in a Social Context: An Introduction." *Journal for General Philosophy of Science* 51:211–17.


**Cite this article:** Fazelpour, Sina, and Daniel Steel. 2022. "Diversity, Trust, and Conformity: A Simulation Study." *Philosophy of Science* 89 (2):209–231. https://doi.org/10.1017/psa.2021.25