Dynamics and Diversity in Epistemic Communities

February 18, 2015

Abstract

Academics and other researchers regularly engage in strategic interactions—bargaining, cooperation, collaboration, etc. Given this strategic setting, we ask: how do the dynamics of social learning in epistemic communities influence outcomes of various actors? We focus, in particular, on minority groups in academia. As we show, evolutionary game theoretic models indicate that such actors may be disadvantaged through social learning. These dynamics, in turn, can impact the course of scientifc inquiry and theory change by preventing the diversification of epistemic communities.

1 Introduction

There are (at least) two reasons why philosophers might care about the diversity of epistemic communities.[1] The first concerns social justice. Women, people of color, and other underrepresented groups deserve access to the same opportunities for academic work as their peers. The second reason concerns the success of epistemic communities. It has been widely argued that diversity improves inquiry. If so, those who care about successful inquiry should care about the make-up of epistemic communities. In this paper we discuss social dynamical effects whereby underrepresented groups may be disadvantaged in academia, potentially impeding the diversification of epistemic communities. We believe these effects are highly relevant to discussions of diversity that focus on social justice issues. In this paper we will focus on their relevance to epistemic progress.[2]

It has been widely observed that many academic communities have failed to diversify with respect to gender and race. For example, women make up only 25% of STEM workers (Beede et al., 2011).[3] Across academia, members of racial minorities are often severely underrepresented. Philosophers and other researchers have successfully argued that a number of factors can help explain the continued uniformity of many fields. Implicit bias and stereotype threat in particular seem to play central roles (Saul, 2011).

[1] By epistemic communities we mean groups of knowledge makers like academics and industry researchers, though our discussion will generally focus on academia.

[2] See Bruner and O'Connor (2015) for similar work focusing on social justice issues.

[3] According to Norlock (2006) women also account for only about 20% of tenured philosophy professors.

As such, there are many ongoing attempts to improve diversity in epistemic communities by mitigating the effects of these phenomena.

As we will argue here, however, even in the absence of explicit bias, implicit bias, and stereotype threat, minority groups can end up disadvantaged in epistemic communities due to the dynamics of social interaction. In particular, we will use evolutionary game theoretic models to show that when majority and minority groups interact, minority types often learn to accommodate majority types, while the majority learns to take advantage of this accommodating behavior. As a result, minority groups end up disadvantaged in strategic scenarios. These outcomes are more likely, in our models, the smaller the minority group. They occur across a variety of strategic settings, including bargaining, cooperation, and collaborative research scenarios. They occur despite assumptions that majority and minority groups do not differ with respect to skill levels or competence of any sort. Furthermore, as we will argue, these disadvantaged outcomes for minority groups may help explain why some groups, including women and racial minorities, tend to cluster into subdisciplines in academia. One of the most striking things about the effects we describe is that this minority disadvantage occurs even in models of populations that can be thought of as well meaning—where actors are not initially biased and do not actively try to disadvantage minority groups. This is not to say that their eventual behavior is unobjectionable, but rather that behavior consistent with bias and discrimination arises as a result of reasonable learning by reasonable agents.

We will further argue that the effects we note may not be limited to minority groups in the traditional sense, but may also arise for what we call *epistemic* minorities—groups that hold minority opinions in an epistemic community, practice methods only used by a minority of practitioners, or focus on questions of interest to a small part of a discipline. In other words, the effects we will describe can potentially impede epistemic progress by 1) preventing the success of gender and racial minorities in epistemic communities 2) preventing the success of epistemic minorities and 3) leading minority groups of both sorts to cluster together, thus preventing discourse.

Our paper will proceed as follows. In section 2, we discuss the senses of diversity which will be adressed here, and discuss why diversity of these various kinds is important to epistemic progress. In section 3, we justify the use of evolutionary game theory to model epistemic communities, and describe this methodology. In section 4, we outline the games we use to represent bargaining, cooperation, and collaboration in academia. In section 5, we describe our results as well as relevant work from other authors. In section 6 we discuss the relevance of these results to epistemic minorities and the potential impacts of these effects on epistemic progress. In the conclusion, we briefly describe policy measures that may help prevent the social dynamical effects we describe. We end with an appeal for the future use of formal methods, like those applied here, to explore topics in feminist philosophy and related areas.

2 Diversity Matters

There are two broad senses of diversity that are relevant to this paper. The first is diversity with respect to gender, race, and other unalterable personal factors such as sexual orientation or cultural background. We will call this *personal* diversity.[4] The second sense of diversity can be referred to as *epistemic* diversity.[5] Epistemic diversity tracks differences in beliefs and behavior that lead epistemic agents to engage in inquiry in different ways. For example, diversity of intellectual beliefs and assumptions, research methodologies, and interests in research problems all fall under this heading. In the rest of this section we will briefly discuss 1) why epistemic diversity is important to epistemic progress and 2) how personal diversity maps onto epistemic diversity.

Philosophers of science and feminist epistemologists have argued convincingly that inquiry is inherently value laden (Longino, 1990, 2002; Potter, 1993; Harding, 1991, 1998; Wylie, 2007). Theories are underdetermined by evidence and so background assumptions and values will play a role in determining theory choice (Longino, 1990, 2002; Nelson, 1993). This means that two different researchers will potentially generate knowledge differently, and thus the make-up of an epistemic community may influence what theories that community generates and supports.

Many have argued that, in particular, epistemically diverse communities are more successful than uniform ones. There are a number of various epistemic traits for which diversity has been championed. Longino (1990, 2002) contends that research communities that hold uniform assumptions and prejudices will lack the critical perspective to challenge these viewpoints. Epistemic communities where agents hold diverse views, on the other hand, will be better positioned to rethink failing assumptions. Zollman (2010) uses decision and game theoretic models to argue that epistemic communities that hold a variety of beliefs about the world may be more likely to eventually converge on successful theories. Kuhn (1977) points out that disagreement among scientists is a crucial part of theory change. He argues that the necessary sorts of disagreements can only occur in groups with some diversity of choice criteria for theories. Kitcher (1990) holds that it is beneficial for a community of agents to work on diverse problems and to use diverse methodologies for similar reasons. Weisberg and Muldoon (2009) consider a model of a community in which a number of different types of epistemic agents explore theory space, and conclude that populations where agents employ mixtures of search strategies are, in some cases, more successful. To summarize these arguments, epistemic diversity may be beneficial with respect to (at least) assumptions and prejudices, theoretic beliefs, criteria for adopting new theories, scientific methods, and problem choice.[6] We do not provide an exhaustive list of the proposed benefits of epistemic diversity. Our intention is rather to give a sense of the breadth of such arguments.

Note that even if one believes that epistemic diversity is important to inquiry, fur-

[4]We focus on gender and racial diversity throughout the paper, though the work may potentially be extended to other sorts of personal diversity.

[5]Fehr (2011) makes a similar distinction between situational and epistemic diversity.

[6]Psychologists have also argued that diverse viewpoints promote problem solving and creativity in groups (Paulus and Nijstad, 2003).

ther arguments are needed to support the claim that personal diversity is beneficial to epistemic progress.[7] It has in fact been argued that gender and racial diversity do, in many cases, correlate with epistemic diversity. Researchers of different personal types, for example, will be interested in different problems and hold different assumptions and prejudices as a result of their varying experiences. A wide variety of arguments have been made along these lines and we will not outline them all here.[8] We will instead make our point by giving a few examples of episodes from science where personal diversity was clearly important.[9] Haraway (1989) outlines how the introduction of female scientists revolutionized the field of primatology, which had previously focused largely on the behavior of male primates. In this case, there is an obvious connection between the gender of the researchers and their choice of research problem. Female primatologists also introduced pioneering methodologies designed to prevent sexist assumptions from creeping into primate research. Longino (1990) describes the debate over 'man-the-hunter' and 'woman-the-gatherer' theories of the evolution of human cognition. These two theories each focus on one gender as being more important to early tool use in humans and thus to subsequent human evolution. In this case gender played a role in leading scientists to challenge a problematic dominant paradigm (albeit with their own value-laden theory).[10]

To sum up, personal diversity often results in epistemic diversity, which in turn is important to inquiry. We will now outline how social dynamical effects, even in the absence of bias, can result in systematic disadvantage to minority groups, thereby potentially decreasing the level of diversity in such communities.

3 Methodology

Evolutionary game theory is a methodology first developed by biologists to explore the evolution of strategic behavior in animals. This methodology has since been adopted by social scientists and philosophers to inform how strategic behavior in humans changes through learning and social learning. Since our aim is to investigate how communities of epistemic agents learn to interact strategically with each other, evolutionary game theory is an appropriate methodology to employ.

Before continuing, it will be useful to describe evolutionary game theoretic methods.[11]

[7] It might be the case that epistemic and personal diversity are not correlated. It might also be the case that upon recognizing the value of epistemic diversity, a group of uniform agents can choose to explore different theories, use different methodologies, etc. and so gain the benefits of epistemic diversity. Kitcher (1990) provides models where actors have the option of something like this.

[8] For more on this, see work in standpoint epistemology such as Harding (1991, 1998).

[9] Because the literature we are referencing here is largely from feminist epistemology, these examples focus on gender. Similar cases can be found for race and other personal factors.

[10] It is not lost on us that in these clear cut examples there are humanistic aspects of the research such that gender based stereotypes directly influence the researchers involved. It is harder to find clear cases where personal diversity is important in areas like physics or mathematics, though many have argued for such effects (Harding, 1991, 1998; Fehr, 2011).

[11] This is intended to be a cursory overview, and readers who wish to engage more thoroughly with the details of our results, but who are not familiar with the methods used, may wish to read Weibull (1997) or Gintis (2009).

A *game* in game theory is a model of a strategic interaction between agents. Games are usually defined by a set of *players*, or actors in the game, a set of *strategies*—available actions—for each player, and *payoffs* for each possible combination of strategies.[12]

Traditionally, game theorists have focused on rational decision making by actors, as defined by payoff maximization where payoff tracks actors' preferences. The evolutionary game theoretic approach, in contrast, attempts to understand behavior not by appeal to rationality, but by appeal to evolution, cultural evolution, or learning. This is done by employing what are called *dynamics* to games. Dynamics are rules that determine how actors or populations of actors playing a game will change based on past interactions. Throughout this paper, we will employ the *replicator dynamics* to model behavioral change in epistemic communities.[13] The basic idea behind the replicator dynamics is that in a population with many actors playing different strategies, the strategies that get higher payoffs will proliferate, while those with lower payoffs will decline in number. In the case of an epistemic community, it is natural to think of payoffs as corresponding to the amount of esteem or credit a particular researcher receives, the amount of pay he or she is offered, etc.[14] Individuals in an epistemic community are then assumed to imitate the successful actions that led to this high payoff, leading to the proliferation of such successful actions.

One concept that has been central to game theoretic analyses is that of a *Nash equilibrium*. A Nash equilibrium is a set of strategies where no player may deviate and improve her payoff. These strategy sets are thought of as stable and likely to arise in real world settings for this reason. They also arise in many evolutionary settings, making them relevant to our evolutionary analysis.

4 Epistemic Games and Type-Conditional Behavior

As discussed in the previous section, evolutionary game theoretic models include two things—games and dynamics. In this section we describe the games we use to model strategic interactions in epistemic communities. In the next section we add the dynamics and discuss results.

There are three sorts of strategic interaction that we focus on here—bargaining, cooperation, and collaboration. As we will argue, these three types of interaction are central to communities of researchers. The games we use to represent these scenarios are the Nash demand game, the stag hunt, and the collaboration game, respectively. In particular, we look at versions of these games where actors can be of different types (male/female, black/white, etc.) and can condition their behavior on the type of their

[12]Traditional games also define *information* available to the actors, but in evolutionary game theory, this aspect is downplayed.

[13]These dynamics were intended as a model of change via natural selection, but have subsequently been shown to effectively model both individual learning (Hopkins, 2002; Börgers and Sarin, 1997) and cultural evolution (Weibull, 1997).

[14]This is not an unusual assumption to make in formal models of the social structure of science. Kitcher (1990) and Strevens (2003) both make similar assumptions in a classical game-theoretic framework, while Bruner (2013) assumes something similar in a dynamic model involving cultural evolution.

interactive partners. We describe the three games in turn and discuss their relevance to epistemic communities. We then discuss how these games work when actors can condition on type.

In the Nash demand game, actors each demand a certain portion of a resource. If the demands of the actors are compatible in that they do not exceed the resource, each actor receives the portion he or she requested. If, on the other hand, the demands are greater than the total resource, bargaining fails and the two actors receive what is called the 'disagreement point' (often nothing). Figure 1 displays a *payoff table* of a simplified version of this game where actors can demand 4, 5, or 6 of a total resource of 10. We call these demands 'Low', 'Med', and 'High'.[15] A payoff table lists payoffs to actors for any combination of possible strategies in a game. Note that this game has three pure strategy Nash equilibria—High v Low, Med v Med, and Low v High.[16] In each of these the actors fully divide the resource. This means that if either actor demands more, the disagreement point is reached. If either demands less, she simply gets less.

		Player 2		
		Low	**Med**	**High**
Player 1	**Low**	4,4	4,5	4,6
	Med	5,4	5,5	0,0
	High	6,4	0,0	0,0

Figure 1: A payoff table for a Nash demand game. Rows represent strategy choices for player 1. Columns represent strategy choices for player 2. Entries list the payoff for player 1 followed by that for player 2.

Bargaining may not seem like a central part of research because there are only a few situations—salary and funding negotiations, for example—in which researchers explicitly bargain over resources. A closer look, however, reveals that research communities are full of cases where actors must bargain to divide labor. Consider, for example, academics who engage in any sort of joint project—conference planning, committee work, running a department, or collaborative research. In every example, by some process of implicit or explicit negotiation, these actors must decide who will take on what portion of the work done. Although the Nash demand game is formulated as a situation where two individuals divide a resource, it can just as easily be interpreted as two individuals dividing tasks.[17] As such, this game is an appropriate model both of explicit bargaining over resources such as salary, and of bargaining over workload in joint projects.

[15] Simplified games of this sort are standardly employed in evolutionary analyses of bargaining (Skyrms, 1994, 1996; Skyrms and Zollman, 2010; Young, 1993; Binmore, 2008; Alexander and Skyrms, 1999; Alexander, 2000).

[16] Pure strategies are ones where actors always take the same action rather than randomly mixing over multiple actions. This game (and the following ones) also have mixed Nash equilibria—those where actors use strategies that probabilistically choose actions—but these will be less germane to our evolutionary analyses and so we do not discuss them here.

[17] Assume that the resource is either leisure time or time for individual research.

The stag hunt is a model of cooperation under risk. Suppose that two hunters may either hunt for stag or for hare. If either actor hunts hare, she get a small amount of meat. If the actors hunt stag together, they manage to bring home more meat. But stag is a riskier strategy because if either hunts stag alone, she is unable to catch anything. The payoff table for this game is shown in Figure 2. The two pure strategy Nash equilibria of this game are Stag v Stag and Hare v Hare.

		Player 2	
		Stag	**Hare**
Player 1	**Stag**	3,3	0,2
	Hare	2,0	2,2

Figure 2: A payoff table for a stag hunt. Rows represent strategy choices for player 1. Columns represent strategy choices for player 2. Entries list the payoff for player 1 followed by that for player 2.

Stag hunts occur whenever joint effort is mutually beneficial, but potentially risky. Academics who cooperate to plan conferences, co-organize committees, advise students, and, of course, collaborate on research projects all stand to benefit from these cooperative engagements, but also potentially put themselves at risk. If a co-organizer fails to appropriately carry out her role in such an arrangement, her cooperative partners may be disadvantaged because the endeavor fails (or is less successful than it might have been) despite their having put in a great deal of work. As such, the stag hunt is an appropriate model of many sorts of cooperation in academia.

We call our last model the collaboration game.[18] In this game, two actors first decide whether to enter a collaborative partnership by playing the stag hunt. Should the two actors choose to hunt stag, they then decide how to divide the resource obtained. This negotiation is modeled with the Nash demand game. Figure 3 shows the payoff table for the collaboration game. There are four strategies in this game—the uncooperative Hare strategy and three cooperative Stag strategies where actors demand Low, Med, or High of their partner. Hare always generates a dependable, low payoff. Stag leads to a greater joint payoff, though it is now risky for two reasons: 1) a partner could choose Hare or 2) bargaining could fail. Either of these outcomes leads to a payoff of zero. Even if bargaining does succeed, a stag hunter who ends up with the Low demand might be better off hunting hare.

This model captures a strategic scenario that many collaborating or cooperating academics encounter. Academics typically must first decide whether or not to collaborate with a potential co-author. Like a stag hunt, collaboration is beneficial because it comes with the potential for rewards. For example, Card and DellaVigna (2013) found that in top economics journals co-authored papers were cited significantly more than single

[18]This game is extensively analyzed by Wagner (2012). He notes that this game is equivalent to a Nash demand game with an outside option.

Player 2

Player 1		Stag - Low	Stag - Med	Stag - High	Hare
	Stag - Low	3,3	3,5	3,7	0,4
	Stag - Med	5,3	5,5	0,0	0,4
	Stag - High	7,3	0,0	0,0	0,4
	Hare	4,0	4,0	4,0	4,4

Figure 3: A payoff table for a collaboration game. Rows represent strategy choices for player 1. Columns represent strategy choices for player 2. Entries list the payoff for player 1 followed by that for player 2.

authored papers. Co-authored papers are more likely to be accepted to top journals in many fields (Laband, 1987; Gordon, 1980; Beaver and Rosen, 1979). And many authors have argued that collaboration improves overall academic productivity (Morrison et al., 2003; Landry et al., 1996; Lee and Bozeman, 2005). Collaboration is risky, however, because if a collaborative partner fails to fulfill her part of the collaboration properly, her co-authors may be disadvantaged.[19]

Once collaboration has begun, actors also must bargain either implicitly or explicitly to decide how the efforts and rewards of collaboration will be divided. Who will be first author on a paper? Who will be last author? Who will do what portion of the work? Who will present joint work at prestigious conferences? These negotiations may end at equitable divisions (Med v Med) as when, for example, one author puts in significant effort and acts as first author. They may, however, end at inequitable divisions (Low v High) as in cases where one actor does little work relative to author position.[20]

This concludes our description of the base games that we employ in this paper. We will now discuss how such games work when actors can condition on the type of their interactive partner. Typical game theoretic analyses consider interacting agents who behave the same way upon meeting any other agent. In real epistemic communities, however, empirical evidence suggests that actors often treat different types of people differently. For example, it has been argued that tenure and promotion decisions are made differently for women and men (Perna, 2001). Emails to professors asking for mentoring help tagged with male and/or white sounding names receive more and better responses than those with female and/or ethnic sounding names (Milkman et al., 2014). Black applicants are less likely to receive NIH funding than white applicants controlling for other factors (Ginther et al., 2011). Researchers, when assessing otherwise identical male and female academic job candidates, are more likely to believe males are more qualified, more likely to hire the male, and more likely to offer him a higher salary

[19] In cases where a collaborative partner engages in academic dishonesty, this disadvantage may be very serious.

[20] Recent work on ghost authorship implies that in some disciplines such outcomes are common (Bennett and Taylor, 2003).

(Moss-Racusin et al., 2012; Steinpreis et al., 1999).[21]

In order to capture this sort of behavior, we investigate games where actors are of two different types and condition their behavior on the type of their interactive partner. Importantly, Nash equilibria for the type-conditional versions of the games described above are significantly altered. Instead of strategies like 'Stag' or 'High', conditioning actors choose strategies like 'play Stag with my in-group, play Hare with my out-group'. For simplicity sake, we will refer to these strategies using a list where the first item on the list is the strategy played with one's in-group and the second with one's out-group as in: <Stag; Hare>.

With these new strategies come new equilibria where actors play one Nash equilibrium against their in-group members and another against out-group members. Nash equilibria, remember, are important here because actors playing them will not be incentivized to switch strategies. Consider an academic community where women and men regularly bargain and where women play the strategy <Med; High> and men play <Med; Low>. In this case, both men and women make fair bargaining demands against like types, but when men and women interact women demand greater resources and men concede. Despite the fact that, in a case like this, men might prefer to receive better bargaining outcomes, because women are all demanding High, men cannot switch strategies and improve their lot. Or suppose that white researchers play <Stag; Hare> and researchers of color play <Hare; Hare> in the stag hunt. In this case, white researchers cooperate amongst themselves, but not with researchers of color, and researchers of color always choose Hare. Again, researchers of color cannot make unilateral changes to reach a cooperative outcome with white researchers. In the type-conditional collaboration game, suppose in a population of male and female academics men play <Stag - Med; Stag - High> and women play <Hare; Stag - Low>. In this case, collaboration and fair division are the norm when two men interact. Individual work is pervasive among women. Collaboration occurs between men and women, but an unequal bargaining outcome has emerged.

These examples of new equilibria in these games are significant to the current discussion because they can be interpreted as involving discriminatory norms (Axtell et al., 2000; Bruner, 2014). At these equilibria, both populations treat out-group members differently than in-group members and in both cases, as a result, one group is disadvantaged. Notably, two types of discriminatory norms can occur in the collaboration game. Actors can refuse to cooperate with out-group members and disadvantage them in this way, and they can also divide cooperative products differently with out and in-group members. As we will show in the next section, these sorts of discriminatory norms can emerge endogenously in populations of interacting epistemic agents.

[21]Similar findings, some outside of academia, have been garnered for job candidates who are LGBTQ or racial minorities (Tilcsik, 2011; Bertrand and Mullainathan, 2003). Outside of academia, assertive bargaining behavior by women has been found to meet with more resistance than assertive bargaining behavior by men (Bowles et al., 2007; Tinsley et al., 2009). And experimental subjects are more likely to hire men than women to perform arithmetic tasks regardless of performance (Reuben et al., 2014).

5 Social Dynamics and Minority Disadvantage

We have now described the games we use to model strategic interaction in epistemic communities. In this section we look at what happens when groups of researchers learn to interact in these strategic scenarios. Unless otherwise specified, the dynamics employed here will be the replicator dynamics which, as discussed in section 3, can be used to model social and individual learning.[22] The best way to interpret these dynamics is as an approximate representation of academics who copy the behaviors of successful colleagues, and repeat successful behaviors of their own. Compelling research indicates that such social learning is ubiquitous in humans and that in particular humans tend to imitate successful and prominent peers (Richerson and Boyd, 2008; Henrich and Henrich, 2007).[23] There is also strong research support for the importance of individual learning in strategic situations. We start by describing previous results on populations playing the Nash demand game and the stag hunt. We then present new results on groups playing the collaboration game.

5.1 Previous Results: Bargaining and Cooperation

Imagine a group of researchers that includes two different types. This group of researchers must bargain regularly, and, in doing so, learn to adopt bargaining practices that work well for themselves and their peers, as just described. Axtell et al. (2000) find that discriminatory norms commonly evolve in a Nash demand game model of such a situation.[24] Perhaps surprisingly, without types, the fair division, Med v Med, is seen most often in such models (Skyrms, 1996, 1994; Ellingsen, 1997; Young, 1993). Without type-conditioning, demanding High is risky because one is going to meet others with high demands, and as a result sometimes reach the disagreement point. Those who demand Low lose out on the resource when they meet others demanding Med or Low. In a type-conditional model, though, if one demands High against a different type that always demands Low, one is guaranteed to always receive High—the most preferable bargaining outcome. The division into types allows for an advantageous (or disadvantageous)

[22] We use the discrete time replicator dynamics. These dynamics give updates of a population distribution at discrete steps or generations. In particular, we use a version of the two population replicator dynamics where one population may be smaller than the other, and where all actors interact with both populations. They are formulated as follows. Let x and y represent the two populations so that $x+y=1$ and $x \leq y$. Strategies for population x update according to the following dynamics, $x'_i = x_i(\frac{f_i(x)}{\sum_{j=1}^{n} f_j(x)x_j})$ where x_i is the proportion of the x population playing strategy i, $f_i(x)$ is the fitness of type i in x given the population states of x and y, and $\sum_{j=1}^{n} f_j(x)x_j$ is the average population fitness for x given the states of x and y. Strategies for population y update according to the same dynamics.

[23] Because we divide our populations into two types, where actors treat different types differently, we also assume that they learn differently from the two types. Actors in our models imitate peers of their own type, rather than the other. When it comes to choosing models for social learning humans clearly are sensitive to social divisions like gender and race, though we think this assumption, like many aspects of the models we discuss, is an idealization from a more complex reality.

[24] Axtell et al. (2000) do not employ the replicator dynamics, but rather use a different sort of dynamics where actors copy group members. Bruner (2014) similarly finds that, for the replicator dynamics, such norms arise regularly in the Nash demand game. Skyrms and Zollman (2010) find similar results.

strategy that is unavailable in a uniform population.[25]

Axtell et al. looks only at situations with equally sized groups. Notably, when Bruner (2014) extends these results to cases where one population is in a minority, he finds that the minority is at a significant disadvantage as far as bargaining outcomes go. The smaller the minority population, the more likely it is to end up at the unfavorable bargaining outcome—where the minority demands Low against the majority who demand High—and the less likely it is that the majority population ends up at this outcome. Bruner likewise finds that in populations playing the stag hunt, as minority population size decreases, it becomes increasingly likely that populations end up in a scenario where the majority population cooperates with itself, but refuses to cooperate with the minority population.

Why the disadvantage for minority groups? This effect occurs because there is a difference, in these models, in how strategically important each group is to the other. Majority members meet minority types relatively rarely, whereas minority types are constantly interacting with majority members. It is often the case that for the minority population, the least risky and therefore best strategy is to demand Low (or, in the stag hunt, Hare) of the majority. In simulations, the minority group thus quickly learns to adapt to the majority by playing Low or Hare. Majority strategies in response to the minority shift more slowly, and as a result the majority does best to demand High against the minority population, or to play Hare.[26]

It should be noted that this sort of effect does not always end in minority disadvantage—in the right sort of game it will be advantageous on average for minority members to quickly adapt to majorities by learning a strategy that benefits them (such as demanding High in the Nash demand game) (Bruner, 2014). We do not focus on these situations since our aim is to investigate cases where minority groups are potentially disadvantaged by social dynamics in the absence of other factors. Furthermore, as we will discuss in section 5.2.4, when one looks at realistic populations minority disadvantage is seen even in these games.

5.2 New Results: Collaboration

We now consider what happens when groups of academics learn to collaborate with each other. At the end of this section we will discuss the relevance of all the results discussed here to epistemic communities.

5.2.1 Minority Groups and Collaboration

Imagine a population of researchers of two types playing the collaboration game. We begin by considering what happens when researchers learn to collaborate with those of the other type. When the two types each make up 50% of the population, they are most

[25] D'Arms et al. (1998) similarly point out that allowing anti-correlated interaction between bargainers who make high and low demands allows for the evolution of these two types.

[26] For a biological corollary, see work by **?** on the Red King effect.

likely to reach fair bargaining outcomes.[27] When one type is in the minority, however, it becomes increasingly likely that the groups reach unfair bargaining outcomes, and increasingly likely that the minority is disadvantaged. By this we mean that minority types end up demanding Low against majority types who demand High. Figure 4 shows results for 10,000 runs of replicator dynamics simulations of this scenario. The payoff associated with Hare (which we will call G from now on) was held fixed at 2.5. Low, Med, and High were set at 4, 5 and 6, respectively. As can be seen in this figure, when the minority is very small, they end up disadvantaged in more than half of simulations.

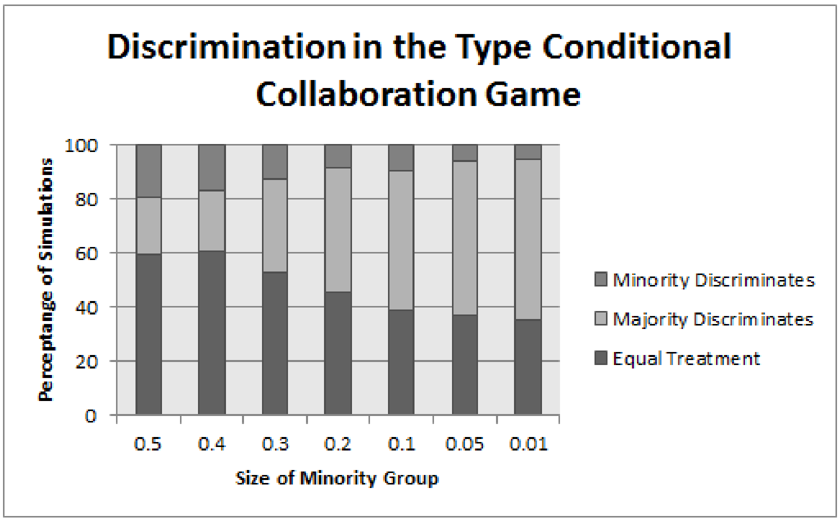

Figure 4: Simulations of populations playing the type-conditional collaboration game. Each bar summarizes outcomes for 10,000 simulations. The x-axis represents different sizes of minority populations. The y-axis represents the proportions of simulations that arrived at different outcomes.

As described in section 5.1, these effects are a result of different learning environments for minority and majority types. Minority types learn quickly that when collaborating with majority types, they are safest demanding Low. Majority types then learn to demand High. It should be noted that these outcomes are consistent with certain empirical observations. Both West et al. (2013) and Sugimoto (2013), for example, find that women are less likely than men to hold coveted author positions when collaborating.

One thing that should be noted is that when a minority ends up at a disadvantage with respect to a majority type in these simulations this outcome can be particularly damaging because minority members tend to interact with majority members more often. Figure 5 shows the average payoff to minority and majority groups when they collaborate and are disadvantaged in the collaboration game. As is evident, the smaller the minority population the more it is disadvantaged by discrimination, and the less the majority

[27] And when they reach unfair outcomes, they are each equally likely to discriminate against the other.

population is.[28]

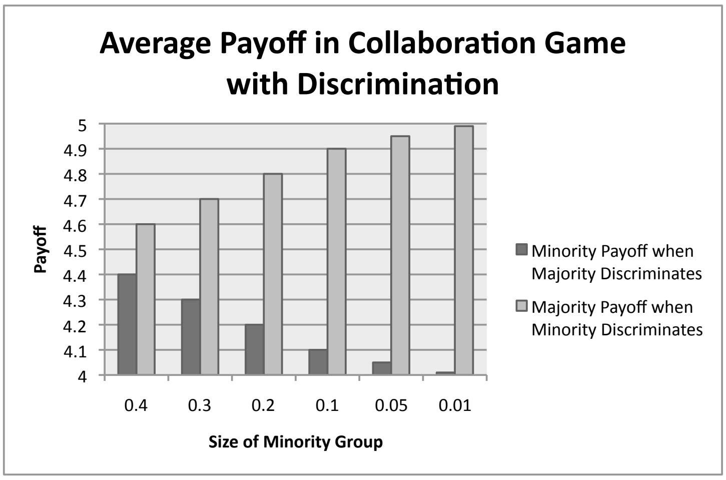

Figure 5: Average payoffs for minority and majority groups with discriminating norms in the type-conditional collaboration game. Note that the y-axis is scaled to highlight the phenomenon of interest.

5.2.2 When Collaboration is Necessary

In the social sciences and humanities, collaborative work can provide benefits and so may be desirable, but in most cases researchers in these areas can successfully perform individual work as well. In the lab sciences, however, group work is often essential to success. To expand our models to capture such situations, we vary the payoff of individual work, or hare hunting (G). Note that when G = 0, the collaboration game is essentially equivalent to the Nash demand game.[29] This represents a case where individual work is completely useless and actors must always collaborate and bargain. Slightly higher Gs represent cases where individual work may not be as beneficial as collaborating, but can still provide some benefit.

We find that as G decreases, collaboration increases. This is unsurprising as the payoff to individual work is poor when G is low. When G is low, however, it also becomes increasingly likely that bargaining favors members of the majority group. Figure 6 illustrates these results. The minority constitutes 5% of the total population in this figure, and the possible bargaining demands are 4, 5, and 6. As is evident, when G is low the majority discriminates against the minority in over 60% of simulations and the minority almost never discriminates against the majority.

[28]This finding is similar to something noted in recent blog posts. Computer scientist Karen Petrie pointed out that in a population where men and women make equally many sexist remarks, but where the ratio of women to men is $1 : n$, women will receive n^2 as many sexist remarks.

[29]Actors still have the option to hunt hare. But since this individual work pays nothing, in order to obtain any payoff they must hunt stag and bargain to divide it.

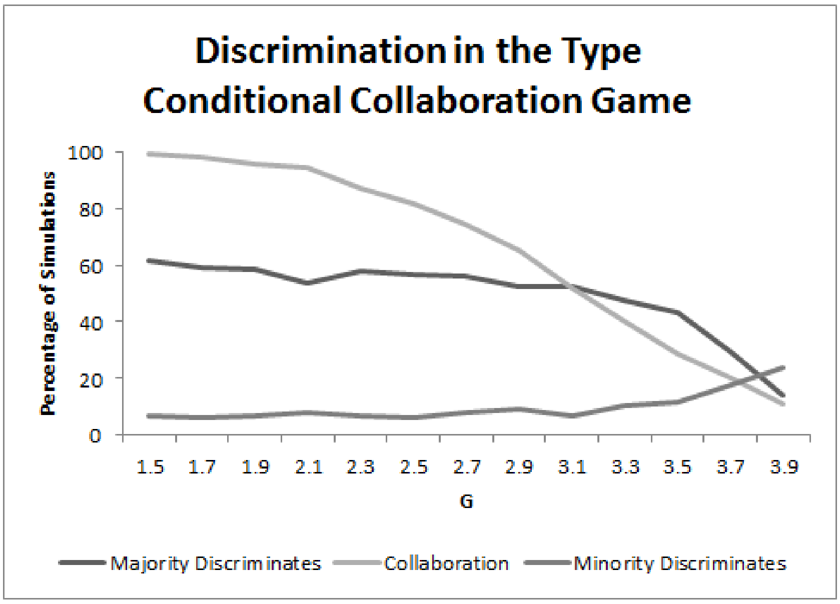

Figure 6: Simulations of populations playing the type-conditional collaboration game. The x-axis represents different values for G. The y-axis represents the proportions of simulations that arrived at inter-group collaboration with different bargaining outcomes.

These results occur because when G is low, minorities are unable to successfully opt out of collaborating. In other words, they learn to collaborate even when discriminatory norms governing collaboration arise. These results may indicate that minority groups in areas where collaboration is necessary are at particularly high risk of disadvantage. Note that as in the last results, when the minority group is disadvantaged here, it is especially damaging because of the prominence of majority types.

5.2.3 Discrimination and Deterrence

We now explore cases in which norms of division have already become cemented in a discipline. In other words, agents still must choose whether to collaborate or not, but if collaboration does occur, the individuals involved do not engage in a round of bargaining, but instead fall back on extant norms to determine how to divide labor and rewards. Our aim is now to determine whether an existing unfair division will disincentivize collaboration by minority members. Note that in cases like this, because actors do not need to bargain, the collaboration game is essentially equivalent to a stag hunt, but where the benefit for hunting stag is different for the two partners.

This further model is particularly important because it will not always be possible for collaborative partners to determine their payoffs for collaboration. Collaborating partners decide author order. Author order, however, does not totally determine how an academic community, hiring, or promotional bodies react to collaborative work. Consider, for example, a case where a hiring body assumes that the white author on a joint paper must have contributed more to the collaboration than a black partner, and gives credit accordingly. The models here capture both scenarios where groups bargain differently over collaboration due to existing societal norms, and situations where societal norms undervalue the contributions of underrepresented groups to collaborative work.

Figure 7 shows the proportion of instances in which collaboration emerges between minority and majority groups in these models. We hold the payoff associated with playing Hare fixed at 2, while Low is varied from 2.2 to 5. We find that unfair norms of division disincentivize collaboration between types. The greater the inequity (the lower Low), the less likely it is that collaboration emerges across groups. This effect is more pronounced the smaller the minority population. When Low = 2.2 and the minority group makes up 1% of the population, for example, collaboration only occurs between types in about 10% of simulations. Again, this is particularly disastrous to the minority for by and large many of their interactions are with members of the majority, meaning successful collaboration is relatively unlikely. Note that while it may seem like a *good* thing for minority agents to avoid collaboration with majority agents if they receive poorer outcomes, in these models there is still always an advantage to collaboration (G is always less than Low).[30]

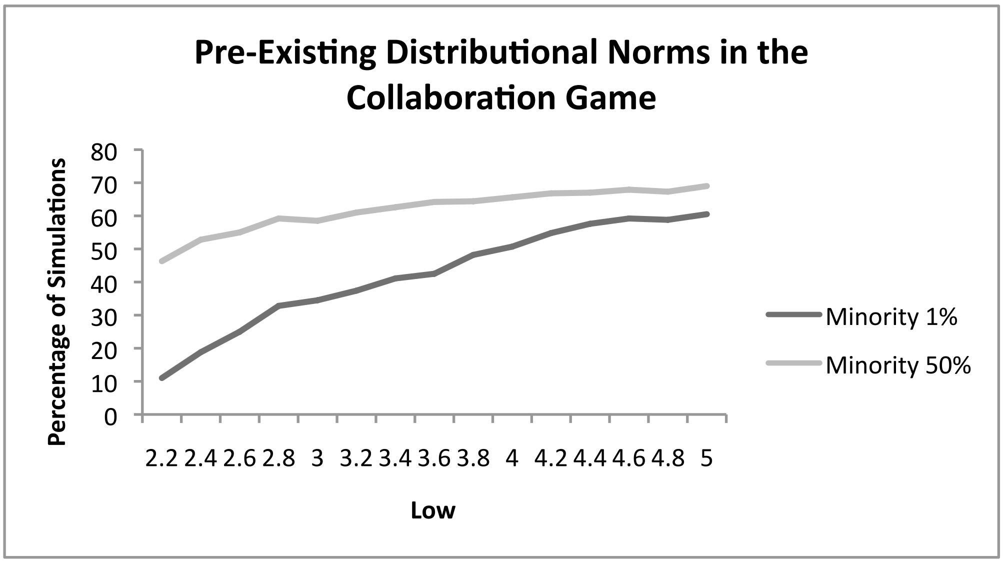

Figure 7: Simulations of populations playing the type-conditional collaboration game with existing norms of division. The x-axis represents different values for Low (the payoff to minority types for collaboration). The y-axis represents the proportions of simulations that arrived at inter-group collaboration.

These results again arise due to differential learning across populations. When actors

[30]These results are similar to some from Wagner (2012).

play the collaboration game with existing norms of division, Hare is the safest strategy as it always guarantees a payoff. When minority types get relatively little for bargaining, they are disincentivized from taking the risk of collaborating. This means that they quickly learn to play Hare in these simulations, and the smaller the minority, the more quickly this happens. Majority types then slowly learn to play Hare against them as well.

Some empirical observations are consistent with the results just presented. In many fields, women are less likely to collaborate and more likely to collaborate with other women (McDowell and Smith, 1992; Ferber and Teiman, 1980; West et al., 2013; Boschini and Sjögren, 2007).[31] McDowell and Smith (1992) have argued that the 'productivity gap' between male and female economists could be explained by the fact that female economists lose the benefits of collaborative endeavors. Del Carmen and Bing (2000) find that African American authors in criminology are highly likely to work alone.

One phenomena that may be partially explained by these outcomes is the clustering of certain minority groups into academic subdisciplines. Botts et al. (2014), for example, find that black philosophers, who they calculate to make up about 1.3 % of U.S. philosophers, tend to work disproportionally in certain subfields. If minority types do, in fact, learn to avoid collaborating and cooperating with majority types, this may lead minority types to seek out areas where they are more likely to encounter others like themselves. Alternatively, in subdisciplines where minority types have higher numbers, they may be less likely to be disadvantaged in strategic interactions. If so, this may attract other minority types to such subdisciplines.

5.2.4 Minority Disadvantage and Existing Bias

The work we have presented to this point all involves simulations where we do not make assumptions about the starting proportions of strategies in a population. In other words, our simulations randomize over the starting points epistemic communities might take—half of actors make fair demands, or 1/3, or 90% of actors make High demands, etc.—and then look at what proportions of outcomes arise from this random collection.[32]

For the phenomena we are studying, however, there is little reason to think that the starting points of epistemic communities are random. Epistemic communities are embedded in societies where biased behavior against women and racial minorities already exists. For this reason, one might be especially interested in what happens in these simulations when one starts with a relatively high proportion of majority types who behave in a discriminatory way towards minority types.

In these restricted cases, our results are significantly strengthened. In other words, when one looks only at populations that already have some tendency towards bias, minority disadvantage due to learning speed is more significant. Why would this be the

[31] Some of these results are outdated, but we mention them here because they still potentially support the effects we describe.

[32] In doing so, we are measuring what are called the 'basins of attraction' for populations playing games under the dynamics we use. These basins specify which starting proportions of strategies evolve to which equilibria.

case? When a significant proportion of a majority population is already demanding High or playing Hare against a minority population, the best thing the minority population can do is to respond by demanding Low or playing Hare themselves. If the minority population does this more slowly, the majority population has time to learn other behaviors, meaning that the group may arrive at a non-discriminatory norm. If the minority population does this quickly, and the majority population keeps discriminating because they learn slowly, the group will end up at the discriminatory norm.

We mentioned in section 5.1 that there are games where the minority speed effect leads to minority advantage because generally it benefits the minority to quickly learn to play a strategy that benefits them. Even for these games, though, if one looks only at starting points where there is a significant amount of discrimination in the other population, learning speed instead hurts the minority group. Figure 8 shows outcomes for a Nash demand game where actors can demand 2, 5, or 8. This figure shows the proportion of outcomes where the minority is discriminated against as a function of the size of the minority population.

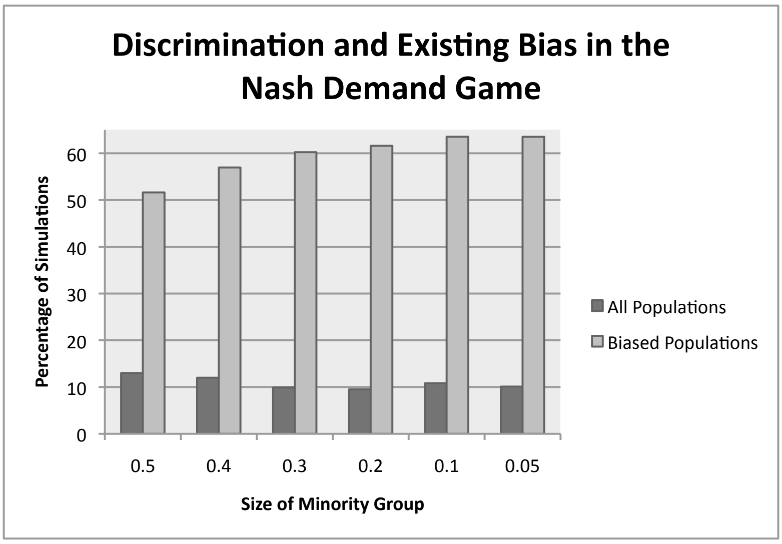

Figure 8: Simulations of populations playing the Nash demand game either starting at any population proportion, or only populations where ≥60% of majority types demand 8. The y-axis represents simulations where the minority is discriminated against.

As this figure shows, over all population starting points, the smaller the minority the more likely it is (by a bit) that they discriminate. Over starting points where more than 60% of the majority demand 8, the smaller the minority the more likely it is that they are discriminated against. In other words, by looking only at realistic starting points, in this case, we see a reversal of the minority speed effect, which now disadvantages minorities. This figure is meant to demonstrate a general observation, which is that the more the majority population discriminates, the worse learning quickly is for the minority population.

5.3 Summary

To summarize the results presented in sections 5.1 and 5.2: minority groups in academia can end up disadvantaged in bargaining, cooperative, and collaborative scenarios. This disadvantage is more likely to arise for small groups as a result of a differential in learning environments for the majority and minority types. This disadvantage is also more likely to arise in situations where actors do not have the option to work alone. When it does arise, it is more harmful to small groups than large ones. And furthermore, it can deter minority types from collaborating with majority types, possibly leading minorities to cluster in academic subdisciplines. And, lastly, if a majority group is already discriminating to some degree, the fact that minority groups quickly learn to deal with them is more detrimental to the minority type.

Significantly, the effects we describe can arise without what one would think of as particularly pernicious behavior on the part of either group. In particular, actors need not be implicitly or explicitly biased against minority types for these effects to arise. The preconditions for these effects are that 1) actors condition behavior on their type of interactive partner, 2) actors learn to behave in ways that benefit themselves, and 3) one type is in the minority. The only one of these conditions that might be thought of as ethically questionable is conditioning behavior on one's type of interacting partner. While this is arguably problematic, it does not have the same ethical character as, say, purposefully tanking a black colleague's tenure case, or harassing female graduate students. It is not clear that actors need even be aware that they are learning to treat types differently in the scenarios these models represent.

Note that this does not mean that the resulting behavior in these models is ethically unproblematic—it is. The point is that from relatively neutral starting points, well-meaning actors can slowly learn to behave in ways consistent with discrimination and bias, and that this is more likely to negatively affect minority groups.[33]

6 Epistemic Minorities and Epistemic Implications

Our analysis thus far has focused on racial and gender minority groups. We will now discuss the possibility of extending these results to epistemic minorities—groups of agents in epistemic communities who hold uncommon views, use uncommon methods, etc. Are epistemic minorities likely to end up disadvantaged by the effects we have outlined?

The question is whether or not researchers condition their bargaining, cooperative, and collaborative behavior on the *epistemic* type of their interactive partner. We certainly think this is plausible when it comes to cooperation and collaboration. It seems

[33] One may question the claim that our models are 'neutral' given that we assume actors condition on traits like gender and race. Why not hair color? Or shoe size? Gender and race tend to be the visible markers which matter to type-conditioning in many real-world communities, and this is why we focus on them. While this feature (which is an explanandum in its own right) is built into the models, the discriminatory norms that arise are not.

likely that epistemic agents decide whether or not to engage in joint projects and joint work based on the views and methods of their potential partner.[34] If so, our results suggest that epistemic majorities may learn to collaborate amongst themselves, but not with epistemic minorities, and that this should be more likely the smaller the minority population. To some degree, outcomes like this are the norm in research communities. It is normal for researchers working on the same problems to work together. (And it is hard to imagine that this is not a good thing for epistemic progress.) On the other hand, if such results occur for epistemic minorities who use different methods or starting assumptions, this may be problematic. As discussed, Longino (1990) argues that a diversity of assumptions is important to creating critical debate about inquiry. If epistemic agents avoid those unlike themselves, this sort of debate will be hampered. Furthermore, if arguments from Kitcher (1990) and Weisberg and Muldoon (2009) are correct, there may be synergistic effects that occur when researchers use diverse methodologies. If researchers avoid interaction with those who use different methods, again this might impact inquiry.

We will now ask: what are the general implications of our work for epistemic progress? As discussed in section 2, diversity is important to epistemic communities in a number of ways. Our work here indicates several processes by which diversity in epistemic communities may be impacted. First, if gender and racial minorities are hurt by discriminatory norms of bargaining, cooperation, and collaboration they may be less likely to stay in epistemic communities, and less likely to gain influential positions in these groups. This may have the effect of preventing attempts to diversify disciplines where minority groups are underrepresented. Second, if gender and racial minorities learn to only collaborate with like types, this may lead them to cluster in subdisciplines. Although this clustering may preserve diversity in the discipline as a whole, it may negatively impact the diversity of research groups. Lastly, as just outlined, the effects we describe may impact epistemic minorities both by hurting their careers, and potentially driving them to avoid interaction with other types.

7 Conclusion

We will conclude by first saying something about the explanatory power of the models provided. We will then discuss how the effects we note can be mitigated and prevented in real world populations. Lastly, we present an appeal to other formal philosophers to work on topics related to diversity.

In many cases it is simply not possible to directly examine real world phenomena using empirical methods. In particular, phenomena that involve many complex moving parts (such as many actors in a research community) and are extended in time (such as social change) cannot easily be directly observed. In cases like this, mathematical

[34] With respect to bargaining behavior, this sort of conditioning might also occur, though it is difficult to assess this empirically. Suppose, for example, that 90% of a scientific community hold an established view and 10% support a less common theory. It is plausible that in joint efforts established types will expect the uncommon types to do more work, or to reap fewer rewards.

models provide important tools in that they allow exploration of causal possibilities that empirical and philosophical methods do not. In a model, one can alter conditions in a controlled way and see how these alterations influence outcomes (Weisberg, 2013). That said, models are, by necessity, simplifications of the real world, and miss important aspects of real world target phenomena. For example, in our models we assume that academics always learn to act in their own best interest. Clearly, in many cases, this assumption will not hold. Sometimes academics are bound by discipline-wide conventions or rules that prevent self-beneficial behavior. Sometimes academics actively take the preferences of their peers into account when deciding how to act, even when this may prevent them from gaining rewards. In using models to explore complicated phenomena, then, one should be careful. In this case it is appropriate to ask: Are the models here explanatory? Do the social dynamical effects we describe help to explain why some groups continue to be underrepresented in some disciplines?

There are at least two reasons to think that the models presented may have genuine explanatory power. First, the general results described are robust. When actors condition on types, the dynamics of an evolving population are completely altered. These type-conditional effects occur across games and over many different parameter settings in each of these. The minority effects outlined are likewise robust. They occur because of a disparity in the learning situations that minority and majority groups find themselves in. These effects again happen across models and for myriad parameter settings. Robust results are arguably more likely to be genuinely explanatory of real world phenomena since the real world situations being modeled are more likely to fall under the set of conditions where these results arise.

Second, the aspects of the models here that lead to disadvantage are features of the real world. As discussed, minority disadvantage occurs in these models whenever actors 1) learn to behave in their own best interest, 2) condition on types, and 3) one type is in the minority. Experimental work on humans indicates that they (unsurprisingly) learn to do what benefits them in games. The second condition, as discussed, has been widely verified across academic settings. And obviously in many academic communities minority groups exist.

One last point should be made before continuing. Our models *do not* capture behaviors that are widely thought to disadvantage certain groups in epistemic communities, including stereotype threat, implicit bias, and explicit bias. This could be thought of as a limitation of our models, but we take this to provide a sort of explanatory power instead. What our models show is that even if these factors are eradicated in epistemic communities, if there are minority groups in these communities, disadvantage can arise for these minorities groups.

As argued, the results presented here show how minority populations can be disadvantaged as a result of the underlying social dynamics of interactions between populations of different sizes. This said, these dynamical effects are not inevitable. We now discuss ways for mitigating or preventing these effects. The first thing to note is that minority disadvantage is less likely to arise in the models presented when the sizes of the interacting populations are closer to equal. For this reason, efforts to increase the presence of underrepresented groups in epistemic communities can be expected to lessen the potential for these effects. Of course, simply increasing numbers of minority groups may not be enough to change established discriminatory norms, but it should at least make the establishment of future such norms less likely.

Even in populations with equal proportions of types, though, discriminatory norms arise in our simulations. The real driver of the results we see is type-conditional behavior. Without it, discriminatory norms like those discussed are not possible. It may not be particularly novel to argue that researchers should treat each other equally. This said, our results give further support to the intuitive argument that type-conditional treatment (even if one attempts to make it separate but equal) can be damaging.

What can be done to decrease type-conditional behavior? Although our models, as discussed, do not explicitly represent damaging behaviors that arise from bias, the literature on implicit bias explores how to reduce type-conditional behavior in a way that may be helpful here. One option is to shape institutional settings in ways that prevent such conditioning from arising in the first place. Lee (forthcoming) includes a detailed discussion of what sorts of environments diminish biased behavior. As she describes, actors can be motivated to prevent prejudiced behavior by their peers groups and by leadership at universities (Hopkins, 2006). Ensuring that actors have enough time and attention to carefully make potentially biased judgements also decreases the effects of bias (Hofmann et al., 2005). A second option is for institutions to adopt specific policies that stop the effects of type-conditional behavior, such as blinding in journal reviewing and merit reviews. There are many possibilities for such policies. For example, West et al. (2013) find that traditionally women have been less likely to hold coveted first and last author positions in collaborative work. In mathematics, however, authorship is determined alphabetically, and the effects noted by these authors do not occur. This is a case where an institutional convention (unintentionally) evened the playing field for women.[35,36]

A number of philosophers have begun using formal methods like those employed here to explore the effects of epistemic diversity in epistemic communities (Zollman, 2010; Weisberg and Muldoon, 2009). We believe that philosophers are also ideally positioned to formally investigate phenomena related to personal diversity in epistemic communities (and elsewhere). Philosophers are traditionally interested in epistemic and social justice issues around diversity. Furthermore, formal modeling is increasingly practiced in philosophy, meaning that the tools for such explorations are aligned with interests in pursuing them. Some potential areas of future research related to the work presented here include tests of how epistemic agents of diverse personal types form collaborative networks, and how effects like those we see occur in communities with intersectional populations.

[35] Of course, this may be an unsatisfactory practice for other reasons. Einav and Yariv (2006) find that authors in economics with earlier surnames tend to be more successful than those with later surnames. They do not observe this effect in disciplines where author order is not typically alphabetical.

[36] When it comes to decreasing these effects for epistemic minorities, things are more tricky. Conditioning behavior on an agent's epistemic type is often justified, and generally seems less ethically problematic.

Given the importance of diversity to epistemic communities, promoting the presence of underrepresented groups is not only important as a step towards social justice, but as an epistemic good. The research presented in this paper points at ways that social dynamical effects in epistemic communities can lead to systematic disadvantage for minority groups, even among well-meaning actors. This may help explain why some epistemic communities continue to fail to diversify. The proposals we present for mitigating these effects are not new, but we hope our results give more weight to those advocating for said proposals.

Bibliography

Alexander, J McKenzie (2000). "Evolutionary explanations of distributive justice." *Philosophy of Science*, 490–516.

Alexander, Jason and Brian Skyrms (1999). "Bargaining with neighbors: Is justice contagious?." *The Journal of philosophy*, 588–598.

Axtell, Robert, Joshua M Epstein, and H Peyton Young. "The emergence of classes in a multi-agent bargaining model.".

Beaver, D deB and Richard Rosen (1979). "Studies in scientific collaboration Part III. Professionalization and the natural history of modern scientific co-authorship." *Scientometrics*, *1*(3), 231–245.

Beede, David, Tiffany Julian, David Langdon, George McKittrick, Beethika Khan, and Mark Doms (2011). "Women in STEM: A Gender Gap to Innovation. ESA Issue Brief# 04-11.." *US Department of Commerce*.

Bennett, Dianne M and David McD Taylor (2003). "Unethical practices in authorship of scientific papers." *Emergency Medicine*, *15*(3), 263–270.

Bertrand, Marianne and Sendhil Mullainathan. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. Technical report, National Bureau of Economic Research, (2003).

Binmore, Ken (2008). "Do Conventions Need to be common knowledge?." *Topoi*, *27*, 17–27.

Börgers, Tilman and Rajiv Sarin (1997). "Learning through reinforcement and replicator dynamics." *Journal of Economic Theory*, *77*(1), 1–14.

Boschini, Anne and Anna Sjögren (2007). "Is team formation gender neutral? Evidence from coauthorship patterns." *Journal of Labor Economics*, *25*(2), 325–365.

Botts, Tina Fernandes, Liam Kofi Bright, Myisha Cherry, Guntur Mallarangeng, and Quayshawn Spencer (2014). "What Is the State of Blacks in Philosophy?." *Critical Philosophy of Race*, *2*(2), 224–242.

Bowles, Hannah Riley, Linda Babcock, and Lei Lai (2007). "Social incentives for gender differences in the propensity to initiate negotiations: Sometimes it does hurt to ask." *Organizational Behavior and Human Decision Processes*, *103*(1), 84–103.

Bruner, Justin P (2013). "Policing epistemic communities." *Episteme*, *10*(04), 403–416.

Bruner, Justin P (2014). "Racists and Minorities in Population Games." working paper.

Bruner, Justin P and Cailin O'Connor (2015). "Power, Bargaining, and Collaboration." *Scientific Collaboration and Collective Knowledge*. Ed. Conor Mayo-Wilson Thomas Boyer and Michael Weisberg. Oxford University Press.

Card, David and Stefano DellaVigna. Nine facts about top journals in economics. Technical report, National Bureau of Economic Research, (2013).

D'Arms, Justin, Robert Batterman, and Krzyzstof Górny (1998). "Game theoretic explanations and the evolution of justice." *Philosophy of Science*, 76–102.

Del Carmen, Alejandro and Robert L Bing (2000). "Academic productivity of African Americans in criminology and criminal justice." *Journal of Criminal Justice Education*, *11*(2), 237–249.

Einav, Liran and Leeat Yariv (2006). "What's in a surname? The effects of surname initials on academic success." *The Journal of Economic Perspectives*, 175–188.

Ellingsen, Tore (1997). "The evolution of bargaining behavior." *The Quarterly Journal of Economics*, 581–602.

Fehr, Carla (2011). "What is in it for me? The benefits of diversity in scientific communities." *Feminist Epistemology and Philosophy of Science*. Ed. Heidi E. Grasswick. Springer.

Ferber, Marianne A and Michelle Teiman (1980). "Are women economists at a disadvantage in publishing journal articles?." *Eastern Economic Journal*, 189–193.

Ginther, Donna K, Walter T Schaffer, Joshua Schnell, Beth Masimore, Faye Liu, Laurel L Haak, and Raynard Kington (2011). "Race, ethnicity, and NIH research awards." *Science*, *333*(6045), 1015–1019.

Gintis, Herbert (2009). *Game theory evolving: A problem-centered introduction to modeling strategic interaction.* Princeton University Press.

Gordon, Michael D (1980). "A critical reassessment of inferred relations between multiple authorship, scientific collaboration, the production of papers and their acceptance for publication." *Scientometrics*, *2*(3), 193–201.

Haraway, Donna Jeanne (1989). *Primate visions: Gender, race, and nature in the world of modern science.* Psychology Press.

Harding, Sandra G (1991). *Whose science? Whose knowledge?: Thinking from women's lives.* Cornell University Press.

Harding, Sandra G (1998). *Is science multicultural?: Postcolonialisms, feminisms, and epistemologies.* Indiana University Press.

Henrich, Joseph and Natalie Henrich (2007). *Why humans cooperate: A cultural and evolutionary explanation.* Oxford University Press, USA.

Hofmann, Wilhelm, Bertram Gawronski, Tobias Gschwendner, Huy Le, and Manfred Schmitt (2005). "A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures." *Personality and Social Psychology Bulletin*, *31*(10), 1369–1385.

Hopkins, Ed (2002). "Two competing models of how people learn in games." *Econometrica*, *70*(6), 2141–2166.

Hopkins, Nancy (2006). "Diversification of a university faculty: Observations on hiring women faculty in the schools of science and engineering at MIT." *MIT Faculty Newsletter*, *18*(4), 25.

Kitcher, Philip (1990). "The division of cognitive labor." *The Journal of Philosophy*, 5–22.

Kuhn, Thomas S (1977). "Objetivity, value judgment, and theory choice.".

Laband, David N (1987). "A qualitative test of journal discrimination against women." *Eastern Economic Journal*, 149–153.

Landry, Réjean, Namatie Traore, and Benoît Godin (1996). "An econometric analysis of the effect of collaboration on academic research productivity." *Higher Education*, *32*(3), 283–301.

Lee, Carole J. (ming). "Revisiting Current Causes of Women's Underrepresentation in Science." *Implicit Bias and Philosophy Volume 1: Metaphysics and Epistemology.* Ed. J. Saul and M. Brownstein. Oxford University Press.

Lee, Sooho and Barry Bozeman (2005). "The impact of research collaboration on scientific productivity." *Social studies of science*, *35*(5), 673–702.

Longino, Helen E (1990). *Science as social knowledge: Values and objectivity in scientific inquiry.* Princeton University Press.

Longino, Helen E (2002). *The fate of knowledge.* Princeton University Press.

McDowell, John M and Janet Kiholm Smith (1992). "The effect of gender-sorting on propensity to coauthor: Implications for academic promotion." *Economic Inquiry*, *30*(1), 68–82.

Milkman, Katherine L, Modupe Akinola, and Dolly Chugh (2014). "What Happens Before? A Field Experiment Exploring How Pay and Representation Differentially Shape Bias on the Pathway into Organizations." *A Field Experiment Exploring How Pay and Representation Differentially Shape Bias on the Pathway into Organizations (April 23, 2014)*.

Morrison, Philip S, Gill Dobbie, and Fiona J McDonald (2003). "Research collaboration among university scientists." *Higher Education Research & Development*, *22*(3), 275–296.

Moss-Racusin, Corinne A, John F Dovidio, Victoria L Brescoll, Mark J Graham, and Jo Handelsman (2012). "Science faculty's subtle gender biases favor male students." *Proceedings of the National Academy of Sciences*, *109*(41), 16474–16479.

Nelson, Lynn Hankinson (1993). "Epistemological communities." *Feminist epistemologies*, 121–160.

Norlock, Kathryn. Women in the Profession: a more formal report to the CSW. Technical report, Lady Eaton College, (2006).

Paulus, Paul B and Bernard A Nijstad (2003). *Group creativity: Innovation through collaboration.* Oxford University Press.

Perna, Laura W (2001). "Sex and race differences in faculty tenure and promotion." *Research in Higher Education*, *42*(5), 541–567.

Potter, Elizabeth (1993). "Gender and epistemic negotiation." *Feminist epistemologies*, 161–186.

Reuben, Ernesto, Paola Sapienza, and Luigi Zingales (2014). "How stereotypes impair women's careers in science." *Proceedings of the National Academy of Sciences*, *111*(12), 4403–4408.

Richerson, Peter J and Robert Boyd (2008). *Not by genes alone: How culture transformed human evolution.* University of Chicago Press.

Saul, Jennifer (2011). "Implicit bias, stereotype threat and women in philosophy." *Women in philosophy: What needs to change.*

Skyrms, B. (1996). *Evolution of the Social Contract.* Cambridge: Cambridge University Press.

Skyrms, Brian (1994). "Sex and justice." *The Journal of philosophy*, 305–320.

Skyrms, Brian and Kevin JS Zollman (2010). "Evolutionary considerations in the framing of social norms." *Politics, Philosophy & Economics*, *9*(3), 265–273.

Steinpreis, Rhea E, Katie A Anders, and Dawn Ritzke (1999). "The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: A national empirical study." *Sex roles*, *41*(7-8), 509–528.

Strevens, Michael (2003). "The role of the priority rule in science." *The Journal of philosophy*, 55–79.

Sugimoto, Cassidy R. "Global gender disparities in science.".

Tilcsik, András (2011). "Pride and Prejudice: Employment Discrimination against Openly Gay Men in the United States1." *American Journal of Sociology*, *117*(2), 586–626.

Tinsley, Catherine H, Sandra I Cheldelin, Andrea Kupfer Schneider, and Emily T Amanatullah (2009). "Women at the bargaining table: Pitfalls and prospects." *Negotiation Journal*, *25*(2), 233–248.

Wagner, Elliott O (2012). "Evolving to Divide the Fruits of Cooperation*." *Philosophy of Science*, *79*(1), 81–94.

Weibull, Jörgen W (1997). *Evolutionary game theory.* MIT press.

Weisberg, Michael (2013). *Simulation and similarity: using models to understand the world.* OUP USA.

Weisberg, Michael and Ryan Muldoon (2009). "Epistemic Landscapes and the Division of Cognitive Labor*." *Philosophy of Science*, *76*(2), 225–252.

West, Jevin D., Jennifer Jacquet, Molly King, Shelley Correll, and Carl Bergstrom (2013). "The Role of Gender in Scholarly Authorship." *PLoS ONE*, *8*(7).

Wylie, Alison (2007). "The constitution of archaeological evidence." *The Archaeology of Identities: a Reader*, 97–18.

Young, H Peyton (1993). "An evolutionary model of bargaining." *Journal of Economic Theory*, *59*(1), 145–168.

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