Dynamics and Diversity in Epistemic Communities

Cailin O'Connor and Justin Bruner

This is a post-peer-review, pre-copyedit version of an article published in Erkenntnis. The final authenticated version is available online at: http://dx.doi.org/10.1007/s10670-017-9950-y

Abstract

Bruner (2017) shows that in cultural interactions, members of minority groups will learn to interact with members of majority groups more quickly—minorities tend to meet majorities more often as a brute fact of their respective numbers—and, as a result, may come to be disadvantaged in situations where they divide resources. In this paper, we discuss the implications of this effect for epistemic communities. We use evolutionary game theoretic methods to show that minority groups *can* end up disadvantaged in academic interactions like bargaining and collaboration as a result of this effect. These outcomes are more likely, in our models, the smaller the minority group. They occur despite assumptions that majority and minority groups do not differ with respect to skill level, personality, preference, or competence of any sort. Furthermore, as we will argue, these disadvantaged outcomes for minority groups may negatively impact the progress of epistemic communities.

1 Introduction

Bruner (2017) shows that in cultural interactions, members of minority groups learn to interact with members of majority groups more quickly—minorities tend to meet majorities much more often as a brute fact of their respective numbers—and, as a result, may come to be disadvantaged in situations where they divide resources. In this paper, we discuss the implications of this effect for epistemic communities.[1] In particular we ask: when authors develop norms for bargaining and collaboration, can this effect disadvantage minority groups? What about groups in an *epistemic minority*—those who take minority views, or use less popular methodologies?

We use evolutionary game theoretic methods to show that minority groups *can* end up disadvantaged in academic interactions as a result of this effect. These outcomes are more likely, in our models, the smaller the minority group. They occur despite assumptions that majority and minority groups do not differ with respect to skill level, personality, preference, or competence of any sort. Furthermore, as we will argue, these

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

disadvantaged outcomes for minority groups may help explain why some groups, including women and racial minorities, tend to cluster into subdisciplines in academia. As we will point out these effects may impact epistemic progress both by decreasing gender and racial diversity in epistemic communities and so potentially decreasing the epistemic diversity of these communities, and also by directly impacting those in epistemic minorities.

One of the most striking things about the effects we describe is that this minority disadvantage occurs even in models of populations where actors do not display the usual psychological effects associated with minority disadvantage in epistemic communities. They do not experience stereotype threat. They do not have implicit biases. There is no question that these psychological phenomena are an important part of explaining the failure of some academic fields to diversify (Saul, 2011). But our results imply that simply tackling these particular psychological effects may not be enough to eradicate inequitable norms in epistemic groups.

Our paper will proceed as follows. In section 2, we justify the use of evolutionary game theory to model epistemic communities, and describe this methodology. In section 3, we outline the games we use to represent collaboration and bargaining in academia. In section 4, we describe our results as well as relevant work from other authors. In section 5 we discuss the relevance of these results to epistemic progress. In the conclusion, we assess the epistemic role these models can play, and briefly describe policy measures that may help prevent the social dynamical effects we describe.

2 Methodology

Evolutionary game theory is a methodology used 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 briefly describe evolutionary game theoretic methods.[2] 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.[3]

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 by appeal to evolution, cultural evolution, or learning. This is done by applying what are called *dynamics* to games—rules that determine how populations of actors playing a game will change based on past interactions. Throughout this paper, we will employ the

[2]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).

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

replicator dynamics to model behavioral change in epistemic communities.[4] 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.[5] Individuals in an epistemic community are then assumed to imitate and repeat successful actions that gained credit in the past, leading to the proliferation of such successful actions.

One further concept we must mention before continuing is that of a *Nash equilibrium*. A Nash equilibrium for a game is a set of strategies where no player may deviate and improve her payoff. These strategy sets tend to be stable and to arise in real world settings for this reason. They also arise in many evolutionary settings, making them relevant to our evolutionary analysis.

3 Epistemic Games and Type-Conditional Behavior

As just discussed, 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 two sorts of strategic interaction that we focus on here—bargaining and collaboration. We start with bargaining. In the Nash demand game, actors each demand a certain portion of a resource. If the demands of the actors do not exceed the resource, each receives the portion they 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'.[6] A payoff table lists payoffs to actors for any combination of possible strategies in a game. This game has three (pure strategy) Nash equilibria—High v Low, Med v Med, and Low v High.[7] 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.

We will use this game to model certain interactions between epistemic agents. Bargaining may not seem like a central part of research because there are only a few

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

[5]Kitcher (1990) and Strevens (2003) appeal to the 'credit economy' to model academics in a classical game-theoretic framework, while Bruner (2013) uses similar assumptions in a dynamic model involving cultural evolution.

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

[7]Pure strategies are ones where actors always take the same action rather than randomly mixing over multiple actions. This game (and the following one) 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.

		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.

situations—salary and funding negotiations, for example—in which researchers explicitly bargain over resources. In any sort of joint project, however—conference planning, committee work, running a department, or collaborative research—by some process of implicit or explicit negotiation, 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. 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.

The second game is a model of combined joint action and bargaining. We call it the collaboration game. Two actors first decide whether to enter a collaborative partnership (or joint project), or work alone. Should the two actors choose to work together they gain greater joint payoff for mutual work, but must decide how to divide the resource obtained. As is clear from figure 2, there are four strategies in this game—the uncollaborative Solo strategy and three collaborative strategies where actors demand Low, Med, or High of their partner. Solo always generates a dependable, low payoff. Collaboration leads to a higher joint payoff, though it is now risky for two reasons: 1) a partner could choose Solo or 2) bargaining could fail. Either of these outcomes leads to a payoff of zero. This game has up to four Nash equilibria depending on the payoffs. Solo v Solo is always an equilibrium (if your partner will not work with you, there is no point in trying). If the Solo option does not yield too high a payoff, the three compatible bargaining demands—Low v High, Med v Med, and High v Low—are equilibria as well.[8]

This model captures a strategic scenario that many collaborating or cooperating academics encounter. Academics typically must decide whether or not to collaborate with a potential co-author. Collaboration is beneficial because it comes with the potential for rewards.[9] It is risky, however, because if a collaborative partner fails to fulfill her part of the project properly, or engages in academic dishonesty, her co-authors may be

[8]This game is extensively analyzed by Wagner (2012) who calls it the stag hunt/divide the dollar game to capture the sense that it combines joint action and bargaining. It is equivalent to a Nash demand game with an outside option.

[9]Co-authored papers are more likely to be accepted to top journals in many fields, and are cited more (Laband, 1987; Gordon, 1980; Beaver and Rosen, 1979; Card and DellaVigna, 2013). Many authors have argued that collaboration improves overall academic productivity (Morrison et al., 2003; Landry et al., 1996; Lee and Bozeman, 2005).

		Player 2		
	Low	**Med**	**High**	**Solo**
Player 1 **Low**	3,3	3,5	3,7	0,4
Med	5,3	5,5	0,0	0,4
High	7,3	0,0	0,0	0,4
Solo	4,0	4,0	4,0	4,4

Figure 2: 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.

disadvantaged.

Once collaboration begins, actors also must bargain either implicitly or explicitly to decide how the efforts and rewards of collaboration will be divided. What will the author order be? Who will do what portion of the work? 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.[10]

Typical game theoretic analyses consider interacting agents who behave the same way upon meeting any other agent. In order to investigate the situation of minority groups in epistemic communities, though, we need models where actors have the option to treat each other differently based on aspects of social identity, like gender and race.[11] (Later in the paper, we will also ask what happens when actors can condition behavior on aspects of epistemic identity—like methodology used, or background assumptions made.) In order to capture this, we investigate games where actors are of two different types and condition their behavior on the type of their interactive partner. Importantly, strategies for the type-conditional versions of the games described above are significantly altered. Instead of strategies like 'Solo' or 'High', conditioning actors choose strategies like 'play Solo with my in-group, play High with my out-group'. We will use the notation <Solo; High> to denote a joint strategy with the in-group strategy listed first and the out-group second.

With these new strategies come new equilibria where actors play one Nash equilibrium against their in-group members and another against out-group members. Consider

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

[11]This occurs in real academic communities. 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 (Moss-Racusin et al., 2012; Steinpreis et al., 1999). Similar findings, some outside of academia, have been garnered for job candidates who are LGBTQ or racial minorities (Tilcsik, 2011; Bertrand and Mullainathan, 2003).

an academic community where women and men regularly bargain and where women play the strategy <Med; Low> and men play <Med; High>. In this case, both men and women make fair bargaining demands against like types, but when men and women interact men demand greater resources and women concede. Despite the fact that, in a case like this, women might prefer to receive better bargaining outcomes, because men are all demanding High, women cannot switch strategies and improve their lot. Or suppose that white researchers play <Med; Solo> and researchers of color play <Solo; Solo>. In this case, white researchers collaborate amongst themselves, but not with researchers of color, and researchers of color always work alone. Again, researchers of color cannot make unilateral changes to reach a cooperative outcome with white researchers.

These examples of new equilibria in these games are significant for the current discussion because they can be interpreted as involving discriminatory norms (Axtell et al., 2001; Bruner, 2017). At these equilibria, both populations treat out-group members differently from 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 collaborative work and 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.

4 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. The dynamics employed here will be the replicator dynamics which, as discussed in section 2, can be used to model social and individual learning.[12] 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).[13] We start by describing previous result germane to this discussion. We then present new results on groups

[12]We use a version of the discrete time, two population replicator dynamics where one population may be smaller than the other, and where all actors interact with both populations. Let x and y represent the two populations so that $x + y = 1$ and $x \leq y$. Strategies for population x update according to $x'_i = x_i(\frac{f_i(x,y)}{\sum_{j=1}^{n} f_j(x,y)x_j})$ where x_i is the proportion of the x population playing strategy i, $f_i(x,y)$ is the fitness of type i in x given the population states of x and y, and $\sum_{j=1}^{n} f_j(x,y)x_j$ is the average population fitness for x given the states of x and y. Strategies for population y update according to the analogous dynamics.

[13]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 have indeed been found to be sensitive to social identity (Henrich and Henrich, 2007; Killian, 1990; Losin et al., 2012).

playing the collaboration game.

4.1 Previous Results

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. (2001) find that discriminatory norms commonly evolve in a Nash demand game model of such a situation.[14] Perhaps surprisingly, without types, the fair division, Med v Med, is seen most often in such models (Skyrms, 1996; 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) strategy that is unavailable in a uniform population.[15]

Axtell et al. look only at situations with equally sized groups. When Bruner (2017) looks at cases where one population is in a minority, he finds that the minority can be at a significant disadvantage as far as bargaining outcomes go. For some Nash demand games, 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.

Why the disadvantage for minority groups? 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 of the majority. In simulations, the minority group thus quickly learns to adapt to the majority by playing Low. 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.[16]

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). We do not focus on these situations since our

[14]They use dynamics where actors best respond to memory of past play. Bruner (2017) finds that for the replicator dynamics such norms likewise arise. Skyrms and Zollman (2010) find similar results.

[15]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 strategies.

[16]This effect is analogous to what is called the Red King effect in biology. When one of two mutualistic biological species evolves more slowly than the other, under the right conditions they can gain an advantage over their mutualistic partners (Bergstrom and Lachmann, 2003). For this reason, this minority/majority effect has been described elsewhere as the cultural Red King effect (Bruner, 2017; Rubin and O'Connor, 2017; O'Connor, 2017a,b).

aim is to investigate cases where minority groups are potentially disadvantaged by social dynamics in the absence of other factors. Furthermore, O'Connor (2017a,b) argues that in the cultural case, minority disadvantage as a result of learning speed is much more likely to occur.

4.2 New Results: Collaboration

We now consider how the effect described in the last section might disadvantage minorities in epistemic communities where actors develop conventions around collaboration.

4.2.1 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 likely to reach fair bargaining outcomes. And when they reach unfair outcomes, they are each equally likely to discriminate against the other. When one type is in the minority, however, it is increasingly likely that the groups reach unfair bargaining outcomes when they collaborate, and increasingly likely that the minority is disadvantaged in that they demand Low against majority types who demand High. Figure 3 shows results for 10,000 runs of replicator dynamics simulations of this scenario. The payoff associated with Solo was held fixed at 2.5. Low, Med, and High were set at 4, 5 and 6, respectively. The x-axis shows the size of the minority population, and the y-axis shows which percentages of runs that ended in collaboration reached each of the possible outcomes. As can be seen in this figure, when the minority is very small, they end up disadvantaged in more than half of these simulations. One thing that should be noted is that when a minority ends up at a disadvantage with respect to a majority type this outcome can be particularly damaging because minority members tend to interact with majority members more often and so are most often receiving the Low payoff.

4.2.2 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 Solo (S). Note that when S = 0, the collaboration game is essentially equivalent to the Nash demand game—individual work is completely useless so actors must always collaborate and bargain. Slightly higher Ss represent cases where individual work may not be as beneficial as collaborating, but can still provide some benefit.

We find that as S decreases, collaboration increases. This is unsurprising as the payoff to individual work is poor when S is low. When S is low, however, it also becomes increasingly likely that bargaining favors members of the majority group. Figure 4 illustrates these results. The minority constitutes 5% of the total population in this

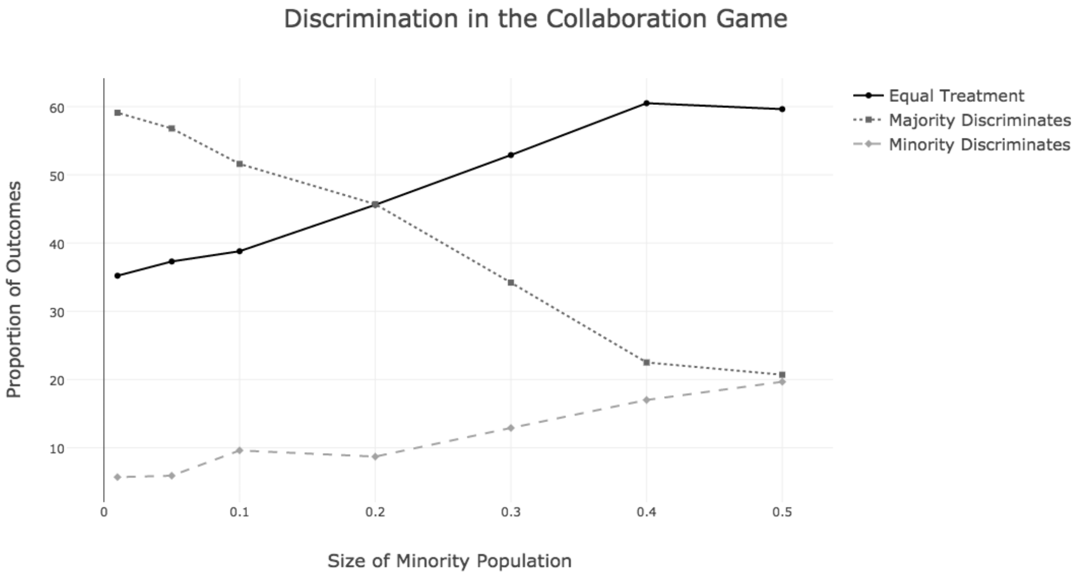

Figure 3: Simulations of populations playing the type-conditional collaboration game. The x-axis represents different sizes of minority populations. The y-axis represents the proportions of simulations that arrived at different outcomes.

figure, and the possible bargaining demands are 4, 5, and 6. As is evident, when S is low the majority discriminates against the minority in over 60% of simulations and the minority almost never discriminates against the majority.

These results occur because when S 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.

4.2.3 Discrimination and Deterrence

We now explore cases in which norms of division have already become cemented in a discipline. Agents still must choose whether to collaborate or not, but if collaboration does occur, the individuals involved do not engage in bargaining, but 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.[17]

[17] 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.

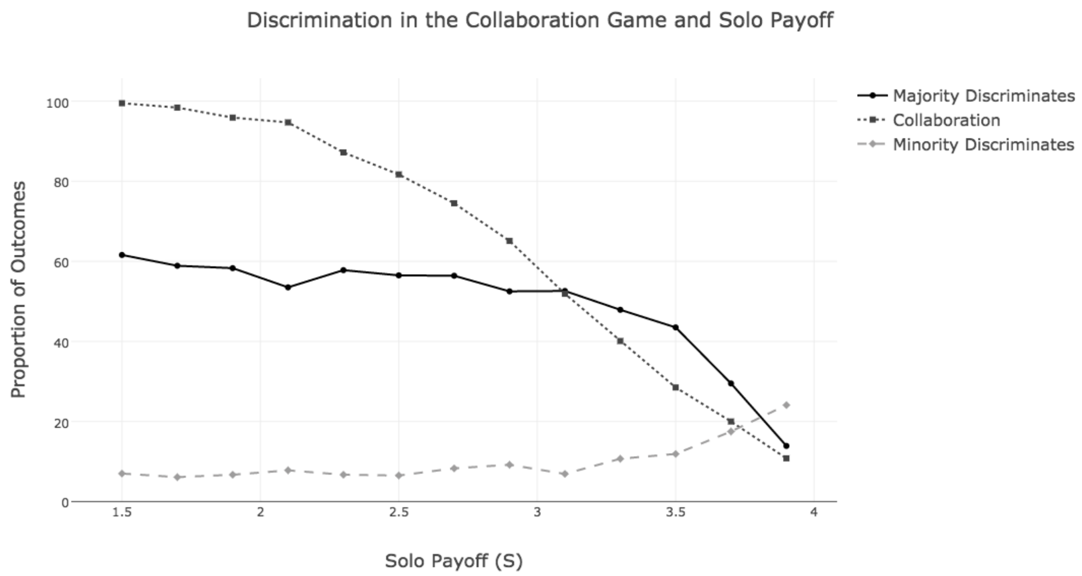

Figure 4: Simulations of populations playing the type-conditional collaboration game. The x-axis represents different values for S. The y-axis represents the proportions of simulations that arrived at inter-group collaboration with different bargaining outcomes.

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, but cannot control how an academic community reacts 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. Sarsons et al. (2015) show that in the field of economics this is commonplace in the case of gender, and may explain differential tenure rates for men and women.[18] The models here thus capture both scenarios where groups bargain differently over collaboration due to existing social norms, and situations where norms undervalue the contributions of underrepresented groups.

Figure 5 shows the proportion of instances in which collaboration emerges between minority and majority groups in these models. We hold the payoff associated with playing Solo 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

[18]Amazingly, they find that when women co-author with each other, the effect is not as strong, indicating that cross gender collaboration results in particularly little assigned credit for women.

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 (Solo is always less than Low).[19] It also should be noted that when collaboration does not emerge between groups the average payoff for the entire community drops because solo work generates less credit than collaborative work. Although the minority group is hurt more by this, the majority group is also hurt when they do not collaborate with minority members.

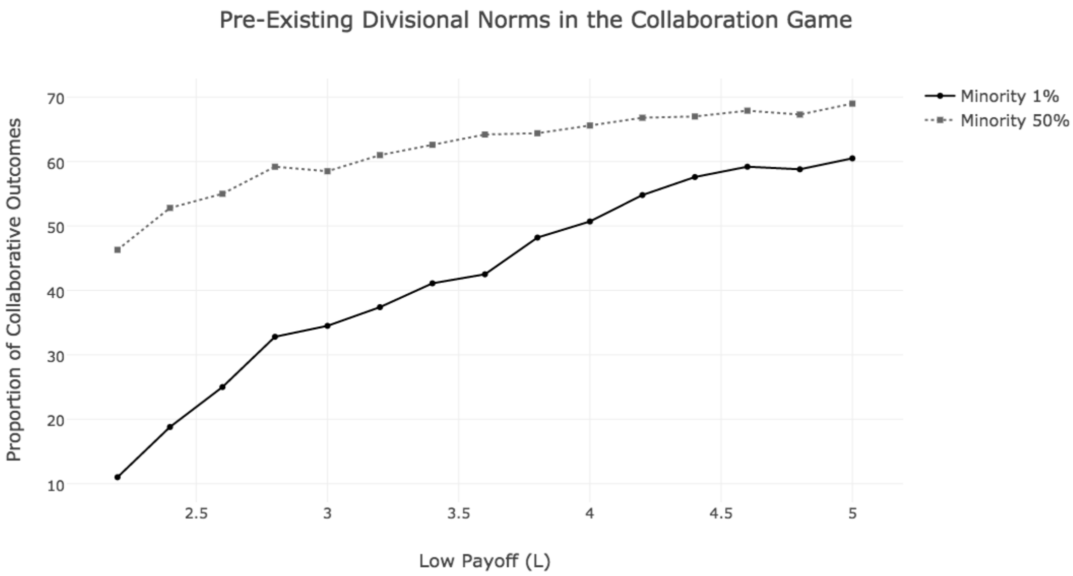

Figure 5: Simulations of populations playing the type-conditional collaboration game with existing norms of division. The x-axis represents different values for Low (the payoff to minority types for collaboration). The y-axis represents the proportions of simulations that arrived at inter-group collaboration.

These results again arise due to differential learning across populations. When actors play the collaboration game with existing norms of division, Solo is a safe 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 Solo in these simulations, and the smaller the minority, the more quickly this happens. Majority types then slowly learn to play Solo 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

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

and Sjögren, 2007). 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 phenomenon 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 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.

4.2.4 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.[20]

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 often embedded in societies where discriminatory 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. Why would this be the case? When a significant proportion of a majority population is already demanding High or playing Solo against a minority population, the best thing the minority population can do is to respond by demanding Low or playing Solo 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 that there are games where the minority 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 populations where there is a significant amount of discrimination in the other population, learning speed instead hurts the minority group.

[20]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 proportions of population states evolve to which equilibria.

4.3 Summary

To summarize the new results presented: minority groups in academia can end up disadvantaged in collaborative scenarios as a result of differential learning speed. This disadvantage is more likely to arise for small groups. 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. 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 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. Note that this does not mean that the resulting behavior in these models represents something ethically unproblematic—it does. The point is that from relatively neutral starting points, actors can slowly learn to behave in ways consistent with discrimination and bias, and that this is more likely to negatively affect minority groups.

5 Epistemic Minorities and Epistemic Implications

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. Longino (1990, 2002) contends that research communities that hold uniform assumptions and prejudices will lack the critical perspective to challenge these viewpoints. Zollman (2010) uses models to argue that epistemic communities that hold a variety of beliefs may be more likely to eventually converge on successful theories. Kuhn (2013) argues that disagreement among scientists is a crucial part of theory change and that this will 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.

Thoma (2015) uses models to conclude that populations where agents employ mixtures of problem choice strategies are, in some cases, more successful than uniform groups.[21] 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.

It has further been argued that gender and racial diversity, in many cases, correlate with epistemic diversity because social identity influences the research problems, assumptions, and prejudices of individuals.[22] For example, 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. 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. In this case gender played a role in leading scientists to challenge a problematic dominant paradigm (albeit with their own value-laden theory).[23]

This connection between personal diversity and epistemic diversity implies that cultural Red King effects may impact the progress of epistemic communities where minority groups leave the field, or do not collaborate with majority types. A further question is whether this effect extends 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 and collaborative behavior on the *epistemic* type of their interactive partner. We certainly think this is plausible in the case of collaboration. 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, 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 outcomes occur for epistemic minorities who use different methods or starting assumptions, this may be problematic. If epistemic agents avoid those unlike themselves, the sort of critical debate championed by Longino (1990) may be hampered and the synergistic effects that occur when researchers use diverse methodologies argued for by Kitcher (1990) and Thoma (2015) might be lost.

6 Conclusion

It may not always be possible to empirically examine processes that involve complex moving parts (such as many actors in a research community) and that are extended in

[21]Weisberg and Muldoon (2009) introduced an earlier model with similar results, but see Thoma (2015) and Alexander et al. (2015).

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

[23]It is harder to find clear cases where personal diversity is important in areas like physics or mathematics, though some have argued for such effects (Harding, 1991, 1998; Fehr, 2011).

time (such as social change). In cases like this, mathematical models 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 one must take care when using them to inform complex processes. In this case it is appropriate to ask: Are the models here explanatory? What epistemic roles can they play?

There are at least two reasons to think that the results described here at least have the potential to occur in real communities. 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 dynamics. 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, and, again, happen across games and dynamics (Bruner, 2017; O'Connor, 2017a,b; Rubin and O'Connor, 2017). 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.

This points to an epistemic role for these models as telling 'how-potentially' stories (Rosenstock et al., 2017). They can direct our attention to an effect that has the potential to really occur, and prompt further investigation, both theoretical and empirical, into this effect. There is an another epistemic role for these models as well, which is to outline the surprisingly sparse preconditions necessary to produce minority disadvantage. Our models do not capture behaviors that are widely thought to disadvantage underrepresented groups in epistemic communities, including stereotype threat and implicit bias. But even if these factors are eradicated in epistemic communities, if the above preconditions hold, disadvantage can arise for these minorities groups. This is a version of 'how-possibly' modeling—it specifies under what conditions a surprising effect can possibly arise.

Before finishing, we would like to briefly discuss ways for mitigating or preventing detrimental minority effects in epistemic communities. The real driver of the results we see is type-conditional behavior. Without it, discriminatory conventions 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 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 to prevent type conditioning from occurring at all. Actors can be motivated to prevent prejudiced behavior by their peers groups and by leadership at universities (Hopkins, 2006; Lee, 2016). 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.[24]

In this paper, we argue that the minority effects may have implications for epistemic progress. We do this by modeling the emergence of conventions of bargaining and collaboration in these communities. While philosophers have extensively modeled aspects of epistemic communities related to theory choice and research patterns, there has been relatively little modeling of the impact of conventional and normative patterns of behavior on epistemic progress. There are many promising extensions for this work, for example, into the emergence of collaboration in groups with intersectional social identities and into the effects of such conventions on the formation on collaboration networks.

Bibliography

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

Alexander, Jason McKenzie, Johannes Himmelreich, and Christopher Thompson (2015). "Epistemic Landscapes, Optimal Search, and the Division of Cognitive Labor." *Philosophy of Science*, *82*(3), 424–453.

Axtell, Robert, Joshua M Epstein, and H Peyton Young (2001). "The emergence of classes in a multi-agent bargaining model." *Social Dynamics*, 191–211.

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.

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

Bergstrom, Carl T and Michael Lachmann (2003). "The Red King effect: when the slowest runner wins the coevolutionary race." *Proceedings of the National Academy of Sciences*, *100*(2), 593–598.

[24]For an interesting 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.

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.

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

Bruner, Justin P. (2017). "Minority Disadvantage in Population Games." *Synthese*, (DOI:10.1007/s11229-017-1487-8).

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.

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.

Killian, Janice N (1990). "Effect of model characteristics on musical preference of junior high students." *Journal of Research in Music Education*, *38*(2), 115–123.

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

Kuhn, Thomas S (2013). "Objectivity, value judgment, and theory choice." *Arguing About Science*. Ed. Alexander Bird and James Ladyman. Routledge.

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. (2016). "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.

Losin, Elizabeth A Reynolds, Macro Iacoboni, Alia Martin, and Mirella Dapretto (2012). "Own-gender imitation activates the brain's reward circuitry." *Social cognitive and affective neuroscience*, *7*(7), 804–810.

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.

O'Connor, Cailin (2017a). "The cultural red king effect." *The Journal of Mathematical Sociology*, *41*(3), 155–171.

O'Connor, Cailin (2017b). *The Dynamics of Inequity.*

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.

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

Rosenstock, Sarita, Justin Bruner, and Cailin O?Connor (2017). "In Epistemic Networks, Is Less Really More?." *Philosophy of Science*, *84*(2), 234–252.

Rubin, Hannah and Cailin O'Connor (2017). "Discrimination and Collaboration." *Philosophy of Science*, (forthcoming).

Sarsons, Heather et al. Gender differences in recognition for group work. Technical report, (2015).

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 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.

Thoma, Johanna (2015). "The Epistemic Division of Labor Revisited." *Philosophy of Science*, *82*(3), 454–472.

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

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.