Inequality and Inequity in the Emergence of Conventions

Calvin Cochran and Cailin O'Connor

January 11, 2019

Many societies have norms of equity—that those who make symmetric social contributions deserve symmetric rewards. Despite this, there are widespread patterns of social inequity, especially along gender and racial lines. It is often the case that members of certain social groups receive greater rewards per contribution than others. In this paper, we draw on evolutionary game theory to show that the emergence of this sort of convention is far from surprising. In simple cultural evolutionary models, inequity is much more likely to emerge than equity, despite the presence of stable, equitable outcomes that groups might instead learn. As we outline, social groups provide a way to break symmetry between actors in determining both contribution and reward in joint projects.

1 Introduction

It has been widely observed that cross-culturally women tend to do more overall work supporting households, and tend to have less free time, than men do (Coltrane, 2000; Bianchi et al., 2006; Treas and Drobnic, 2010). Despite this, women are poorer, on average, than men even in highly developed nations (Casper et al., 1994; Pressman, 2002). When it comes to collaborative research in science, it also seems to be the case that in some disciplines women's contributions to the production of research are undercompensated. They get less credit for collaborative projects than men do, even when doing a large portion of work (West et al., 2013; Sugimoto, 2013; Feldon et al., 2017).

These are examples of inequality in the sense that the division of social resources favors those in one social group. They are also examples of inequity, by which we mean that symmetric contributions to a social good are not met with symmetric rewards. Women garner less economic compensation per hour contributed to household production. Women get less academic credit per time spent on scientific research. These observations are perhaps surprising given the widespread acceptance of social norms supporting equity (Adams and Freedman, 1976; Freedman and Montanari, 1980).[1] One question we might ask is: despite explicit endorsement for equitable reward structures, what sorts of social processes might lead to inequitable ones?

Previous authors have used evolutionary game theoretic models to help illuminate how inherently unequal conventions can emerge between social groups.[2] In particular, these models show that when societies are divided into social categories (men and women, black and white people, Christians and Muslims) the dynamics of social learning and cultural evolution can lead to unequal divisions of resources for no particular reason.[3] In other words, in these models completely identical groups with an option to divide resources equally often end up doing it unequally simply as a result of their different group memberships.

In this paper, we look to explore not the emergence of inequality, but the emergence of inequity. We ask: can these sorts of simple cultural dynamics drive groups to reward

[1] Starting with Adams (1963), equity theory—which posits that in many societies people find inequity highly unpleasant and seek to minimize it—has had many explanatory successes. See Van den Bos (2001) for an overview.

[2] See, for example, Axtell et al. (2000); Poza et al. (2011); Gallo (2014); Bruner (2017); Bruner and O'Connor (2015); O'Connor and Bruner (2017); Rubin and O'Connor (2017); O'Connor (2017a).

[3] To be clear, we are not making any claims here about the similarities of these various social categories—race is not the same as gender is not the same as religious affiliation. This said, the highly simplified models we will explore here capture broad, general aspects of social categories, and so can be applied to various cases.

labor inequitably on the basis of meaningless social identities? To address this question, we introduce a game where actors first produce some social good, and where their contributions to this production may vary. They then must divide the products of their joint labor. We model the emergence of conventions to regulate this sort of interaction in a group with social categories like gender or race.

These models allow us to pull apart unequal conventions from inequitable ones. Outcomes where two groups receive unequal levels of resource may be equitable if one side did more work in the first place. Outcomes where the two groups receive equal resources may be inequitable if one side did more work. We find that there are many stable outcomes where members of one group do more work per level of compensation. In other words, explicit inequity can emerge via cultural evolution under very minimal conditions, even between groups that are completely identical in terms of skills, preferences, etc. In fact, in most models we investigate, we find that inequity is the much more likely outcome than equity.

Notably, this is the case even when we give actors the ability to condition their demands for compensation based on the contributions made by their partners in joint labor. Our model is the first of this sort to admit this possibility. But, as we show, the ability to condition demands in this way does little to ameliorate the emergence of inequity. The take-away is that, despite stated norms of equity, outcomes like those described above, where members of one social group receive inequitable rewards for labor, should be expected to arise under minimal conditions from simple dynamics of social interaction.

The paper will proceed as follows. In section 2 we introduce the main models that will be used in this paper—variations on the Nash bargaining game that involve a production stage and a division of resources stage. We also describe previous results on the emergence of unequal conventions in models of the evolution of bargaining. In sections 3 and 4 we describe the main results of the paper. These two sections consider models where individuals do and do not condition their demands for compensation based on the level of work done by their interactive partners. As we will elaborate, both variations allow for the robust emergence of inequitable conventions. In addition, as we outline, these results are robust across various modeling choices. In section 5 we conclude.

2 The Bargaining Game and the Produce and Partition Game

The Nash demand game was introduced by Nash (1950) to represent scenarios where two individuals divide a resource (money, goods, free time), where there are different plausible divisions that are more or less preferable to each of the actors, and where highly aggressive individuals will fail to successfully agree on a division. In representing divisions of social resources, we will start with a simplified version of his model, a 'mini-game', that captures these features, and is computationally tractable.[4]

Suppose two actors divide a resource of value 10, and each can make a low, medium, or high demand, corresponding to a request for 4, 5, or 6 units of the good. Further suppose that if these requests are compatible, each actor gets what they ask for. If they over-demand the resource, though, each gets a low payoff, sometimes called the *disagreement point*, of 0. The *payoff table* of this game is pictured in figure 1. Rows represent the possible demands for player 1, and columns for player 2, while entries to the table list what payoffs each player gets for some combination of demands.

This game has three *Nash equilibria*.[5] This solution concept refers to sets of strategies where no actors can change behaviors and improve their payoff. For this reason, Nash equilibria tend to be stable in the sense that no one is incentivized to change. As we will see they also tend to be the endpoints of evolutionary processes. These equilibria

[4]See Sigmund et al. (2001) for more on the mini-game approach. There is a long tradition of using mini versions of the bargaining game in evolutionary models (Young, 1993b; Skyrms, 1994). The full models here involve a mini Nash demand game of this sort embedded in a larger game where actors first produce the resource via another mini Nash demand game.

[5]We will only worry about pure strategy Nash equilibria in this paper since they are the only equilibria relevant to evolutionary dynamics in this case.

		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: Payoff table for a mini Nash demand game.

are the strategy pairings where the actors perfectly divide the resource: Low vs. High, Med vs. Med, or High vs. Low. At these pairings, an actor who changes to demand less simply gets less, while one who tries demanding more over-demands the resource and gets nothing.

Now imagine a model where actors in a society play this game with each other repeatedly, and where these actors belong to two different social identity groups. Suppose that they are able to choose how aggressively to bargain based on the identity group membership of each partner they encounter. (For example, a latinx person might choose medium demands with other latinx people, but low demands when meeting white people.) Further suppose that over time, this group culturally evolves—individuals update their behaviors in ways that benefit themselves, so that, eventually, some stable pattern of group behavior emerges. This model might represent actors of two different races learning how to bargain over salary in the workforce, or men and women developing conventions to divide household labor.[6]

Under many dynamics—rules that model cultural change or learning—there are three outcomes that tend to emerge between groups in this sort of model, corresponding to the three equilibria. Either group A demands High and B Low, or they make medium

[6] As noted, social categories, like gender and race, are importantly different from each other, and the processes that govern interactions between these categories will be very different as well. We are working within a tradition of social modeling that privileges simplicity, tractability, and causal clarity over realism. (See Weisberg (2012) for an analysis of the trade-offs between these various modeling virtues.) This allows us to illustrate minimal conditions for the emergence of inequitable conventions in the sense outlined by O'Connor (2017c). It also allows us to apply the same models to social processes that may have different details, as long as we understand this application to be a course grained one.

demands, or group A demands Low and B High. Notice that one of these outcomes looks something like a 'fair' convention of bargaining, and the other two are discriminatory in the sense that individuals treat in- and out-group members differently to the detriment of one out-group. Indeed, starting with Axtell et al. (2000), this sort of model has been used as a bare bones representation of the emergence of discrimination and of inequality between groups.[7] The remarkable thing about this model is that we see inequality emerge endogenously among actors who simply engage in reasonable learning given their environments. This is despite any sort of justifying asymmetry between the groups in terms of skills, preferences, or starting conditions, and without any assumptions about biases, stereotypes, or the psychology of in-group/out-group interaction. Follow up results have demonstrated the robustness of this emergence of inequality to modeling choices, and have proven the flexibility of this framework to illuminate issues surrounding inequality (Poza et al., 2011; Gallo, 2014; Bruner, 2017; O'Connor and Bruner, 2017; Rubin and O'Connor, 2017).

A suggestion raised by Wagner (2012) is that the simplified Nash demand game can, with slight modifications, be taken to represent a situation where actors do not divide a windfall resource, but instead divide the fruits of joint labor. In particular, Wagner suggests that a combination of the stag hunt game (where actors choose between mutually beneficial, but risky, joint action, and risk-free solo production) and the Nash demand game can represent a case where two actors first decide whether to produce a good together and then decide who gets how much of it.[8]

The game we explore in this paper is similar to the stag hunt/Nash demand game, but adds the feature that actors can make differential contributions to the jointly produced good, rather than just deciding to opt in or opt out of it. We'll call this the produce and

[7]Their model is very similar to those introduced by Young (1993a,b).

[8]O'Connor and Bruner (2017) explore the emergence of unequal conventions between groups in this stag hunt/Nash demand combo, which they call 'the collaboration game'. Inequality emerges endogenously in this model, and is particularly likely when actors have low payoffs for hare hunting, or solo work.

partition game. We suppose that actors first play a Nash demand game to divide labor. We might think of the resource as representing free time, or time away from the joint product which can be used for personal gain. An aggressive demand, then, is a demand for a small amount of labor, while an accommodating demand represents a willingness to make a large contribution to the good produced. If actors reach the disagreement point in this part of the game, their project fails. They did not jointly contribute enough time to succeed. This choice assumes that the good is either produced or not, rather than varying in benefit based on the level of contribution as in, for example, a public goods game. While this is a simplifying assumption, it also corresponds to many realistic scenarios. In an actual stag hunt, for example, you either get the stag or you do not. In many work collaborations, you either land the big client or you don't.[9] This said, one natural extension to the work we present here is to models where greater levels of contribution correspond to a more valuable good.

If actors do enough labor to produce a good, they then have to decide how to divvy it out. This is done via a second round of the Nash demand game where the demands are now for an amount of the resource produced. Payoffs for the entire interaction then represent a combination of preferences for less work/external work in stage one and more reward in stage two.[10] Notice that even actors who do not produce a good get some payoff, from lazing around in stage one, or else from using their extra time to produce solo payoffs.

To be concrete, assume there are three levels of contribution in the first stage: Shirk, Work, and Toil. These represent small, medium, and large effort levels, respectively. Because actors are dividing labor, shirk is the aggressive demand, and generates a payoff

[9]For other cases, there are levels of contribution below which essentially no payoff is generated and above which extra effort produces small differences. In building a house, the amount of effort which makes it livable creates a large payoff, and extra effort to improve the dwelling will generate smaller surpluses. A model like the one just described is a decent match to such scenarios.

[10]To our knowledge, this kind of two-part bargaining game has never appeared in the literature before as a model for equity (evolutionary or otherwise). The closest replication might be Kazemi et al. (2017) in which participants produced a public good which could be divided unevenly. This split was determined unilaterally by a predetermined leader, however, and not via a bargaining game.

of 6. Work and Toil contributions generate payoffs of 5, and 4, respectively. The players fail to complete the project if both players shirk, or if one shirks and the other works. Otherwise they invest enough for the joint project to reach completion—generating a resource of value 10. At this point, they each make a Low, Medium, or High demand (for 4, 5, or 6) of this resource. If their demands sum to 10 or less, then the payoff for each player is the sum of their effort payoff and their demand. Otherwise, the agents cannot agree on how the resource should be split, and each walks away with only their effort payoff. Each player then has nine distinct (Contribution, Demand) strategies: (Shirk,L), (Shirk,M), (Shirk,H), (Work,L), (Work,M), (Work,H), (Toil,L), (Toil,M), and (Toil,H). Figure 2 gives the payoff table for two individuals playing produce and partition.

		Player 2								
		Shirk, L	**Shirk, M**	**Shirk, H**	**Work, L**	**Work, M**	**Work, H**	**Toil, L**	**Toil, M**	**Toil, H**
	Shirk, L	6, 6	6, 6	6, 6	6, 5	6, 5	6, 5	10, 8	10, 9	**10, 10**
	Shirk, M	6, 6	6, 6	6, 6	6, 5	6, 5	6, 5	11, 8	**11, 9**	6, 4
	Shirk, H	6, 6	6, 6	6, 6	6, 5	6, 5	6, 5	**12, 8**	6, 4	6, 4
	Work, L	5, 6	5, 6	5, 6	9, 9	9, 10	**9, 11**	9, 8	9, 9	9, 10
Player 1	**Work, M**	5, 6	5, 6	5, 6	10, 9	**10, 10**	5, 5	10, 8	10, 9	5, 4
	Work, H	5, 6	5, 6	5, 6	**11, 9**	5, 5	5, 5	11, 8	5, 4	5, 4
	Toil, L	8, 10	8, 11	**8, 12**	8, 9	8, 10	8, 11	8, 8	8, 9	8, 10
	Toil, M	9, 10	**9, 11**	4, 6	9, 9	9, 10	4, 5	9, 8	9, 9	4, 4
	Toil, H	**10, 10**	4, 6	4, 6	10, 9	4, 5	4, 5	10, 8	4, 4	4, 4

Figure 2: Payoff table for produce and partition.

In the basic Nash demand game, the equilibria are the strategy profiles where players divide the resource without waste. Given that produce and partition is akin to two Nash demand games strung together, one might expect its Nash equilibria to be the strategy sets where the two resources (effort and reward) are divided efficiently. This intuition turns out to be correct in many cases. The structure of the game is such that given another player's work contribution, the best response usually involves doing just enough work to complete the project (e.g., choosing Shirk when the other player chooses Toil).[11] And once a project is completed, a player does best to match their expectations for compensation to the demand of their partner. The Nash equilibria for produce and partition are then the strategy pairs which waste neither effort nor any of the produced good: (Shirk,M)/(Toil,M), (Shirk,L)/(Toil,H), (Shirk,H)/(Toil,L), (Work,L)/(Work,H), and (Work,M)/(Work,M), and all the flipped versions of these. (I.e., both (Shirk,M)/(Toil,M) and (Toil,M)/(Shirk,M) are equilibria. In the first, player 1 gets a preferable outcome, and in the second player 2 does. From an equity standpoint, they are equivalent.) The payoffs for these equilibria are bolded in figure 2.

As we will see, these equilibria will be the endpoints of our evolutionary models, meaning that they represent the possible, stable social arrangements between social groups. We can take these as representations of social conventions—stable patterns of behavior which might have been otherwise, but which each actor will prefer to adhere to given that the rest of the group does.[12] Axtell et al. (2000) refer to such patterns of behavior in similar models as norms, though typical accounts of norms require actors to believe that a pattern of behavior ought to be followed. This sort of belief is not captured in these simple models.

Before moving on, we would like to pull out in more detail the characters of these various possible conventions. In the first three types of equilibria, ((Shirk,M)/(Toil,M), (Shirk,L)/(Toil,H), and (Shirk,H)/(Toil,L)), one group is systematically bearing more of the work. In the middle three ((Shirk,L)/(Toil,H), (Shirk,H)/(Toil,L), and (Work,L)/(Work,H)), one group is consistently taking home more of the spoils. These unequal partitions of labor and rewards might be tolerable if they are at least equitable, however. Most academics, for instance, would agree that if one co-author does more work than the other, then he deserves the more prestigious author position. This scenario might be captured

[11] An exception to this occurs when a player expects such a low payoff from later bargaining that they do better to just slack off in round one and never complete the project at all. We will return to this possibility later.

[12] For more on the use of games to represent conventions, see Lewis (1969). See O'Connor (2017b) for more on using bargaining games to illustrate inequitable and unequal conventions in particular.

by equilibrium (Shirk,L)/(Toil,H), where one group consistently invests and reaps less and the other invests and reaps more. While such an equilibrium is unequal in the sense that one group earns less, it is at least equitable. This and the (Work,M)/(Work,M) equilibrium are actually the only equitable outcomes. The equilibrium (Work,L)/(Work,H), for instance, resembles the plight of women described in the introduction: equal work reaps unequal reward. Anyone who has participated in a group project at school is familiar with (Shirk,M)/(Toil,M): industrious students put forth more effort than lazy ones, but all group members receive the same grade. Finally, (Shirk,H)/(Toil,L) is the most inequitable of these conventions, with one party working hard to receive little pay and the other barely working to obtain a fortune. One might imagine this is the relation some CEO's have with their employees.

So we see that this game has equilibria that are inequitable in two different ways—actors work equally hard to different rewards, or actors make different contributions to equal rewards (or even to unequal rewards that do not correspond to their levels of contribution). Now we ask: do these equilibria emerge endogenously between social groups under circumstances of learning or cultural evolution? And: in such scenarios, how likely is it that we see inequity emerge?

3 The Emergence of Inequity

In the models we focus on now, as mentioned above, populations are divided into two groups that are identical modulo some arbitrary marker. We assume that actors can play produce and partition with all members of the population, and that they condition their strategies on the marker of their opponent. In other words, they choose a (Contribution, Demand) combination based on what sort of individual they interact with. The stable group level equilibria of this model will involve three conventions: one for interaction within group A, one for interaction within group B, and one for interaction between the two groups. Since we are interested in the emergence of inequitable conventions between those in different social categories, we will focus here on the between-group equilibria. These are exactly those equilibria described in the last section, but extended to an entire group. In this model, (Shirk,H)/(Toil,L) would correspond to a scenario where, for example, whenever women and men form a household women contribute more labor than men, and reap less reward. This tells us that, at the very least, the sorts of inequitable patterns described in the introduction are plausible, stable social conventions.

While the Nash equilibrium solution concept is one of the hallmarks of classical game theory, it is deficient in that is does not specify *how* these equilibria are reached. In addition, while some games like the ubiquitous prisoner's dilemma have exactly one Nash equilibrium, others may have several. Ours, in fact, has nine (of five different types). Even if we observe that the populations will always converge to the set of Nash equilibria under some dynamic, this says nothing about which of these outcomes are more or less likely to be realized.

In order to develop a model for the emergence of equity/inequity, then, we look to *evolutionary game theory*. This branch of modeling was originally developed to capture the evolution of competitive behavior in animals (Maynard-Smith and Price, 1973), but it has since found applications in the social sciences to study cultural evolution in humans. Agents in these models play a game repeatedly and update their strategy over time based on past success. This gives us a compelling story for how agents might actually move towards equilibria and gives the equilibrium selection problem some tractability. We will use this methodology to estimate which of our equilibria are likely to be reached.

Imagine our two populations locked in repeated play of produce and partition. Since, as mentioned, we focus on between-group interactions, suppose that during every round, each agent from the first population is randomly paired with an agent from the second to play the game once. For those familiar with evolutionary game theory, this is a standard two-population model. After each round, players in population 1 look around to see how the rest of the agents in their group did. Players who did worse than average will imitate the strategies of players who performed well in the previous round. This imitation is done in proportion to relative success, i.e., players who perform well below average are more likely to abandon their strategy and players who perform well above average are more likely to be copied. Players use their new strategies in the next round, and the process repeats. These are the fundamental updating rules for the discrete time *replicator dynamics*, the most common evolutionary dynamics employed in evolutionary game theory. Strategies that perform above average proliferate, while those that underachieve are gradually abandoned. While this is almost certainly an oversimplification of how strategy updating actually occurs in human groups, the replicator dynamics provide a plausible and computationally simple learning rule that represents some realistic aspects of human cultural change.[13]

One important quality of the replicator dynamics is that they are deterministic. This essentially means that there is no randomness to the dynamics. One can repeatedly start the populations in strategy distribution state p and the replicator dynamics will always carry the system to the same Nash equilibrium.[14] One method of estimating the frequency with which each equilibrium is realized, then, is to initialize our two populations repeatedly with random proportions of strategies.[15] Recording the end point at which the players arrive each time gives us an estimate of the *basins of attraction* for each equilibrium: the probability that players will settle on this convention in the

[13]Weibull (1997) shows that the replicator dynamics can be used to model cultural change via differential imitation of successful group members. This is because they are the mean field dynamics of explicit imitation learning dynamics. Lancy (1996); Fiske (1999); Henrich and Gil-White (2001); Henrich and Henrich (2007); Richerson and Boyd (2008) provide evidence that this sort of imitation occurs in real human societies.

[14]Unless the dynamics never settle at an equilibrium, but this type of outcome is beyond the purview of this paper.

[15]For the purposes of this paper, strategies are initially chosen with uniform probability over the nine previously identified. There is nothing special about this choice, however, and there may in some cases be grounds to assume some other probability distribution over initial strategy selection. A future study might investigate the potential effects of outgroup bias on equity, for instance, by increasing the probability of demanding high initially.

Table 1: Basins of attraction for produce and partition

Equilibrium	Work,M/ Work,M	Work,L/ Work, H	Shirk,H/ Toil,L	Shirk,M/ Toil,M	Shirk, L/ Toil,H
Basin of attraction	20.00%	16.64%	18.55%	33.43%	11.38%

long run given random starting places. Since there are many (in fact infinite) potential strategy initializations, we simulate a random sample of 10^8 strategy initializations and record the basins of attraction in table 1.[16] We collapse the nine possible equilibria into their five types.

Two major results can be gleamed from this table. Perhaps most immediately, the basin of attraction percentages sum to 100%, meaning that all simulations eventually approached one of the Nash equilibria. While not entirely unexpected, this result is important because it suggests that two social classes repeatedly given the opportunity to produce some shared joint good will almost always converge to some convention that allows them to do so. In the long run, one group will come to uniformly contribute a set amount and demand a set amount, as will the other group. Unilateral deviation from this arrangement will only lead to a lower payoff, so such conventions are generally hard to leave.[17]

The other notable result is the ubiquity of unequal and inequitable outcomes. As noted in section two, (Work,M)/(Work,M) is the only convention where all players put forth equal effort and reap equal rewards. The probability of arriving at this outcome is estimated at 20.00%, meaning that work and pay are not equally split in 80.00% of cases. What about equitability? Nash equilibria (Work,M)/(Work,M) and (Shirk,L)/(Toil,H) are the only perfectly equitable outcomes and have combined basins

[16]All simulations were run in Eclipse, a Java based IDE. Code for this model is available online at https://github.com/llamajones24/Inequity-sims.

[17]The Nash equilibria of the game in figure 2 are strict, meaning that an individual player who deviates from equilibrium is guaranteed a lower payoff (provided no other players deviate). This also makes each equilibrium an *evolutionarily stable strategy* profile and thus asymptotically stable under the replicator dynamics (Accinelli and Carrera, 2008). This means that states very near the Nash equilibria are drawn to them.

of attraction of 31.38%. Inequitable outcomes, then, emerge in over two-thirds of all runs and can be further broken into somewhat inequitable outcomes (Work,L)/(Work,H) and (Shirk,M)/(Toil,M) occurring 50.07% of the time and the extremely inequitable outcome (Shirk,H)/(Toil,L) emerging in 18.55% of trials. Overall these results suggest that inequity is not only possible but is in fact quite likely to emerge endogenously between two social groups interacting over time.

The payoffs used in this game are not, of course, unique. Utilities may be tailored to the particular joint project. Depending on the payoffs, it may not be worthwhile for an individual to toil if they expect Low compensation. Our current payoff table does not fully capture this situation. Consider instead the payoffs in figure 3.

		Player 2								
		Shirk, L	**Shirk, M**	**Shirk, H**	**Work, L**	**Work, M**	**Work, H**	**Toil, L**	**Toil, M**	**Toil, H**
	Shirk, L	5, 5	5, 5	5, 5	5, 3	5, 3	5, 3	8, 4	8, 6	**8, 8**
	Shirk, M	5, 5	5, 5	5, 5	5, 3	5, 3	5, 3	10, 4	**10, 6**	5, 1
	Shirk, H	5, 5	5, 5	**5, 5**	5, 3	5, 3	5, 3	12, 4	5, 1	5, 1
	Work, L	3, 5	3, 5	3, 5	6, 6	6, 8	**6, 10**	6, 4	6, 6	6, 8
Player 1	**Work, M**	3, 5	3, 5	3, 5	8, 6	**8, 8**	3, 3	8, 4	8, 6	3, 1
	Work, H	3, 5	3, 5	3, 5	**10, 6**	3, 3	3, 3	10, 4	3, 1	3, 1
	Toil, L	4, 8	4, 10	4, 12	4, 6	4, 8	4, 10	4, 4	4, 6	4, 8
	Toil, M	6, 8	**6, 10**	1, 5	6, 6	6, 8	1, 3	6, 4	6, 6	1, 1
	Toil, H	**8, 8**	1, 5	1, 5	8, 6	1, 3	1, 3	8, 4	1, 1	1, 1

Figure 3: Produce and partition with modified payoffs

In this new payoff table, payoffs for contributions are 5, 3, and 1 for Shirk, Work, and Toil investments, respectively. Similarly, 3, 5 and 7 are the available demands for resource produced. While this game retains most of the Nash equilibria from the first edition, it loses (Shirk,H)/(Toil,L), since (Toil,L) is a dominated strategy. (One can always earn higher by playing (Shirk,H).) A new equilibrium arises at (Shirk,H)/(Shirk,H), where agents from both populations invest very little into the project, which is never completed.[18] Given that (Shirk,H)/(Toil,L) was the most inequitable outcome and is

[18]While (Shirk,H)/(Shirk,H) is a pure Nash equilibrium, it is not a strict one, meaning that each

Table 2: Basins of attraction for produce and partition with modified payoffs

Equilibrium	Work,M/ Work,M	Work,L/ Work, H	Predominantly Shirk,H/Shirk,H	Shirk,M/ Toil,M	Shirk, L/ Toil,H
Basin of attraction	9.05%	4.42%	11.89%	52.17%	22.47%

no longer a Nash equilibrium, one might expect the basins of attraction for inequitability to shrink and equitability to become more likely. Inspection of table 2 reveals that this is partially true. While the basin of attraction for inequitable outcomes has indeed decreased from 68.62% to 56.59%, the roughly 12% difference has been funnelled not into the basin of attraction for equitability but instead into the non-cooperative (Shirk,H)/(Shirk,H) equilibrium's basin. The probability of ending up at an equitable outcome remains at just above 31%. Despite modifications to the payoffs and the removal of the least equitable solution, inequitability persists more often than not.

One might wonder whether there exist payoffs which consistently lead the populations to an equitable equilibrium. While a complete sweep of the payoff space would be impractical, we do simulate a variety of payoffs to test the robustness of our findings under the replicator dynamics. In particular, we vary the worth of the produced good relative to the cost of effort. As the value of the joint product increases (with fixed effort costs), the basin of attraction for equity increases and the basin for inequity decreases. The change is small, however, and inequity remains the more likely outcome.[19] Unsurprisingly, less valuable joint goods typically lead to equilibria in which all players shirk, preferring to slack off rather than produce something. We also investigate whether the particular choices for partition demands might affect our results. We find that more disparate high and low demands lead to slightly higher basins of attraction for equity.

actor can change strategies and get an equal payoff to their expected one. As a result, the system seldom converges to two populations of (Shirk,H) players but rather results in two populations made up predominantly of (Shirk,H) players with a scattering of (Shirk,L) and (Shirk,M) players. This is sustainable because contributing Low, paired with any demand, is a best response to (Shirk,H). The project is never completed and bargained over, so these strategies all earn the same payoff.

[19] Even when the joint good produced is extremely valuable (worth 400), the basin of attraction for equity is only 44%.

For instance, when the available demands are 10%, 50% and 90% of the joint good, the basin of attraction for equity is 40%. In contrast, when possible demands are 45%, 50%, and 55% of the good, the probability of arriving at an equitable outcome is 28%. This is in part because certain inequitable strategy pairs (e.g. Shirk,H/Toil,L) are not Nash equilibria when partition demands are disparate, precluding them as realizable conventions.[20] We again find that the impact of varying these payoffs is minor. In all of these simulations, the basin of attraction for equity never surpasses 44%. Thus, we have reason to suspect that inequitable conventions are probable outcomes for a large range of payoffs.

4 Inequity and Conditioned Demands

In the previous section we assumed that players select both their contribution level and demand before ever encountering their opponent. Actors make the same demand regardless of what their opponent contributes. In many cases, however, actors in real life choose demands for compensation based in part on an interactive partner's contribution. Imagine that two college roommates Amy and Brenda love throwing parties at their apartment. In addition, the apartment must be at a minimal level of cleanliness for any parties to be thrown. Both women refraining from cleaning will result in a filthy, uninhabitable apartment. If Amy spends the week scrubbing the house furiously while Brenda slacks off and plays video games, Amy may feel more deserving and might request that she get to use the apartment to party that weekend (a high demand). How might this sort of conditional demand framework affect the probability of reaching an equitable convention? Do actors resolve to reward those who work hard, conforming to our intuitions about equitability? Or is inequity robust across these models?

[20] In addition, other authors note a similar pattern for evolution in the Nash demand game (O'Connor, 2017b). There are fewer states where populations move towards low demands when those demands are very low.

Table 3: Basins of attraction for conditional produce and partition

Equilibrium	Work,M/ Work,M	Work,L/ Work, H	Shirk,H/ Toil,L	Shirk,M/ Toil,M	Shirk, L/ Toil,H
Basin of attraction	31.44%	4.33%	5.83%	46.19%	2.97%

A major difference in this conditional strategy framework is that player strategies have four parts: (Contribution, Demand vs Shirking opponent, Demand vs Working opponent, Demand vs Toiling opponent). A player using strategy (Shirk, M, L, H) would Shirk, demand Medium if her opponent Shirks, demand Low if her opponent Works, and demand High if her opponent Toils. There are a total of $3^4 = 81$ distinct strategies of this form, so we omit the payoff matrix here. Using the original contribution payoffs (6, 5, 4) and demands (4, 5, 6), one might expect that the increase in strategies will lead to an increase in Nash equilibria. While this is true, most of the equilibria are equivalent to one of the original five from section 3: (Shirk,M)/(Toil,M), (Shirk,L)/(Toil,H), (Shirk,H)/(Toil,L), (Work,L)/(Work,H), and (Work,M)/(Work,M).[21]

Inspecting the basins of attraction for the conditional produce and partition in table 3 yields two final results. The first is that inequitable outcomes still occur with high probability (56.35%). This phenomenon has manifested itself throughout all variations of produce and partition considered in this paper. The other result is that the Nash equilibria basins of attraction in table 3 only sum to about 91%. To illustrate what happens the other 9% of the time, imagine an employee and employer running a small business. Both must put forward some minimum amount of effort at work for the business to stay afloat. An employee who invests the minimum amount of effort for the business to succeed will walk home with their promised paycheck. An employee who goes above

[21]Consider two strategies (Work, L, L, L) and (Work, H, L, H) composing population 1, while population 2 plays (Work, L, H, M) uniformly. Although the strategies (Work, L, L, L) and (Work, H, L, H) in population 1 are technically distinct based on their demands, observe that both strategies implement contribution level "work" and both of them only ever demand low, since their population 2 opponents uniformly contribute medium. This state looks like equilibrium (Work,L)/(Work,H), though it is composed of many technically distinct strategies.

and beyond to excel at her tasks may warrant a bonus from the employer who hopes to incentivize the employee's good work (and not lose the employee to another company).

This is the essence of what occurs in these 9% of runs. One population is split between two strategies: one investing smaller and demanding smaller, the other investing larger and demanding larger. The second population discerns these discrepancies in contribution and adjusts their demands for resources accordingly for each population 1 agent encountered. They always work, so that the good is produced, and then demand less resource from harder workers and more from slackers. While these cases are relatively rare, one might classify them as among the more equitable outcomes: greater effort is being rewarded with greater spoils.

4.1 Agent Based Modeling and Robustness

To this point, we have explored only infinite population models where change is represented via the replicator dynamics. As we have pointed out, the replicator dynamics are often used successfully as a model of cultural imitation. But there are other aspects of human cultural change. We now briefly discuss results from a model that makes very different assumptions about population size and dynamics. Of course, no model that is simple enough to analyze will capture all the relevant ways that humans update their strategic behavior. Instead, our goal is to provide a robustness check on our results. We do this by altering a number of features of the model, but still maintaining the key elements meant to correspond to our target systems. If we see similar results emerge, we gain confidence that our results are not a relic of our modeling choices.

In particular, we simulate agent-based models with finite populations. The paradigm we use is adapted from one employed by Young (1993b) to study the evolution of bargaining between groups, and developed by Axtell et al. (2000) to explore the emergence of inequitable conventions. Since then, it has been used by several authors to explore the emergence of inequity and inequality (Gallo, 2014; O'Connor, 2017a; LaCroix and

O'Connor, 2018). Our version of this model assumes that each round one agent is chosen from each group to play the produce and partition game. Each agent has a finite string of memories of past play by opponents.[22] They use these memories to choose strategies. In particular, each agent chooses whatever strategy would yield the highest payoff on average if played against their memories, or *best responds* to their memories. This is a boundedly-rational response rule. Agents make calculations about what strategy will work best, on the assumption that their memories reflect the true distribution of opponent strategies. This is arguably more cognitively complex than simply imitating successful group members. But agents do not have infinite memories, and neither do they make complicated inferences about how their opponents might behave.[23]

Because our agents play a produce and partition game, we actually model them as having four strings of memories. The first corresponds to memories about past production strategies of opponents. The next three correspond to conditional memories about what partition demands opponents made after contributing some amount to production. Strategy choices are made by first calculating a best response for production.[24] Then, if agents contribute enough to produce the resource, they consult their conditional memories of how agents made partition demands based on their production contributions, and best respond.

This model has a few parameters. The first is group size, and the second memory length (we assume equal memory length for each of the four strings). The last parameters we can vary are the particular levels of contribution and demand in the two stages of the game. Simulations of these models end up at the same, stable equilibria that simulations of the replicator dynamics do. As with our previous models, we find that all five equilibria

[22]We start agents with no memories, and have them select their first memory randomly.

[23]This version of the learning rule is the one employed by Axtell et al. (2000), with the small variation that we start with no memories rather than random strings. Young (1993b), on the other hand, considers a group of agents who share recent memories and best respond to a random sample of these memories.

[24]In order to make this simple, we constrain payoffs so that it is always worthwhile for agents to produce the resource. This means that they do not have to take the partition stage into account in calculating best responses.

Table 4: Outcomes for an agent based model of conditional produce and partition

Equilibrium	Work,M/ Work,M	Work,L/ Work, H	Shirk,H/ Toil,L	Shirk,M/ Toil,M	Shirk, L/ Toil,H
Basin of attraction	34.44%	22.87%	3.59%	22.02%	17.08%

arise with significant probability. To give an example, table 4 shows the likelihood of each equilibrium for a population of size 10, memory length 10, and both contribution payoffs and demands of 4, 5, and 6.[25] Equitability emerges in 51% of simulations, and inequitable outcomes the rest of the time.

Varying the parameters influences the probability that each outcome emerges. In particular, increasing the size of the population makes the Work,M/Work,M outcome increasingly likely. Smaller populations are more likely to end up at inequitable, and unequal, outcomes. Making the Low and High demands of the partition game more disparate increases the likelihood of outcomes where actors partition fairly.[26] But, in general, across parameter values, we again find that the probability of inequity emerging in this model is always high, even though actors condition demands on the production inputs of their partners.

5 Conclusion

One might wonder at this point: given the robustness with which inequitable conventions of various sorts emerge in these simple models, how do we explain the prevalence of equity norms? Remember that the models we have considered always involve a population divided into two groups or social categories. Things turn out differently in a group without these sorts of divisions.[27] If we consider a single population, the symmetric,

[25]Code for this model is available online at https://github.com/cailinmeister/inequityinequality. We ran simulations for population sizes 4-50, and memory length 10-15 to check robustness.

[26]We did not make the demands more disparate for the production game to avoid cases where actors were not always incentivized to produce, as noted.

[27]By this we simply mean a one population model.

equitable outcome (Work,M)/(Work,M) emerges 60% of the time. Another 27.4% of the time, all agents end up demanding medium in the end, and some shirk while others toil. The rest of the time, a number of equilibria emerge that involve a mixture of shirkers and toilers, and different demands. Notice that at these equilibria, though, despite the fact that individual interactions will be inequitable, it is nonetheless the case that all actors have the same expected payoffs. Otherwise, of course, they would not be equilibria in a single population.

Why does the simple addition or subtraction of categories from the model so radically alter the cultural evolution that occurs? One way to explain it appeals to symmetry and symmetry breaking. In the contribution part of the model, actors are most efficient if they perfectly divide the labor. (I.e., no extra work is done to produce the good, and the opportunity to create a joint surplus is taken advantage of.) In a population without categories, the only way to guarantee that every pairing of actors will efficiently divide labor is for everyone to choose contribution level "work". Otherwise, sometimes shirkers will meet each other and fail to produce the good, and sometimes toilers will meet each other and put too much work into the project. With categories, actors can use category membership as a symmetry-breaking mechanism. There is an extra piece of information in interactions between those in different categories (i.e., I am type A and you are type B), which allows them to efficiently break symmetry with respect to contributions (type A always contributes more, type B less).[28]

This same sort of reasoning applies to bargaining over rewards from joint labor. With one exception—when it comes to this stage of the process, information about earlier contributions can be used as a symmetry breaker. But, of course, a dependence

[28]See O'Connor (2017b) for an in-depth discussion of this sort of symmetry breaking. Similar reasoning can be applied to work in philosophy of science explaining the emergence of fairness norms. Skyrms (1994); Alexander and Skyrms (1999); Skyrms (2014); Alexander (2007) show that in simplified Nash demand games the equal outcome is special from an evolutionary point of view. Because it is the only symmetric outcome, it is more likely to evolve. In contrast, when Axtell et al. (2000); Bruner (2017); Bruner and O'Connor (2015); O'Connor and Bruner (2017); Rubin and O'Connor (2017) add categories to the same sort of model, the fairness norm is no longer special because the categories break symmetry between actors of different types.

between contribution and reward is what we expect from an equitable convention.[29] When we have two different categories of actors, there is a symmetry breaker available at the reward stage that has nothing to do with contribution. In other words, once we get to the second stage, there is no particular reason to coordinate reward based on contribution compared to coordinating reward based on irrelevant group membership.

This is perhaps the central insight of the paper. From a standpoint where we think of conventions and norms as facilitating social coordination, equity is special in groups where everyone is the same, but its specialness disappears as soon as any sort of further social information in the form of social category membership is added. What we see here is that it is quite easy to evolve conventions that do not involve equitable divisions of jointly produced social resources if we have groups divided into categories.

Acknowledgements

This research was supported by NSF grant 1535139. We would also like to thank the members of the Social Dynamics Seminar at UC-Irvine for their insightful comments. And thanks to the audience at the 2018 PSA meeting for comments and feedback. Finally, we appreciate the feedback from three anonymous referees which helped improve the quality of this paper.

Bibliography

Accinelli, Elvio and Edgar J. Sánchez Carrera (2008). "On evolutionarily stable strategies and replicator dynamics in asymmetric two-population games." *Dynamics, Games*

[29]In the models just described, we do not see this because actors have no good way to coordinate their first stage contributions besides demanding medium. We also do not shape payoffs so that there is enough fine-grainedness to get exactly equal payoffs through different contribution/reward combinations, which is necessary to reach equilibria in this sort of model.

and Science I. Ed. David A. Rand Mauricio Matos Peixoto, Alberto Adrego Pinto. Springer, Berlin, Heidelberg.

Adams, J Stacy (1963). "Towards an understanding of inequity.." *The Journal of Abnormal and Social Psychology*, *67*(5), 422.

Adams, J Stacy and Sara Freedman (1976). "Equity theory revisited: Comments and annotated bibliography." *Advances in experimental social psychology*, *9*, 43–90.

Alexander, J McKenzie (2007). *The structural evolution of morality.* Cambridge University Press Cambridge.

Alexander, Jason and Brian Skyrms (1999). "Bargaining with neighbors: Is justice contagious?." *The Journal of philosophy*, *96*(11), 588–598.

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

Bianchi, Suzanne M, John P Robinson, and Melissa A Milke (2006). *The changing rhythms of American family life.* Russell Sage Foundation.

Bruner, Justin P (2017). "Minority (dis)advantage in population games." *Synthese*, doi 10.1007/s11229–017–1487–8.

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.

Casper, Lynne M, Sara S McLanahan, and Irwin Garfinkel (1994). "The gender-poverty gap: What we can learn from other countries." *American sociological review*, 594–605.

Coltrane, Scott (2000). "Research on household labor: Modeling and measuring the social embeddedness of routine family work." *Journal of Marriage and family*, *62*(4), 1208–1233.

Feldon, David F, James Peugh, Michelle A Maher, Josipa Roksa, and Colby Tofel-Grehl (2017). "Time-to-credit gender inequities of first-year PhD students in the biological sciences." *CBE-Life Sciences Education*, *16*(1), ar4.

Fiske, Alan Page (1999). "Learning a culture the way informants do: Observing, imitating, and participating." *Unpublished manuscript, University of California, Los Angeles.*

Freedman, Sara M and John R Montanari (1980). "An integrative model of managerial reward allocation." *Academy of Management Review*, *5*(3), 381–390.

Gallo, Edoardo. Communication networks in markets. Technical report, Faculty of Economics, University of Cambridge, (2014).

Henrich, Joseph and Francisco J Gil-White (2001). "The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission." *Evolution and human behavior*, *22*(3), 165–196.

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

Kazemi, Ali, Daniel Eek, and Tommy Gaerling (2017). "Equity, Equal Shares or Equal Final Outcomes? Group Goal Guides Allocations of Public Goods." *Frontiers in Psychology*, *8*(11).

LaCroix, Travis and Cailin O'Connor (2018). "Power by Association.".

Lancy, David F (1996). *Playing on the mother-ground: Cultural routines for children's development.* Guilford Press.

Lewis, David (1969). *Convention: A philosophical study.* John Wiley & Sons.

Maynard-Smith, J and George R Price (1973). "The logic of animal conflict." *Nature*, *246*(5427), 15–18.

Nash, John F (1950). "The bargaining problem." *Econometrica: Journal of the Econometric Society*, 155–162.

O'Connor, Cailin (2017a). "The Cultural Red King Effect." *The Journal of Mathematical Sociology, doi: 10.1080/0022250X.2017.1335723.*

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

O'Connor, Cailin (2017c). "Modeling Minimal Conditions for the Emergence of Inequity." *Working Paper.*

O'Connor, Cailin and Justin P Bruner (2017). "Dynamics and Diversity in Epistemic Communities." Working Paper.

Poza, David J, Félix A Villafáñez, Javier Pajares, Adolfo López-Paredes, and Cesáreo Hernández (2011). "New insights on the Emergence of Classes Model." *Discrete Dynamics in Nature and Society, 2011.*

Pressman, Steven (2002). "Explaining the gender poverty gap in developed and transitional economies." *Journal of Economic Issues*, *36*(1), 17–40.

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

Rubin, Hannah and Cailin O'Connor (2017). "Discrimination and Collaboration in Science." (Working Paper).

Sigmund, Karl, Christoph Hauert, and Martin A Nowak (2001). "Reward and punishment." *Proceedings of the National Academy of Sciences*, *98*(19), 10757–10762.

Skyrms, Brian (1994). "Sex and justice." *The Journal of philosophy*, *91*(6), 305–320.

Skyrms, Brian (2014). *Evolution of the social contract.* Cambridge University Press.

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

Treas, Judith and Sonja Drobnic (2010). *Dividing the domestic: Men, women, and household work in cross-national perspective.* Stanford University Press.

Bos, KeesVan den (2001). "Fairness heuristic theory." *Theoretical and cultural perspectives on organizational justice*, *63*.

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 (2012). *Simulation and similarity: Using models to understand the world.* Oxford University Press.

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

Young, H Peyton (1993a). "The evolution of conventions." *Econometrica: Journal of the Econometric Society*, 57–84.

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