Power by Association

Travis LaCroix, Cailin O'Connor

Department of Logic and Philosophy of Science
University of California, Irvine

Abstract

We use tools from evolutionary game theory to examine how power might influence the cultural evolution of inequitable norms between discernible groups (such as gender or racial groups) in a population of otherwise identical individuals. Similar extant models always assume that power is homogeneous across a social group. As such, these models fail to capture situations where individuals who are not themselves disempowered nonetheless end up disadvantaged in bargaining scenarios by dint of their social group membership. Thus, we assume that there is heterogeneity in the groups in that some individuals are more powerful than others.

Our model shows that even when most individuals in two discernible sub-groups are relevantly identical, powerful individuals can affect the social outcomes for their entire group; this results in power *by association* for their in-group and a bargaining disadvantage for their out-group. In addition, we observe scenarios like those described where individuals who are *more* powerful will get less in a bargaining scenario because a convention has emerged disadvantaging their social group.

1. Introduction

Not so long ago in America, it was a rule that black people sit in the back of the bus and white people in the front. Today, women tend to be paid less than men, on average, for the same work, even when the data is adjusted for external factors that may exacerbate such a discrepancy.[1] These are both cases where patterns of behaviour related to the division of resources disadvantage those in one social group compared to those in another.

Email addresses: tlacroix@uci.edu (Travis LaCroix), cailino@uci.edu (Cailin O'Connor)

[1] Moyser (2017) points out that when we consider annual earnings, women make \$0.74 for every dollar earned by men. However, this comparison is confounded by discrepancies in work hours between genders. However, even accounting for this, women earn \$0.87 for every dollar of their male counterparts. See also Baker and Drolet (2010).

In investigating such phenomena, economists have often employed the *Nash demand game*—a simple, game-theoretic model intended to capture situations where two people divide a resource (Nash, 1950). Some of the earliest work on this model showed how power might play a pivotal role in determining who gets more and who gets less in a bargaining scenario. In particular, under Nash's influential axiomatic 'bargaining solution', an actor with more power (we will say more shortly about what this means) is expected to get more of a resource (Nash, 1953). This observation has been taken to inform inequities between social groups like those just described. For example, applying the Nash bargaining game to household bargaining, economists have argued that when women tend to be in a worse economic position in the case of divorce (i.e., if bargaining over who does what in the household fails), they should also be expected to do more work in the household (Manser and Brown, 1980; McElroy and Horney, 1981).

In many cases, though, individuals who are personally powerful nonetheless end up disadvantaged in bargaining scenarios by dint of belonging to oppressed social groups. In the case of household bargaining, for example, it has been widely noted that even women with earnings equal to or greater than their husbands tend to do more household labour (Horne et al., 2017). This seems to be because divisions of resources are not determined solely by bargains between individuals, but also by social conventions and norms—society-wide patterns that specify which types of people get more and which less in various scenarios.[2] To understand these divisions, then, we need to look at the processes by which such conventions and norms might emerge in human societies.

In this paper, we examine how power might influence the cultural evolution of inequitable norms between discernible groups, such as gender or racial groups, in a population of otherwise identical individuals. In this respect, our work extends that of Bruner and O'Connor

[2]We will say more later about what we mean by conventions and norms here. O'Connor (2018) discusses the conventionality of inequitable divisions of resources at greater length.

(2017), who investigate this very question. Their models, though, always assume that power is homogeneous across a social group—i.e., if women are disempowered with respect to household bargaining, for example, then they are all disempowered and all in the same way. This simplification means that their models fail to capture situations like the puzzling ones just described—where individuals who are not themselves disempowered nonetheless end up disadvantaged in bargaining scenarios by dint of their social group membership.

We use tools from evolutionary game theory to build agent-based models where members of different social groups learn and culturally evolve to divide resources. We suppose that there is heterogeneity in the groups in that some individuals are more powerful than others. What we find is that even a single powerful member of a social group will make it more likely that every member of that group ends up advantaged with respect to bargaining conventions. Thus, our model shows that even when most individuals in two discernible sub-groups are relevantly identical, powerful individuals can affect the social outcomes for their entire group; this results in power *by association* for their in-group and a bargaining disadvantage for their out-group. In addition, we observe scenarios like those described where individuals who are *more* powerful will get less in a bargaining scenario because a convention has emerged disadvantaging their social group. These models show that when thinking about the effects of power on bargaining, it is crucial to consider not only the impact of power on the positions of two bargainers, but also the impact of power on the conventional positions of two social groups. As we point out in the conclusion, both aspects seem relevant to the emergence of real-world bargains.

The paper proceeds as follows. Section 2 outlines relevant previous work in the field and lays the technical groundwork for discussion of the models examined herein. Section 3 describes the current model in detail and presents original results with respect to bargaining conventions and the cultural evolution of inequitable norms. Section 4 offers an interpretation and discusses the relevance of these results.

2. Background and Previous Results

As mentioned, a standard way of modelling scenarios where individuals divide resources is with a Nash demand game. This game involves two players who have some resource to divide. Each player makes a demand as to how much of the resource she wants for herself.[3] Suppose, for example, that the resource available is of value 10. If the players make compatible demands—i.e., demands whose sum is at most 10—then both players receive exactly what they demanded. However, if their demands are incompatible—i.e., if they sum to more than 10—then both players receive some poor, though possibly nonzero, payoff called a *disagreement point*. The idea is that if players are jointly too aggressive, they fail to come to an agreement on how to split the resource. We will say more about the disagreement point concept below.

In the case where the resource is infinitely divisible, there exist for this game a continuum of *Nash equilibria*—i.e., strategies where no player can switch and improve her payoff—where the two demands sum to exactly the totality of the resource. To clarify, if have an infinitely divisible 10-unit resource the Nash equilibria are comprised of every outcome where the demands of the two players sum to 10. Notice that if either player demanded more at such an outcome, the total demands would exceed the resource, so they would both get the disagreement point. If either player demanded less, that player would simply get less.

2.1. Power, Bargaining, and Rational Choice

Nash (1950), working in the rational-choice based paradigm of classical game theory, proposed an influential solution to this bargaining problem. Nash's solution is based on four axioms that he thought any solution should satisfy:

1. Pareto efficiency,

[3] Introduced by Nash (1950, 1953), the demand game may also be referred to as a *Nash bargaining game*, *Divide-the-Dollar*, or *Divide-the-Pie*.

2. Symmetry,
3. Invariance to affine transformations, and
4. Independence of irrelevant alternatives.[4]

Pareto efficiency is the condition that no player can make herself better off without making her opponent worse off. That is we do not expect inefficient outcome where players leave behind some of the resource. Symmetry states that if the players are indistinguishable, then the solution should not discriminate between them—i.e., when the players' positions are symmetric, their payoffs should be symmetric. As we will see, this condition is most relevant to our discussion here. Invariance to affine transformations assumes that an affine transformation of the payoffs and disagreement point should not affect the outcome of bargaining.[5] Finally, the independence of irrelevant alternatives (to be a bit hand-wavy) stipulates that if a solution, x, is chosen from a set A, and there is some other set, B, such that $x \in B \subset A$, then that solution must be chosen from the set B—i.e., any solutions in the set-theoretic difference, $A - B$, are irrelevant to bargaining. We might say that if a bargain is made, getting rid of some options that were not chosen should not change the bargain.

The solution derived from these axioms stipulates that for payoffs to the two players, u_1 and u_2, and disagreement points d_1 and d_2, the players maximize $(u_1 - d_1)(u_2 - d_2)$. In other words the solution expects players to maximize the product of their payoffs minus the disagreement point for each player. Notice that when the players have the same disagreement points, this solution will yield equal payoffs.

Nash (1953) re-interpreted the disagreement point in these models to correspond to an issued threat about what would happen should bargaining fail. This is the sense in

[4]Note, if we replace this axiom with the so-called "resource-monotonicity axiom" (which says that no player should be strictly worse off from an increase in the available resource), this new set of axioms gives rise to the Kalai–Smorodinsky bargaining solution. See Kalai and Smorodinsky (1975). See also Moulin (2004).

[5]This sort of transformation involves multiplying all the payoffs for an individual by the same number, or adding a constant to them, or both.

which the disagreement point was first taken to correspond to the power of an individual bargainer. Whoever has the ability to issue a more credible threat, based on their personal situation, can lower their opponent's disagreement point further and reap the benefits in the subsequent bargain. Even without this threat interpretation, though, the disagreement point captures something relevant about the power of an individual—those with more secure fall-back positions are in a better, more powerful place with respect to bargaining in general. They do not care as much about the bargain succeeding and can use this to their advantage. For this reaons, in the models we present, we will operationalize power using differences in disagreement points. This should not be taken as a claim that this is the only way to model power, or the only type of power that matters to bargaining, or even the most important one. It is simply one way to capture an aspect of what it means for bargainers to be more or less powerful.

As briefly mentioned in the introduction, in the influential work of Manser and Brown (1980) and McElroy and Horney (1981) (and many subsequent economists), household bargaining is modelled using a Nash demand game.[6] Household bargaining involves the division of leisure and market or household labour within a household, subject to constraints of total available time. These authors use the Nash solution to predict that women will do more household labour.[7] On their model, u_i is interpreted as marital utility, which is itself a function of home and market goods and leisure time, and d_i is the disagreement point, which is interpreted as payoff in the event of a divorce—i.e., the situation under which bargaining breaks down. This disagreement point is a function of wage rates, household productivity parameters, opportunities outside the marriage (such as remarriage), and other "extrahouse-

[6]These models supersede earlier "common preference" approaches to family economics that treat the household as a 'black box', with income going in and leisure, goods, and children coming out. (See Samuelson (1956) and Becker (1974, 1981).) These earlier models look at household bargaining as an issue of maximizing social welfare and so cannot separate out (possibly disparate) individual utilities for the individual members of the household.

[7]Note, this literature typically focuses on heterosexual households.

hold environmental parameters" (McElroy, 1990). In this sense, individual assets—such as personal wealth, property, or earning ability—determine one's disagreement point insofar as they are linked to one's expected situation when on one's own (Sen, 1982).

Lundberg (2008) suggests that there are several proximate causes of inequity in disagreement points of this sort. These include the fact that women, in general, have lower market wages than men, and so their post-divorce earnings are poorer; women often have primary custodial responsibilities and so must share their earnings with children; and women tend to have worse remarriage prospects relative to divorced men. As such, women's disagreement points are reduced significantly relative to the men's.

Hence, on these models, rational choice predicts asymmetric household labour distributions between men and women in a domestic partnership. When men have higher disagreement points this translates to more bargaining power, and thus to less household labour and more leisure. From this point of view, we should expect that when both partners in a household have an equal degree of bargaining power, they will divide household labour evenly. That is to say, from a rational-choice perspective, symmetric bargaining positions should result in symmetric bargaining outcomes. (In fact, remember that this was one of Nash's axioms.) However, as noted, empirical data shows that women tend to do more household labour than men even when their market compensation is equal to or higher than that of the male counterpart, contra predictions of the Nash model (Horne et al., 2017).[8]

2.2. Evolutionary Game Theory and the Emergence of Inequity

Evolutionary game theory, in contrast with classical game theory, looks at the emergence or evolution of strategic behaviour on a group level, rather than at the level of the rational

[8]Several different models have been proposed to try to account for this discrepancy. We will not examine these in detail here. See Lundberg and Pollak (1996) for an overview of several different economic models of household bargaining in more detail. Several of these are also outlined by O'Connor (2018).

choices of individuals.[9] When it comes to the cultural emergence of bargaining behaviour in particular, this framework has had a number of explanatory successes. For instance, Skyrms (1994, 1996) uses an evolutionary model of actors playing the Nash demand game to explain how a concept of "justice" might evolve. The state wherein an entire population always demands half of a resource is *evolutionarily stable*—i.e., it cannot be invaded by a mutant strategy. This "fair" division is the most common outcome in many evolutionary models.[10] Thus we see why norms for justice and fairness might be so common in human societies.

These evolutionary models can also show how inequitable conventions and norms might emerge in a group. Economists and philosophers of science have decisively demonstrated how once a population is broken into social groups, equity is no longer the expected outcome of such models.[11] Instead, inequitable outcomes, where members of one group get more and the other group less, tend to emerge endogenously. Since these types of models will form the basis of the work discussed here, we will now describe them in detail and discuss relevant results.

To investigate the evolutionary outcomes of bargaining, previous authors have looked at simplified versions of the Nash demand game, as will we. Assume that the resource is always of value 10. Assume that an individual can make one of three possible demands corresponding to a *low* (L), *medium* (M), and *high* (H) amount of the resource, respectively. Assume also that our medium demand corresponds to what we might call an equitable or fair demand. In general, for a resource of value R, we have $0 \leq L < R/2$, $M = R/2$, and $R/2 < H \leq R$. Thus, in our case we have $L < 5, M = 5$, and $H > 5$. In addition, we will assume that L and H are compatible in that they sum to 10 (1 and 9, or 4 and 6, for

[9] These models can be applied to either biological or cultural evolution. For example, there are clear examples in nature where animals appear to play bargaining games. For example, in two-male stallion (*Equus caballus*) coalitions something like a Nash demand game appears to be played to divide mating successes (Feh, 1999).

[10] See also, Ellingsen (1997); Binmore (1998, 2005); Young (1993a); Alexander (2000).

[11] See Axtell et al. (2001); Poza et al. (2011, 011b); Gallo (2014); O'Connor (2017); O'Connor and Bruner (2017); Rubin and O'Connor (2018); Bruner and O'Connor (2017); O'Connor (2018).

example).

In this simplified three-demand Nash bargaining game with a resource of value 10, let us represent the disagreement point for Players 1 and 2 as D and d, respectively. The disagreement points, D and d, are what the players can expect to receive in case bargaining breaks down. Complete information about this game is shown in Table 1. Each entry shows the payoffs for a combination of strategies with player 1's payoff listed first. Note that if

		Player 2		
		L	M	H
	L	L, L	$L, 5$	$\mathbf{L}, \mathbf{H}$
Player 1	M	$5, L$	$\mathbf{5}, \mathbf{5}$	D, d
	H	$\mathbf{H}, \mathbf{L}$	D, d	D, d

Table 1: Simplified Nash Demand Game for 10 Unit Resource

we only consider cases where $D, d < L$, this game has three Nash equilibria.[12] These are highlighted in bold in Table 1 and correspond to the situations in which Player 1 demands low and Player 2 demands high, Player 1 demands high and Player 2 demands low, or both players demand medium. As in the game with a continuum of strategies, demanding more at any of these equilibria would lead to the disagreement point, and demanding less would lead to less. As we will see, in our evolutionary models these three outcomes will be the ones that emerge between social groups.

Besides the game, the other element of an evolutionary game-theoretic model is the dynamics—a set of rules for determining how the strategies of actors in a population will update. This typically occurs over time steps, such that, at each time step, the dynamics determine how the population changes based on assumptions about how the actors learn or evolve. The models we present use a dynamics employed by Axtell et al. (2001), who, as noted, investigate the emergence of inequitable or discriminatory norms. Their model is

[12]We will only consider pure strategies—those where actors always take the same action rather than probabilistically mixing—since these are the relevant ones for an evolutionary analysis.

based upon previous evolutionary models of bargaining from Young (1993a,b). We focus on the work of Axtell et al. (2001) since their version of the model is closest to our own.

Assume a finite population, consisting of N individuals. There are two sub-populations, A and B, of discernible "types". These represent identifiable social groups such as men and women, or white and black people.[13] Each population consists of n_A and n_B individuals such that $n_A + n_B = N$. Following Axtell et al. (2001), we always consider models where $n_A = n_B$, or the groups are equally sized.[14] On each round of play, two agents are chosen to randomly interact with one another. They play the Nash demand game with possible demands H, M, and L, as presented in Table 1. Further, each agent is equipped with a finite memory of length m. An agent's memory consists in the last m strategies that she has encountered. On the basis of this each agent develops a belief—possibly inconsistent with the actual state of the world—about what her opponent will do next time. She chooses her strategy as a best response to this limited memory under the assumption that the probability that her partner will demand H, M, or L is equivalent to the relative frequency of each of these demands' occurrences in her memory. As such, her response is a type of boundedly rational best response to her past experiences. When she is indifferent between strategies, she randomizes.[15]

Axtell et al. (2001) point out that when the agents are entirely symmetric—i.e., when there are no subgroups—the equity norm, wherein every agent in the population demands

[13]For more on this modelling choice and on the fit between such models and real social categories like gender and race see O'Connor (2018).

[14]Bruner (2017) shows, using slightly different models, that when $n_A \neq n_B$, inequitable norms can arise against the minority population solely by dint of the fact that they are indeed a minority. O'Connor (2017) shows that these results are robust under the type of dynamics described here, and O'Connor (2018) provides an overview of these results. This is due to, what Bruner and O'Connor refer to as, the cultural red king effect. This is similar to the red king effect in biology where slower evolving species can gain an adaptive advantage over quickly evolving species (Bergstrom and Lachmann, 2003).

[15]Axtell et al. (2001) further assume that, with some small probability, ϵ, the agent chooses her strategy at random. This small ϵ may be understood as error or experimentation. The addition of an error-rate makes the agents' behaviour a "noisy" best reply to remembered past interactions. We will not include this sort of stochasticity in our versions of these models.

M, is the unique stochastically stable equilibrium (SSE) of the model.[16] This result supports the results of Skyrms (1994, 1996), Binmore (1998, 2005), etc., regarding the evolution of fairness.[17]

Axtell et al. (2001) then test the model with subgroups, which they call "yellow" and "blue". In this version, agents maintain separate memories for what individuals in the two different groups have done in the past, and they condition their behaviour dependent upon whether they are paired with a member of their own group or a member of the other group. As these authors observe, there are three equilibria that emerge between groups in this model, corresponding to the three Nash equilibria of the underlying game. Either the two groups demand medium of each other, one side demands high and the other low, or the first side demands low and the other high.

They take these equilibria to represent the emergence of "norms" in their populations. While these models are probably too simple to capture important relevant aspects of normative behaviour in humans, we do take them to be good representations of social conventions in a similar sense to that outlined by Lewis (1969). We see patterns of behaviour that improve social coordination, that are stable over time, and that are self-reinforcing—i.e., where no individual wants to switch what she is doing given what the group is doing. In particular, notice that the two equilibria in these models where one side demands high and one low can represent something like discriminatory conventions. Individuals treat in- and out-group members differently, to the detriment of one group.

[16] The SSE concept was introduced by Foster and Young (1990). It makes a prediction for play of a game by looking at where an evolutionary dynamics spends its time as the mutation rates in the population go to 0.

[17] In spite of the fact that, asymptotically, the equity norm is significantly more likely than any other outcome, Axtell et al. (2001) show that populations without subgroups may nonetheless spend a significant amount of time in a "fractious" state, wherein some agents demand high and some demand low. This is regardless of the fact that the fractious state is inefficient insofar as a different combination of demands would result in the population being better off on average: The average expected payoff for this fractious state is half of that of the equitable state. The point of their one-population model is that, in spite of the fact that equity is probable in the long run, inefficient divisions may persist for significant periods of time. They show that this waiting time is exponential in memory length and population size.

It is by showing that these sorts of outcomes regularly arise in models with social groups that previous authors have addressed the cultural emergence of inequitable norms and conventions. In spite of the fact that their groups have no inherent social significance, and the individuals in each group are otherwise identical, Axtell et al. (2001) find that some meaningless difference in population type (yellow vs. blue) can give rise to inequity. Further, these conventions arise because of an "accident of history" which becomes reinforced over time. As such, over time the tags come to bear social significance, and the population is divided into a class system that depends upon these (now) discriminatory tags. Since these sorts of processes are self-reinforcing, such patterns continue in spite of the fact that they are initially randomly established—i.e., they have no *a priori* justification.

The robustness of these outcomes has been widely verified. Under different choices of dynamics and different population structures, groups of actors who are of different types, play a Nash demand game, and update their strategies are commonly seen to evolve toward this sort of inequitable arrangement (Poza et al., 011b; Gallo, 2014; Bruner, 2017; Bruner and O'Connor, 2017; O'Connor, 2017; O'Connor and Bruner, 2017; Rubin and O'Connor, 2018).

2.3. Power, Bargaining, and Evolution

Bruner and O'Connor (2017) use a framework much like that employed by Axtell et al. (2001) to investigate how power for one social group might lead to an advantage with respect to the emergence of bargaining conventions. In particular, they frame their model in the context of academic research where the two populations are asymmetric with respect to bargaining power—e.g., junior professors bargaining with tenured professors, or graduate students bargaining with faculty. However, the results are general to bargaining populations with subgroups (O'Connor, 2018).

These authors operationalize power in several ways, including investigating the influence of disagreement points on the emergence of bargaining conventions.[18] They suppose that all members of one group have a disagreement point D, while the other have disagreement point d. When $D > d$, this corresponds to individuals in the one population having more bargaining power than the other insofar as these individuals have less to lose when bargaining breaks down. They show that this addition of power to their model gives rise to a bargaining advantage for individuals in the "powerful" population insofar as it is more likely that the population evolves to a convention where the powerful group demands high.

Bruner and O'Connor (2017) focus on infinite population models using the *replicator dynamic*—a very common dynamic from evolutionary game theory which supposes that strategies yielding above average payoffs proliferate while those that yield below average payoffs diminish and go extinct. In addition, Young (1993a) looks at models much like those we will present here and proves that when two populations play the Nash demand game, the unique SSE of the model is the Nash bargaining solution.[19] In other words, power (as represented by disagreement points) translates to a bargaining advantage in evolutionary models as well as rational-choice ones. Therefore, in both types of models, it pays to be powerful. However, the reason is not the same: in the evolutionary models, the disagreement points influence the population dynamics such that the predicted, population-wide equilibrium is changed, whereas in the rational-choice models the disagreement points influence the actual choices of individuals.

Both Bruner (2017) and Young (1993a) consider social groups that are homogeneous with respect to power. One side has a lower disagreement point, and the other side has a higher disagreement point. However, as noted in the introduction, real social groups are heterogeneous with respect to power, in the sense used here. All women do not face the

[18] They also examine simulations of games wherein the players have different background payoffs, and different outside options (resulting in a collaboration game).

[19] Binmore et al. (2003) expands these results different versions of best response dynamics and for some coordination games.

same, poor outcome if they divorce, neither do all men expect to be in a good situation. If we think about segregated bus seating, while all black people could expect a negative outcome from sitting in the front of the bus, a wealthy black person with an option to hire a lawyer would be in a better situation than a poor person who could not afford representation. The puzzle we address is to say why, despite this variability, we might see an entire social group nonetheless disadvantaged. Previous evolutionary models do not yet settle this point.

In the next section, we turn to agent-based models very similar to those developed by Axtell et al. (2001), but incorporating power differences between individuals, to investigate how this variation influences the emergence of bargaining conventions.

3. The Model

We vary a number of parameters for these models, including population size, N, memory length, m, and possible demands in the underlying game.[20] For each combination of parameters examined, we ran 1000 trials and investigated which of the three equilibria described above emerged.[21] (These remember, were equal demands, one population always getting more, or the other population always getting more.) At the outset of the simulation, agents begin with no memories whatsoever, and they determine their first strategy using a coin flip. Once the agents have at least one memory, they best respond to the memories that they have.[22] We consider two versions of the model. In the first, we add a small number of powerful individuals to one of the two groups. In the second, we consider two groups where

[20]In particular, we look at population sizes, $N = 10, 20, 40, 100$, with sub-populations always equal ($n_A = n_B = N/2$). We set the memory length $m = 5, 10, 15, 20$. And we examine possible demands given by $\langle 4, 5, 6\rangle$, $\langle 3, 5, 7\rangle$ and $\langle 2, 5, 8\rangle$.

[21]We used 1.0×10^4, 3.5×10^4, 3.0×10^6, or 5.0×10^7 time steps, for the population sizes $10, 20, 40$, and 100, respectively. The increase in timescale was to ensure that every (or almost every) trial had complete convergence to one or another equilibrium in order for us to meaningfully compare the results across population size. Since there is no error-rate in our model, and so no error in the agents' strategies, the change in timescale does not make any difference here: once the populations reach some equilibrium, they remain there forever.

[22]Axtell et al. (2001) begin with randomly initialized memories. O'Connor (2017) points out that, for low values of L, more agents will, on average, begin by demanding high, whereas for high values of L, more agents will, on average, begin by demanding low, when the memories are initially randomly implemented.

disagreement points are drawn from distributions with different means. More details are provided below.

Before discussing our results, we want to make clear how a higher disagreement point might shift the dynamic in question. Suppose that our population consists of exactly two individuals—called *Player* 1 and *Player* 2. This means that on each round of play these two players meet and play the Nash demand game. Suppose that for our Nash demand game $L = 4$, $M = 5$, and $H = 6$. Suppose also that the disagreement point for both players are D and d respectively.

Further suppose, in our example, that some game-play has already taken place, with the result that Player 1's memory is $\langle M, H, H, H\rangle$ and Player 2's memory is $\langle L, L, L, L\rangle$. Thus, Player 2 has demanded H thrice and M once, and Player 1 has only demanded L—at least as far as they remember. When $D = d = 0$, the normal best response for Player 1 will be to demand low, since her expected payoff for this demand is 16, whereas her expected payoff for M is 5, and for H is 0. Now suppose that the disagreement points differ. Let $D = 4$ and $d = 0$. Player 1's expected payoff for L is still 16; however, her expectations for demanding M and H are now 17 and 16, respectively. Thus, her best response in this situation is now to demand medium.

Let us examine this situation in more detail. The Players' memories and their respective payoffs for the situation in which $D = d = 0$ are given in Round 1 in Table 2. The players'

	Memory		**Payoff** (L, M, H)	
	Player 1	Player 2	Player 1	Player 2
Round 1	$\langle M, H, H, H\rangle$	$\langle L, L, L, L\rangle$	$(\mathbf{16}, 5, 0)$	$(16, 20, \mathbf{24})$
Round 2	$\langle H, H, H, H\rangle$	$\langle L, L, L, L\rangle$	$(\mathbf{16}, 0, 0)$	$(16, 20, \mathbf{24})$

Table 2: Example Game-Play with Disagreement Points $D = d = 0$

respective best responses based on their memories are the boldfaced payoffs. As such, Player

In this sense, the initial demands are skewed in one direction or the other. Hence why we implement our memories in this way instead.

1 demands L in round 1 and Player 2 demands H. They update their memories for Round 2. At this point, we see that their memories will never change (assuming no error), since they have cemented a situation in which Player 1 always demands low and Player 2 always demands high. This is a stable outcome.

Now, suppose again that $D = 4$ and $d = 0$. An example of one particular path this game might take is shown in Table 3. The difference caused by the disagreement point in round 1

	Memory		**Payoff** (L, M, H)	
	Player 1	Player 2	Player 1	Player 2
Round 1	$\langle M, H, H, H\rangle$	$\langle L, L, L, L\rangle$	$(16, \mathbf{17}, 16)$	$(16, 20, \mathbf{24})$
Round 2	$\langle H, H, H, H\rangle$	$\langle L, L, L, M\rangle$	$(16, 16, 16)$	$(16, 17, \mathbf{18})$
Round 3	$\langle H, H, H, H\rangle$	$\langle L, L, M, H\rangle$	$(16, 16, 16)$	$(\mathbf{16}, 15, 12)$
Round 4	$\langle H, H, H, L\rangle$	$\langle L, M, H, H\rangle$	$(16, 17, \mathbf{18})$	$(\mathbf{16}, 10, 6)$
Round 5	$\langle H, H, L, L\rangle$	$\langle M, H, H, H\rangle$	$(16, 18, \mathbf{20})$	$(\mathbf{16}, 5, 0)$
Round 6	$\langle H, L, L, L\rangle$	$\langle H, H, H, H\rangle$	$(16, 19, \mathbf{22})$	$(\mathbf{16}, 0, 0)$
Round 7	$\langle L, L, L, L\rangle$	$\langle H, H, H, H\rangle$	$(16, 20, \mathbf{24})$	$(\mathbf{16}, 0, 0)$

Table 3: Example Game-Play with Disagreement Point $d = 4$ for Player 1

is that medium, rather than low, is the best reply for Player 1. In Round 2, Player 1 is now *indifferent* between her strategies and so randomizes, giving each strategy 1/3 probability. Assume that she chooses H for her demand. Now, we have the exact same situation in Round 3, except Player 2's best response has shifted from H to L in light of Player 1's randomly chosen strategy from Round 2. Player 1 is still indifferent, so she randomizes over her strategies again. Assume that she chooses H. After Round 4, the game moves deterministically to its stable outcome shown in Round 7. Note that this outcome is the *opposite* of the convention that arose from game-play in Table 2.

This outcome is not determined, since there is an element of stochasticity in Rounds 2 and 3. However, in our example from Table 2, the outcome *was* determined. Changing the disagreement point, in this case, made possible an outcome that was previously impossible, and that greatly advantaged the more powerful player.

3.1. A Few Powerful Individuals

For the first models we consider, we set the disagreement point $d = 0$ for population A. The disagreement point for most of population B is also 0; however, a small handful of individuals in population B have disagreement point D, which ranges from 0 to $L + 0.5$ in intervals of 0.5. (As such, the group itself is not uniformly powerful—most members are identical to the less powerful group.) The two sub-populations are otherwise entirely symmetric.

The main question we are concerned with here is whether a small portion of powerful individuals, perhaps even just one, having a higher disagreement point will affect the outcomes for both of the populations. In every case, what we find is that when D is zero, all members of both populations are in a symmetric position as regards bargaining, and as a result the portion of the runs that end in inequitable norms for population B is effectively equivalent to those that end in inequitable norms for population A. As D increases, though, it is always the case that the group containing powerful members becomes increasingly likely to discriminate and increasingly unlikely to be discriminated against as a result.

Figure 1 shows this effect for a population with 10 members with $m = 10$, and with one single powerful individual. For this, and all the results we will display, the possible demands were 4, 5, and 6. As is evident, two things happen when D increases. First, there are more outcomes where the population with the powerful individual demands High. And second, it becomes increasingly unlikely that the group with the powerful individual will ever demand Low. While alterations to m (the memory length) slightly alter the results, in general they hold across parameter values.

What is driving this result? Members of group A regularly have to interact with the single powerful individual. However, from the point of view of the individuals in population A, the individuals in B are all similar. That is to say, an individual in A who is paired with an individual in B does not know whether that individual's disagreement point is higher. Thus, individuals in A learn to bargain as though the individual with whom they are playing

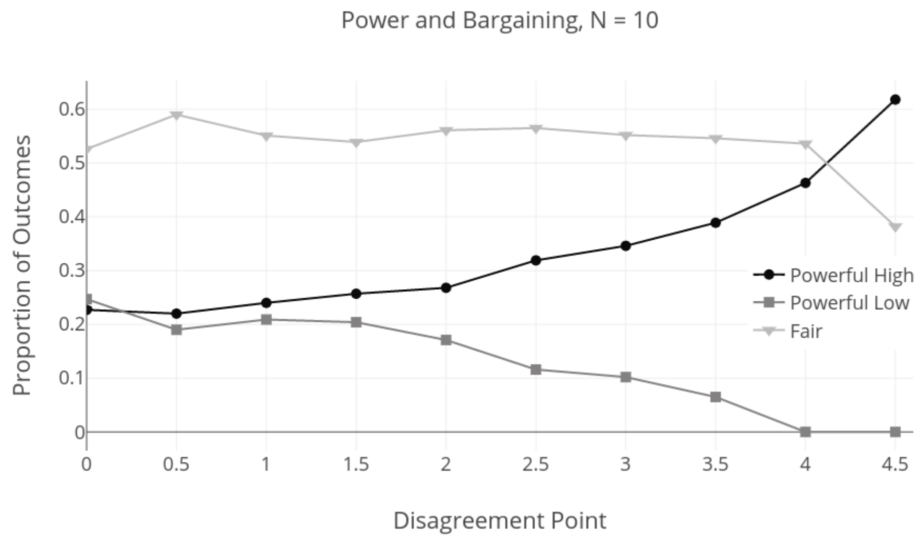

Figure 1: Bargaining outcomes in a population where $N = 10$ as the disagreement point for one powerful individual increases, $m = 10$.

with has a higher disagreement point, regardless of whether or not this is true in fact. As such, they play *as if* all individuals in population B do have a higher disagreement point. Thus, it is as if all individuals in B did have a higher disagreement point.

The remarkable thing about this result is that an entire social group tends to end up in a disadvantaged state as a single individual in the population gains power. The prediction, contra bargaining models grounded in the Nash solution, is that individuals with equal levels of power will often end up at bargaining outcomes where one gets more of a resource due to group level norms. And, in particular, the models predict that power does matter to these outcomes, but in a way that is quite different from in rational choice models. It is the power of social groups that matters here, rather than the individual positions of those involved in a bargain.

This effect of adding one powerful individual is less strong for a larger population for the simple reason that the powerful individual now makes up a smaller proportion of the total group. Suppose that we make a proportionally similar change though, by adding four powerful individuals to one group in a population where $N = 40$. Figure 2 shows these results. The trends observed are stronger than before, so that it becomes very likely the powerful group will demand high at the emerging convention, even though the vast majority of them are identical to their counterpart group.[23]

The effect of adding a few powerful individuals is especially strong when we let $D = 4$ or 4.5. This is because, at these values, the powerful individuals are never incentivised to demand low. When $D = 4.5$, in fact, demanding lower is never a best response for them. It is thus unsurprising that the inclusion of these individuals in a group makes it unlikely for the whole group to end up discriminated against. But notice that we see the effect for much lower values of D. In other words, just a slightly asymmetric bargaining position, such that it is still in the interest of the few powerful individuals to conform to any convention that emerges, still leads to advantage for an entire social group by changing the best responses of these individuals.[24]

[23] As these two figures make clear, when $D = 0$ for all individuals, larger populations are more likely to head to the fair convention than smaller ones. This is due to combinatorics and probability differences between larger and smaller populations. Suppose I am one of a population of 5. Then on a given round of play, there is a 1/5 probability that I am selected to bargain. When I am chosen the first time, I randomize over possible strategies—H, M, and L. Now, if I have one memory, I best respond to that memory. My best response is going to depend on the probability that my opponent chose H, M, or L, which, in turn, is going to depend on whether they were randomizing (likely at the outset) or best-responding to memories (more likely later in the game). There are three possible values for this one memory, each of which will correspond to a unique best response—H for L, M for M, and L for H. When I have two memories, however, 5/9 combinations of memory values have L for a best response; 1/3 combinations of memory values have M as a best response, and 1/9 combinations of memory values have H as a best response. However, having two memory-slots filled with values requires that I have been selected twice, and my being selected twice in any given number of rounds depends inherently on the number of individuals in my population. To simplify somewhat, suppose there are only three rounds to play and suppose that I am chosen to bargain on the third round. When the population consists in 5 individuals in my subpopulation, there is only a 1/25 probability that I was selected for *both* the previous two rounds, meaning there is a 24/25 probability that I was not. However, if my subpopulation consists in 50 individuals, then there is only a 1/2500 probability that I was selected in the previous two rounds (and so have two memories). Therefore, because it is more likely that, at the outset, individuals will have more memories (i.e., will have been selected more frequently) and so will be more likely to choose, e.g., a low strategy, it follows that their partner, who may not have any memories yet, will be more likely to choose fill those memory slots with remembrances of opponents playing low, and thus will be more likely to play high. Since this position is symmetric for every player when the disagreement point is 0, it follows that we should expect less fair play with smaller populations, and so more fair play with larger populations.

[24] Interestingly, when the population is large enough, and memory length long enough, we do sometimes see conventions where an entire social group will demand low but for one powerful individual ($D = 4.5$) who always demands high or medium. Because the other group always demands high, the entire population

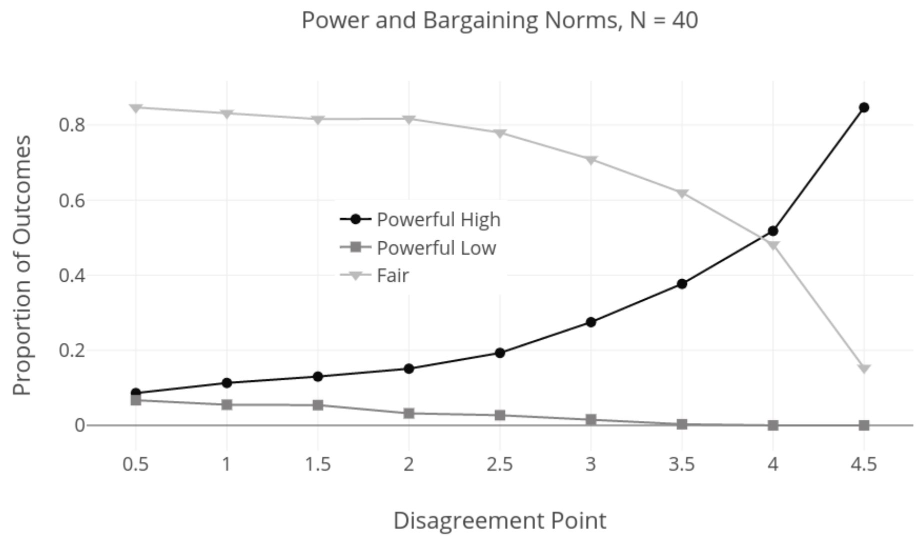

Figure 2: Bargaining outcomes in a population where $N = 40$ as the disagreement point for a few powerful individuals increases, $m = 10$.

3.2. Heterogeneous Power

We also consider models where instead of one individual, or a small handful of individuals, having an unusually high level of power, one population is more powerful on average with variation in both groups. This means that while one population tends to have higher disagreement points and one lower, there are individual interactions that flip this dynamic. Returning to the household bargaining example, these models capture a situation where men tend to have better fall back positions upon divorce, and women worse, but where for some marriages the woman is wealthier and more empowered than the man, and so does better when household bargaining breaks down.

In particular, we assume that disagreement points are sampled from (nearly) normal distributions, with mean 3 for the powerful group and mean 2 for the less powerful group. These distributions are nearly normal because when a value greater than 4.5 or less than 0 is chosen, we re-sample. The standard deviation of these distributions determines how much overlap there is between the two groups. When this value is very low, we approach

stays at the convention.

a situation where the groups are homogeneous, one with a disagreement point of 3 and the other 2. When the value is high, we approach a uniform distribution where the groups are identical in terms of power.

We find that even in simulations where the power relationships between the groups are much more ambiguous, power can advantage one group over another. While details vary based on parameters, in general the effect was stronger for small standard deviations of the distributions used to select disagreement points. It was generally weaker the larger the population, because large populations tended more strongly towards fairness. And it was weaker for longer memory lengths.

Figure 3 shows the effect for distributions with different standard deviations.[25] As described, when the two populations are more relevantly different the powerful one derives a more notable advantage. But in all cases there will be outcomes where, due to the emergence of group level conventions, more powerful individuals will receive low payoffs in bargains with less powerful individuals.

4. Conclusion

From a rational-choice perspective, we have seen that when players are in symmetric positions, we expect bargaining to result in symmetric outcomes. However, this only captures part of the story. In our first set of models, most of the *individual* bargaining situations—i.e., on a given trial, when two individuals are picked at random—are symmetric. In our second set of models, for some interactions either group will be more powerful. Nonetheless, bargaining power by association has a discernible effect on the end results of bargaining situations at a global level. The group dynamics of our models suggest that the global effects of power permeate throughout an entire group. As such, *individuals* in a disadvantaged population are disadvantaged by dint of the fact that they are a part of that population.

[25] For this figure, $N = 20$ and $m = 5$.

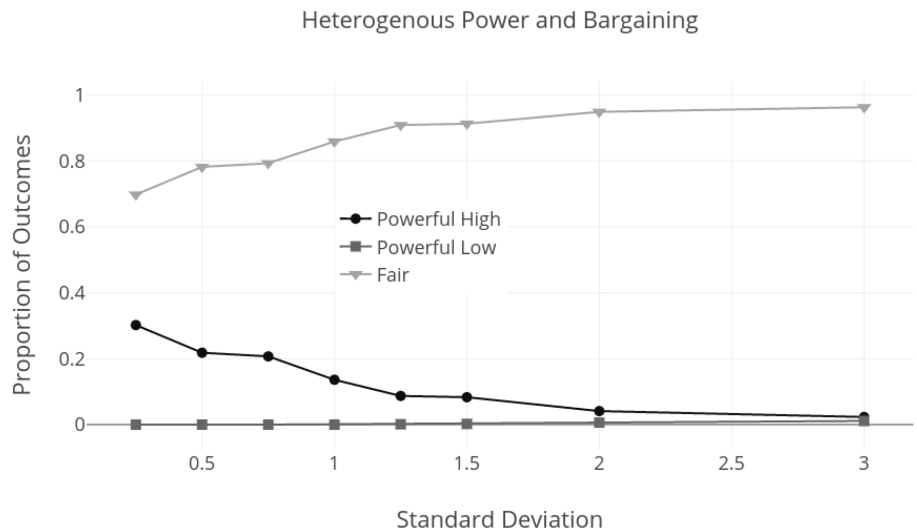

Figure 3: Bargaining outcomes in a population where $N = 20$ with heterogeneous disagreement points, as the standard deviation increases.

Similarly, individuals in the group containing some powerful members are advantaged by dint of the fact that they are *associated* with power. In this sense, they themselves are afforded a sort of *de facto* bargaining power. Because one of their in-group has *actual* power, they come to have power by association.

The models presented here are simple and highly idealized. This means, of course, that one must be careful in applying our results to complex real-world populations. It is clear that they show how, in principle, processes of cultural evolution and learning can lead to outcomes where individual power is less important than group membership in determining resource division. Additionally, in capturing this phenomenon, they show how evolutionary models of bargaining between groups can account for aspects of the real-world that rational choice models generally do not. All this said, it is unlikely that real world populations evolve bargaining conventions using anything particularly close to the processes we outline here. Cultural evolution and human learning are vastly complicated processes. Inasmuch as we can apply lessons from these models, it is largely in a 'how-possibly' way. We see how, contra rational-choice predictions, such patterns of behaviour can emerge. This possibility does mean we should be attentive, in the real world to the chance that this sort of cultural evolutionary process might be contributing to inequitable conventions and norms we may wish to intervene on.

One advantage of the simplicity of these models is that they can be applied widely. We capture a basic bargaining process which might represent salary negotiations, household bargaining, division of tangible resources, or division of credit for co-authored academic publications, for example. Our models do not include details that track the real differences between social identity groups, so the models might be taken to apply to gender, class, race, age, ethnicity—e.g., religion, spoken language, etc.—or any other discernible social grouping (e.g., unions or employers). We understand bargaining power as a higher status quo for a particular individual or group of individuals. In this context, a group has power whenever the individuals in that group have less to lose when negotiations break down—a common condition between such groups. In other words, the simplicity of our models adds some strength to our results in that they can be applied widely to the processes by which groups come to divide resources and the effects of power on this process.

Acknowledgements

Thanks to the UCI Social Dynamics seminar, and especially to Simon Huttegger, Brian Skyrms, and Jeffrey Barrett for comments on this work. This material is based upon work supported by the National Science Foundation under Grant No. 1535139. Many thanks for their support. Thanks to Justin Bruner, whose intellectual work helped inspire this project.

References

Alexander, J. McKenzie (2000). "Evolutionary Explanations of Distributive Justice." *Philosophy of Science*, *67*, 490–516.

Axtell, Robert L., Joshua M. Epstein, and H. Peyton Young (2001). "The Emergence of Classes in a Multi-

Agent Bargaining Model." *Social Dynamics*. Ed. Steven N. Durlauf and Peyton H. Young. Cambridge: The MIT Press, 191–211.

Baker, Michael and Marie Drolet (2010). "A New View of the Male/Female Pay Gap." *Canadian Public Policy*, *36*, 429–464.

Becker, Gary (1974). "A Theory of Social Interactions." *Journal of Political Economy*, *82*, 1063–1094.

Becker, Gary (1981). *A Treatise on the Family*. Harvard, MA: Harvard University Press.

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*, 593–598.

Binmore, Ken, Larry Samuelson, and Peyton Young (2003). "Equilibrium selection in bargaining models." *Games and economic behavior*, *45*(2), 296–328.

Binmore, Kenneth (1998). *Game Theory and the Social Contract Volume 2: Just Playing*. Cambridge: Cambridge University Press.

Binmore, Kenneth (2005). *Natural Justice*. New York: Oxford University Press.

Bruner, Justin (2017). "Minority Disadvantage in Population Games.".

Bruner, Justin and Cailin O'Connor (2017). "Power, Bargaining, and Collaboration." *Scientific Collaboration and Collective Knowledge*. Ed. T. Boyer, C. Mayo-Wilson, and M. Weisberg. Oxford: Oxford University Press.

Ellingsen, Tore (1997). "The Evolution of Bargaining Behavior." *The Quarterly Journal of Economics*, *112*, 581–602.

Feh, C. (1999). "Alliances and reproductive success in Camargue stallions." *Animal Behavior*, *57*, 705–713.

Foster, Dean and Peyton Young (1990). "Stochastic evolutionary game dynamics." *Theoretical population biology*, *38*(2), 219–232.

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

Horne, Rebecca M., Matthew D. Johnson, Nancy L. Galambos, and Harvey J. Krahn (2017). "Time, Money, or Gender? Predictors of the Division of Household Labour Across Life Stages." *Sex Roles*, 1–13.

Kalai, Elud and Meir Smorodinsky (1975). "Other Solutions to Nash's Bargaining Problem." *Econometrica*, *43*, 513–518.

Lewis, David (1969). *Convention*. Oxford: Blackwell.

Lundberg, Shelly (2008). "Gender and Household Decision-Making." *Frontiers in the Economics of Gender*. Ed. Francesca Bettio and Alina Verashchagina. New York: Routledge, 116–134.

Lundberg, Shelly and Robert A. Pollak (1996). "Bargaining and Distribution in Marriage." *The Journal of Economic Perspectives*, *10*, 139–158.

Manser, Marilyn and Murray Brown (1980). "Marriage and Household Decision-Making: A Bargaining Analysis." *International Economic Review*, *21*, 31–44.

McElroy, Marjorie B. (1990). "The Empirical Content of Nash-Bargained Household Behavior." *Journal of Human Resources*, *25*, 559–583.

McElroy, Marjorie B. and Mary Jean Horney (1981). "Nash Bargained Household Decisions: A Generalization of the Theory of Demand." *International Economic Review*, *22*, 333–349.

Moulin, Herve (2004). *Fair Division and Collective Welfare*. Cambridge, MA: MIT Press.

Moyser, Melissa (2017). "Women and Paid Work – Catalogue no. 89-503-X." *Women in Canada: A Gender-based Statistical Report*. . 7 edition. available online at: http://www.statcan.gc.ca/pub/89-503-x/2015001/article/14694-eng.htm. Ottawa: Statistics Canada, 1–38.

Nash, John (1950). "The Bargaining Problem." *Econometrica: Journal of the Econometric Society*, *18*, 155–162.

Nash, John (1953). "Two-Person Cooperative Games." *Econometrica: Journal of the Econometric Society*, *21*, 128–140.

O'Connor, Cailin (2017). "The Cultural Red King Effect." *The Journal of Mathematical Sociology*, 1–17.

O'Connor, Cailin (2018). *Dynamics of Inequity*. Forthcoming. Oxford University Press.

O'Connor, Cailin and Justin Bruner (2017). "Dynamics and Diversity in Epistemic Communities." *Erkenntnis*.

Poza, David J, José I Santos, José M Galán, and Adolfo López-Paredes (2011). "Mesoscopic effects in an agent-based bargaining model in regular lattices." *PloS one*, *6*(3), e17661.

Poza, David J, Félix A Villafáñez, Javier Pajares, Adolfo López-Paredes, and Cesáreo Hernández (011b). "New insights on the Emergence of Classes Model." *Discrete Dynamics in Nature and Society*.

Rubin, Hannah and Cailin O'Connor (2018). "Discrimination and Collaboration in Science." *Philosophy of Science*.

Samuelson, Paul A. (1956). "Social Indifference Curves." *Quarterly Journal of Economics*, *70*, 1–22.

Sen, Amartya (1982). *Poverty and Famines: An Essay on Entitlement and Deprivation*. Oxford: Oxford University Press.

Skyrms, Brian (1994). "Sex and Justice." *The Journal of Philosophy*, *91*, 305–320.

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

Young, Peyton H. (1993a). "An Evolutionary Model of Bargaining." *Journal of Economic Theory*, *59*, 145–168.

Young, Peyton H. (1993b). "The Evolution of Conventions." *Econometrica*, *61*, 57–84.