Ergo AN OPEN ACCESS JOURNAL OF PHILOSOPHY

Power by Association

TRAVIS LACROIX
Department of Philosophy, Dalhousie University
&
CAILIN O'CONNOR
Dept. of Logic and Philosophy of Science, University of California, Irvine

We use tools from evolutionary game theory to examine how power might influence the cultural evolution of inequitable conventions 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. Our models show that even when most individuals in two discernible sub-groups are relevantly identical, powerful individuals can affect the social outcomes for their entire group under a range of conditions; this results in power *by association* for their in-group and a bargaining disadvantage for their out-group.

Not so long ago in America, it was a rule that Black people sit at the back of the bus and white people at the front. Today, women tend to be paid less than men, on average, for the same work, even when the data are adjusted for external factors that may exacerbate such a discrepancy.[1] These are both cases

1. Accounting for discrepancies in work hours between genders, women earn $0.87 for every dollar of their male counterparts (Baker & Drolet 2010; Moyser 2018). Even when women have the same level of education, or work the same jobs, this gap persists (Kroeger, Cooke, & Gould 2016). Controlling for occupation narrows the gap but does not eliminate it; further, the gap is generally more significant for women of colour and LGBTQ+ individuals, and it tends to increase over one's career (Payscale 2019).

Contact: Travis LaCroix <tlacroix@dal.ca>
Cailin O'Connor <cailino@uci.edu>

https://doi.org/10.3998/ergo.2230

where patterns of behaviour related to the division of resources disadvantage those in one social group compared to those in another.

In investigating such phenomena, economists have often employed the *Nash demand game*—a simple, game-theoretic model intended to capture situations where two actors divide a resource (Axtell, Epstein, & Young 2001; Manser & Brown 1980). 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 game or Nash demand game derived scenario. This observation has been taken to inform inequities between social groups like those just described. For example, applying the bargaining model to household bargaining, economists have argued that when women are in a less powerful bargaining position, they should also be expected to do more work in the household (Manser & Brown 1980; McElroy & Horney 1981).

In many cases, though, personally powerful individuals nonetheless end up disadvantaged in bargaining scenarios by dint of belonging to disadvantaged social groups. For example, many of the racial injustices against Black people in the United States appear to have little to do with the circumstances of the individuals in question, and more to do with social-group membership. And returning to household bargaining, as an illustrative case, it has been widely noted that women with earnings equal to, or even greater than, their husbands' still tend to do more household labour (Horne, Johnson, Galambos, & Krahn 2018). 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 types get less, in various scenarios. 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 conventions between discernible groups, such as gender or racial groups, in a population of otherwise identical individuals. Previous similar models have assumed that power is homogeneous across a social group—that is, if women are disempowered with respect to, e.g., household bargaining, then they are all disempowered and all in the same way. These models thus 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. Unlike prior models, 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 affect the social outcomes. 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 their two social groups. As we point out in the conclusion, both aspects seem relevant to the emergence of real-world bargaining.

The paper proceeds as follows. Section 1 outlines relevant previous work in the field and lays the technical groundwork for discussing the models we examine. Section 2 describes our models in detail and presents original results for bargaining conventions and the cultural evolution of inequity. We model several scenarios, illustrating that the phenomenon of power by association is robust across a variety of modelling choices. Section 3 discusses the relevance of these results and how to interpret them.

1. Background and Previous Results

1.1. Bargaining Games and Power

In early game-theoretic work, Nash (1950) introduced a bargaining problem where two actors divide a resource. In this problem, there is conflict in that each prefers to get more. Nonetheless, both actors are incentivised to reach an agreement because doing so improves their payoffs from their baseline, which is sometimes called the *disagreement point*. There are many feasible agreements the actors could reach that would be jointly agreeable—i.e., better than the disagreement point. So, which does the model predict? Nash's famous solution to this problem introduced a set of axioms that, he argued, any solution should satisfy. He then proved that there is one unique division of the resource satisfying these desiderata.

Describing these axioms goes beyond the scope of this paper.[2] But the solution stipulates that for payoffs to the two players, u_1 and u_2, and disagreement points d_1 and d_2, the players should maximise $(u_1 - d_1)(u_2 - d_2)$. In other words, the solution expects both players to maximise the product of the difference between their respective payoffs and disagreement points. Notice that when the play-

2. Nash's axioms are 1) Pareto efficiency, 2) Symmetry, 3) Invariance to affine transformations, and 4) Independence of irrelevant alternatives. Kalai and Smorodinsky (1975) give an influential variation on these axioms. See also Moulin (2004).

ers have the same disagreement points, this solution will yield equal payoffs. If one player has a higher disagreement point, on the other hand, the outcome will tend to favour that player.[3] Subsequent models that explicitly represent a bargaining process, rather than taking an axiomatic approach, have also been shown to predict this division (Binmore 1980; Binmore, Rubinstein, & Wolinsky 1986; Rubinstein 1982).[4]

Nash (1953) re-interpreted the disagreement point in his models to correspond to an issued threat about what would happen should bargaining fail. Under this interpretation, the disagreement point can be taken to correspond to the power of an individual bargainer. Whoever can 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, arguably more powerful, place for bargaining in general. They need not care as much about the bargain succeeding and can use this to their advantage. For this reason, in the models we present, we will operationalise power using differences in disagreement points.

There are other ways to model power in the context of game theory. A powerful actor might be able to

1. take options away from her opponent's choice set,
2. change the relative costs of actions to make specific options unappealing to her opponent,
3. change the likelihood that a particular action will lead to a particular outcome, or

change her opponent's beliefs about costs and benefits associated with different actions in the choice set (Dowding 1991; 1996; 2011).[5]

We are not claiming that disagreement points are the only way to model power in bargaining situations, or even the most important one. It is merely one way to capture an aspect of what it means for bargainers to be more or less powerful.

Also, this choice fits with certain ways of treating power in the literature. Philosophers and social scientists have done a great deal of work on the concept of power. There are roughly two camps (Allen 2016). Some philosophers—e.g., Arendt (1970)—focus on *power-to*, which emphasises the abilities or capacities of

3. These claims hold assuming the feasible set of outcomes does not prevent an equal split of the right sort or one that advantages the player with the higher disagreement point.

4. Though one must set up the details of the static problem in such a way that it captures relevant asymmetries between agents in the dynamic problem (Binmore et al. 1986).

5. See also discussion in Binmore, Morgan, Shaked, and Sutton (1991).

individuals to act in the world. Others have focused on *power-over*, which Weber (1978: 53) defines as 'the probability that one actor within a social relationship will be in a position to carry out his own will despite resistance'.

The notion of a threat point from Nash is more in line with the second conception of power: one actor can gain economic benefits from another by wielding some kind of threat. However, as we shall see, in the models that we present, differences in disagreement points work differently. Actors need not use their advantages to force behavioural changes on an opponent. Instead, their advantages change their own behaviours, and the effects of these changes on the processes of social learning sometimes lead to further advantage. It seems fair to say that power, in this sense, fits well with the first conception: powerful individuals can make choices that would be risky for others with relatively little fear of consequence. In the conclusion, we return to how this particular conception ties into the feminist philosophy literature on what it means to be powerful.

As briefly mentioned in the introduction, in the influential work of Manser and Brown (1980) and McElroy and Horney (1981) (and in the work of many subsequent economists), household bargaining is modelled using the *bargaining game* or *Nash demand game* derived from Nash's problem. 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.[6] On their model, u_i is interpreted as some marital utility, which is itself a function of home and market goods and leisure time, and d_i is the disagreement point, which is construed as the 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 'extra-household 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 in the case of divorce (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 thus 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, these models predict asymmetric household labour distributions between men and women in a domestic partnership. When men have higher

6. Note, this literature typically focuses on monogamous, heterosexual households.

disagreement points, this translates 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, in the sense of equal disagreement points, they will divide household labour evenly. That is to say, symmetric bargaining positions should result in symmetric bargaining outcomes. However, as noted, empirical data show 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 these models (Horne et al. 2018).[7] Again, household bargaining merely provides a salient and well-studied example of the *type* of phenomena in which we are interested here. We will highlight connections to other such cases where relevant.

1.2. Evolutionary Game Theory and the Emergence of Inequity

One shortcoming of classical game-theoretic analysis is that it focuses on rational choice rather than looking at how boundedly-rational individuals might learn from each other, or culturally evolve. Evolutionary game theory, in contrast, looks at the emergence or evolution of strategic behaviour in a group. When it comes to the cultural emergence of bargaining behaviour, in particular, this framework has had many explanatory successes. For instance, Skyrms (1994; 2014) uses an evolutionary model of actors playing a version of 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 (Maynard Smith & Price 1973). This 'fair' division is also the most common outcome in many evolutionary models.[8] Thus we see why conventions for justice and fairness might be so common in human societies.[9]

7. 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 more detail.

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

9. Another standard model for studying the economics of fairness in a game-theoretic context is the *Ultimatum Game* (Güth, Schmittberger, & Schwarze 1982). Here, two players must divide a certain sum of money between them. One player makes an offer, which the other player can accept (in which case both players receive what was proposed) or not (in which case neither player receives anything). From a rational-choice perspective, the proposer should offer a minimal non-zero amount, and the responder should accept it. However, experimental evidence strongly suggests that human players typically reject offers less than 30% (Henrich, Boyd, Bowles, Camerer, & Fehr 2004; Oosterbeek, Sloof, & van de Kuilen 2004). Evolutionary game theory has also been applied in this case to explain the discrepancy between the rational-choice prediction and experimental evidence; see, e.g., Güth and Yaari (1992), Güth (1995), Gale, Binmore, and Samuelson (1995), Skyrms (2014), Huck and Oechssler (1999), Nowak, Page, and Sigmund (2000). The rejection of low-ball offers is often interpreted as an evolved response to these bargaining asymmetries,

These evolutionary models can also show how inequitable conventions might emerge in a group. Economists and philosophers of science have decisively demonstrated that once a population is broken into social groups, equity is no longer the expected outcome of such models.[10] 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 some 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 there is a resource of value 10, which two agents must divide. Each agent demands some portion of the resource for herself. If their demands sum to 10 or less, they each get what they requested. If they over-demand the resource, each receives their disagreement point. Assume that each individual can make one of three possible demands corresponding to a low (L), medium (M), or high (H) amount of the resource, respectively. Assume also that our medium demand corresponds to what we might call an equitable or fair demand. 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 example).

In this simplified game, let us represent the disagreement point for players 1 and 2 as D and d, respectively. Again, these are what the players receive in case bargaining breaks down. This game is shown in Table 1. Each entry shows the payoffs for a combination of strategies with Player 1's listed first. Note that if we only consider cases where D, $d < L$, this game has three *Nash equilibria*—i.e., strategies where no player can switch and improve her payoff.[11] These are bolded 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. In other words, these equilibria consist of the strategy pairings where actors perfectly divide the resource. 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.

limiting the ability of those with superior leverage to extract maximum advantage from that leverage.

10. See Axtell et al. (2001), Bowles (2004), Henrich and Boyd (2008), Stewart (2010), Poza, Santos, Galán, and López-Paredes (2011a), Poza, Villafáñez, Pajares, López-Paredes, and Hernández (2011b), Gallo (2020), Bruner and O'Connor (2017), O'Connor (2017; 2019), Rubin and O'Connor (2018), O'Connor and Bruner (2019).

11. 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 our evolutionary analysis.

Table 1. Simplified Nash Demand Game

		Player 2		
		L	*M*	*H*
	L	*L*, *L*	*L*, 5	**L, H**
Player 1	*M*	5, *L*	**5, 5**	*D*, *d*
	H	**H, L**	*D*, *d*	*D*, *d*

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 change. 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 investigate the emergence of inequitable or discriminatory conventions.[12]

Assume a finite population, consisting of N individuals. There are two sub-populations, A and B. These sub-populations are differentiated by 'tags'—i.e., salient and recognisable (though generally arbitrary) indicators of group membership—and 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 randomly to interact. They play the Nash demand game in Table 1. Furthermore, each agent is equipped with a finite memory of length m. This consists in the last m opponent strategies that she has encountered. Based on 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 the strategy that would have done best against this

12. Their model is based upon previous evolutionary models of bargaining from Young (1993a; 1993b). We focus on the work of Axtell et al. (2001), since their version of the model is closest to our own. There are other frameworks we might have used. This one is useful because, first, it is well understood (meaning we have reasonable control over why different things happen in the model) and, second, the agent-based structure allows us to create variable populations.

13. For more on this modelling choice and the fit between such models and real social categories like gender and race, see O'Connor (2019).

14. Bruner (2019) shows, using slightly different models, that when $n_A \neq n_B$, inequitable conventions can arise that disadvantage the minority population. O'Connor (2017) shows that these results are robust under the type of dynamics described here. We consider only equal groups to study the influence of power absent these effects.

limited memory. In doing so, she assumes that her experiences are a good guide to the strategies played in the other population. As such, her response is a type of boundedly-rational best response to her past experiences. When she is indifferent between strategies, she randomises.[15]

Axtell et al. (2001) point out that when there are no sub-groups, the equity convention (wherein every agent in the population demands *M*) is the unique stochastically stable equilibrium (SSE) of the model.[16] This result supports the results of Skyrms (1994; 2014), Binmore (1998; 2005), etc., regarding the evolution of fairness.

Axtell et al. (2001) then test the model with sub-groups, *A* and *B*. 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 depending upon whether they are paired with a member of their own group or the other group. As these authors observe, three equilibria 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, or one of the two groups demands high and the other low.

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.

It is by showing that these sorts of outcomes regularly arise in models with social groups that previous authors have begun to address the cultural emergence of inequitable conventions. Patterns of behaviour emerge over time that disadvantage one group to the advantage of the other. The robustness of these outcomes has been widely verified. Under different choices of dynamics and different population structures, groups of actors who are (1) of different types, (2)

15. Axtell et al. (2001) further assume that, with some small probability, ϵ, the agent chooses her strategy at random. We do not include this sort of stochasticity in our versions of these models, since we find, using simulations, that it makes little difference to the results.

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.

play a Nash demand game, and (3) update their strategies are commonly seen to evolve toward this sort of inequitable arrangement.[17]

Before continuing to our results, we wish to pull out one assumption all these models make and discuss its empirical justification. They assume individuals condition their bargaining behaviour on group membership—i.e., an agent can employ different strategies when bargaining with, e.g., men versus women. Furthermore, agents generalise their learning over these groups—i.e., an individual who encounters a number of women who are accommodating bargainers will assume that other women will tend to accommodate as well. Clearly, this assumption does not hold of all human interactions. People can differentiate individuals within social groups and treat them differently as such. However, empirical results indicate that humans do, in fact, generalise their learning over social categories and groups. Ridgeway (2011) gives an excellent overview of the extensive literature showing how interactions with individuals in a social group lead to general beliefs about the status and behaviours of that group. Particularly relevant is the finding that *primary social categories*, such as race, gender, and age, are used in many societies to condition behaviours across a broad range of interactive contexts. (In addition, as will become clear, we look at variations of the model where actors generalise over social categories in very minimal ways and find similar results.)

1.3. Power, Bargaining, and Evolution

Bruner and O'Connor (2017) use a framework similar to the one we are discussing to investigate how power for one social group might lead to an advantage for the emergence of bargaining conventions. They suppose that all members of one group have a higher disagreement point. As they show, this addition of power to their model makes it more likely that the population evolves to a convention where the powerful group demands high. 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.[18] In other words, power (as represented by disagreement points) translates to a bargaining advantage in evolutionary models as well as rational-choice ones.[19]

17. See references in Footnote 10.

18. Binmore, Samuelson, and Young (2003) expand these results to different versions of best response dynamics and for some coordination games.

19. 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

Both Young (1993a) and Bruner and O'Connor (2017), though, consider social groups that are homogeneous with respect to power: one side has a (uniformly) lower disagreement point, and the other side has a (uniformly) higher one. However, as noted in the introduction, real social groups are heterogeneous with respect to power. All women do not face the same, poor outcome if they divorce; neither do all men expect to be in a good situation. If we think about segregated bus seating in the mid-twentieth century United States, while all Black people could expect a negative outcome from sitting in the front of the bus, a wealthy Black person with the 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.

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

2. The Model

We model the simplified Nash demand game shown in Table 1 played by a population of N agents with two sub-populations of equal size. Each individual in the population has a fixed memory-length, m. At the outset of a 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. Across simulations, we vary the population size, the memory length, and the values of the demands in the underlying game to compare their results. During a particular simulation, each of these parameters is fixed. In particular, we look at population sizes, $N \in \{10, 20, 40, 100\}$. We set the memory length $m \in \{5, 10, 15, 20\}$. Possible demands are given by $\langle L, M, H\rangle \in \{\langle 4, 5, 6\rangle, \langle 3, 5, 7\rangle, \langle 2, 5, 8\rangle\}$.

For each combination of parameters examined, we ran 1000 trials of our agent-based simulation and investigated which of the three equilibria emerged: both populations always demand medium; population A always demands high, and population B low; or population B always demands high, and population

is changed; whereas, in the rational-choice models, the disagreement points influence the actual choices of individuals.

A low.[20] We consider several versions of the model, each where at least some members of one group have higher disagreement points (or more power) than the rest.

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, these two players meet and play the Nash demand game. Let $L = 4$, $M = 5$, and $H = 6$.

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. 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' respective best responses based on their memories are the boldfaced payoffs. As such, Player 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, since they have cemented a situation in which Player 1 always demands low, and Player 2 high. This is a stable outcome.

Now suppose that $D = 4$ and $d = 0$. An example of one particular path this game might take is shown in Table 3. The different disagreement point means that in round 1 medium, rather than low, is the best reply for Player 1. In Round 2, Player 1 is now *indifferent* between her strategies and so randomises. Assume that she chooses H for her demand. Now, we have the 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,

Table 2. Example game-play with disagreement points $D = d = 0$

	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**, 5, 0)	(16, 20, **24**)
Round 2	$\langle H, H, H, H\rangle$	$\langle L, L, L, L\rangle$	(**16**, 0, 0)	(16, 20, **24**)

20. We used 1.0×10^4, 2.5×10^4, 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 ensures that every trial converges completely to one or another equilibrium for us to compare the results across population size meaningfully. Since there is no error-rate in our model once the populations reach some equilibrium, they remain there forever.

Table 3. Example game-play with disagreement points $d = 4$ for Player 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, **17**, 16)	(16, 20, **24**)
Round 2	$\langle H, H, H, H\rangle$	$\langle L, L, L, M\rangle$	(16, 16, 16)	(16, 17, **18**)
Round 3	$\langle H, H, H, H\rangle$	$\langle L, L, M, H\rangle$	(16, 16, 16)	(**16**, 15, 12)
Round 4	$\langle H, H, H, L\rangle$	$\langle L, M, H, H\rangle$	(16, 17, **18**)	(**16**, 10, 6)
Round 5	$\langle H, H, L, L\rangle$	$\langle M, H, H, H\rangle$	(16, 18, **20**)	(**16**, 5, 0)
Round 6	$\langle H, L, L, L\rangle$	$\langle H, H, H, H\rangle$	(16, 19, **22**)	(**16**, 0, 0)
Round 7	$\langle L, L, L, L\rangle$	$\langle H, H, H, H\rangle$	(16, 20, **24**)	(**16**, 0, 0)

so she randomises 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.

2.1. *A Few Powerful Individuals*

For the first models we consider, the disagreement point is $d = 0$ for population A. The disagreement point for most of population B is also 0; however, a small handful of individuals in B have disagreement point D, which can range from 0 to $L + 0.5$. 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 is whether a small portion of powerful individuals will affect the outcomes for both of the populations. When D is zero, both populations are in a symmetric position and, as a result, are equally likely to be discriminated against. 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.

Figure 1 shows this for a population with 10 members and with one single powerful individual. For this, and all the results we will display, the possible

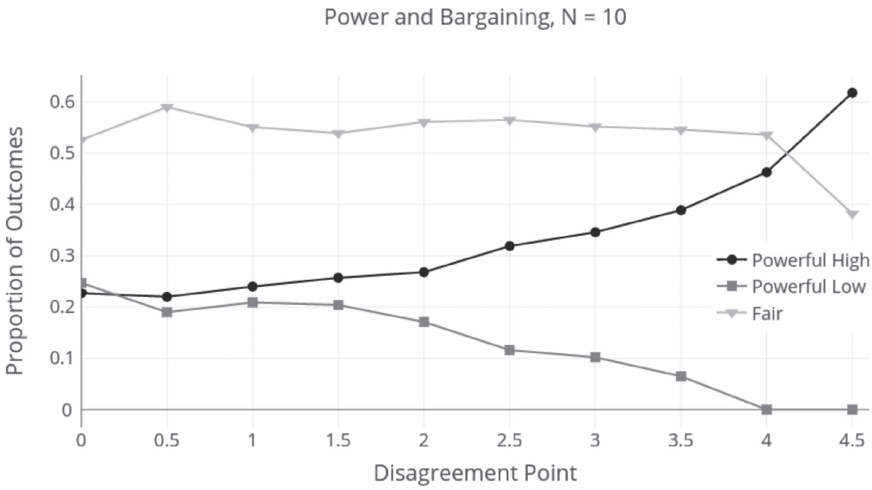

Figure 1. Bargaining outcomes in a population where $N = 10$ as the disagreement point for one powerful individual increases, $m = 10$.

demands were 4, 5, and 6. As D increases, there are more outcomes where the population with the powerful individual demands high. And, it becomes increasingly unlikely that this group will ever demand low. While alterations to m (the memory length) slightly alter the results, qualitatively they hold across parameter values.

What drives this result? Members of group A regularly interact with the single powerful individual. However, they also generalise what they have learned to other individuals in population B. Thus, individuals in A learn to bargain as though their interactive partners in B have higher disagreement points, regardless of whether or not this is, in fact, true.

The remarkable thing about this result is that an entire social group tends to end up in a disadvantaged state because a single individual in their out-group is powerful. 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 conventions. And, in particular, the models predict that power does matter to these outcomes, but in a way that is quite different from rational choice based 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 whole group. However, if we make a proportionally similar change—say, by adding four powerful individuals to one group in a population where $N = 40$—we find even stronger effects where the addition of a few powerful individuals greatly advantages one social group.

The effect of a few powerful individuals is especially strong when $D = 4$ or 4.5. This is because, at these values, powerful individuals are never incentivised to demand low. When $D = 4.5$, demanding low is never the 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 an advantage for an entire social group by changing the best responses of these individuals.[21]

2.2. *Heterogeneous Power*

We also consider models where one population is, *on average*, more powerful than the other, although individual levels of power within each population vary. As a result, although one population tends to have higher disagreement points and the other 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.[22] The standard deviation of these distributions determines how much overlap there is between the two groups. When this value is very low, we approach 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. In general, the effect is stronger for small standard deviations of the distributions used to select disagreement points.[23]

21. 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 stays at the convention.

22. These distributions are nearly normal because when a value greater than 4.5 or less than 0 is chosen, we re-sample.

23. It is generally weaker the larger the population because large populations tend more strongly towards fairness. And it is weaker for longer memory lengths.

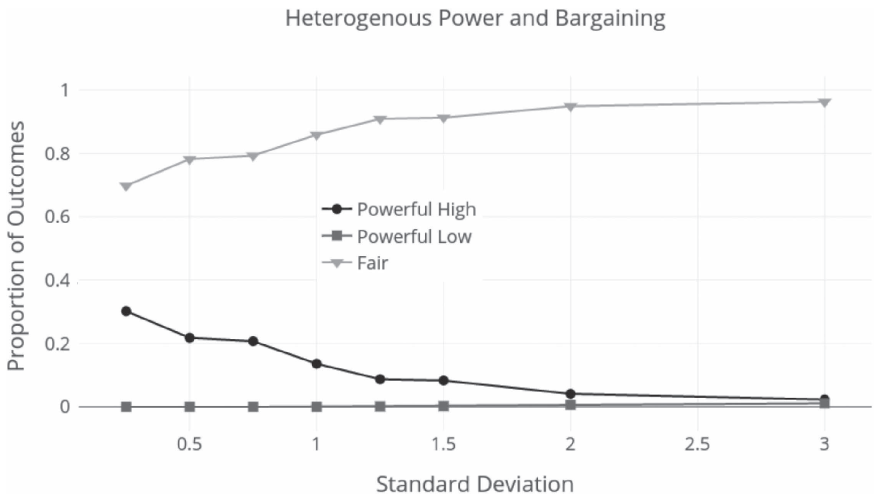

Figure 2. Bargaining outcomes in a population where $N = 20$ with heterogeneous disagreement points, as the standard deviation increases, $m = 5$.

Figure 2 shows the effect for distributions with different standard deviations. 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.

2.3. *Out-Group Bias*

In introducing the model, we mentioned that at the outset of simulation agents begin with no memories whatsoever. Their first strategy is determined with a coin flip. There is an obvious sense in which this is not realistic, since individuals may possess some psychological bias against their out-group. For this reason, we also consider a version of this model where actors have varying degrees of out-group bias. Assume that with some probability, β, an agent with no out-group memories will demand high of their out-group partner, and with probability $1 - \beta$, they flip a coin to determine their first strategy as before. Everything else is as described in Section 2.1.[24]

We examined a range of biases $\beta \in \{0.00, 0.05, 0.10, 0.15, 0.20, 0.25\}$. The results of this addition are relatively unsurprising. When the powerful population, B,

24. This is obviously a minimal sense of bias since it only influences the first interactions with an out-group. However, it alone is enough to impact outcomes significantly.

is biased toward the weak population, *A*, but not vice-versa, the weak population is greatly disadvantaged. When the weak group is biased, but the reverse is not true, the effects of power are somewhat mitigated. We find, however, that in many cases power can still advantage the powerful group, even given this initial discrimination against them. When *both* populations are biased toward each other, the power differential confers a clear bargaining advantage to the powerful group. We do not show figures for these results since they are unsurprising given the previously stated results. The main takeaway, though, is that our central findings are robust to this perhaps more realistic addition to the model.

2.4. *Individual Treatment*

An important concern one might have about the models thus far is that agents treat all members of a group as identical. Of course, in reality, it is possible to tailor interactive strategies to individuals. And, one might worry that the main results from the model would disappear if the agents could do so.

For this reason, we consider a version of the model where actors have separate memories for each possible interactive partner. We make the minimal assumption, though, that when interacting with a new partner, actors draw on general memories in deciding how to treat them. For example, suppose memory length is 10. If I have only 4 memories of interaction with you, I also consider my 6 most recent interactions with others of your type in choosing a best response. Once I have interacted with you enough, I treat you fully as an individual. We consider versions of this model with a few powerful individuals, like those in Section 2.1.[25]

Under this alteration, each pair of individuals across groups can develop their own, separate conventions. That is, actors can learn to make low demands of powerful individuals, but high demands of others. For this reason, we measure not the proportions of group-level conventions, but the proportions of low, medium, and high demands at the end of simulation across all individuals.

We find that even in this model, the power of a few significantly advantages others of their type. If *B* has a few powerful individuals, across parameter values, the weak members of *B* are less advantaged than the powerful ones. But still, their behaviour is much more similar to powerful members of *B* than to members of group *A*. This is because *A*s who have interacted with powerful

25. We also ran simulations with an added parameter, *h*, corresponding to the portion of individual versus group memories attended to in choosing a best response throughout the entire simulation—i.e., extrapolating between the previous models and the one just described.

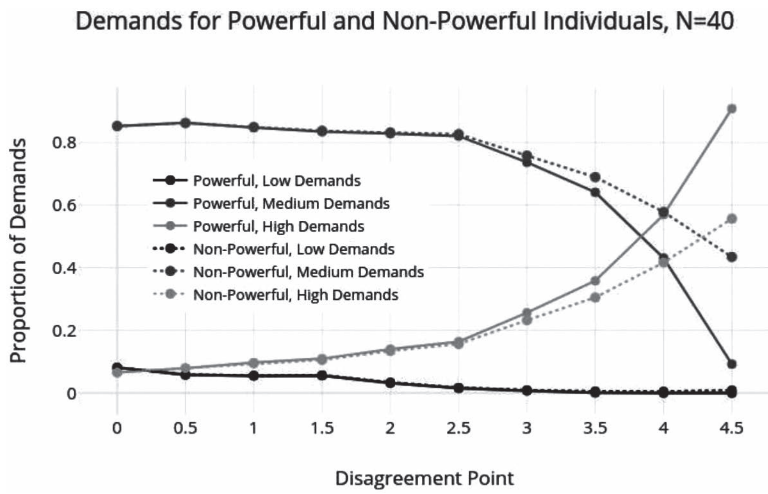

Figure 3. Final demands for powerful and non-powerful agents in a model with individualised memories. $N = 40$, $m = 10$.

*B*s transfer this lesson to new partners, making it more likely that they learn to make accommodating demands of other *B*s as well.

This is especially true for larger population sizes, where agents are less familiar with others as individuals for a longer time-frame. In these models, non-powerful *B*s tend to be almost as advantaged as powerful *B*s. This is clear in Figure 3. We show proportions of final low, medium, and high demands across simulations for both powerful (solid lines) and non-powerful (dashed lines) *B*s. (We do not show *A* strategies to keep the figure readable. They play complementary strategies.) The outcomes for powerful and non-powerful *B*s are almost identical until the disagreement point becomes so high that powerful *B*s

But since we see power by association even in the version where known actors are treated fully as individuals, we do not report these results. are never willing to demand Low. But even at this point, non-powerful *B*s end up demanding High more often than any other strategy, although their disagreement points are identical to those in the *A* group. In other words, even when actors can treat each other individually, the minimal assumption that they generalise previous learning to brand new partners leads to power by association.

2.5. *Fixed Partnerships*

Another worry about the models we have presented is that interactions in the real world are not entirely random. This is especially true in the context of household

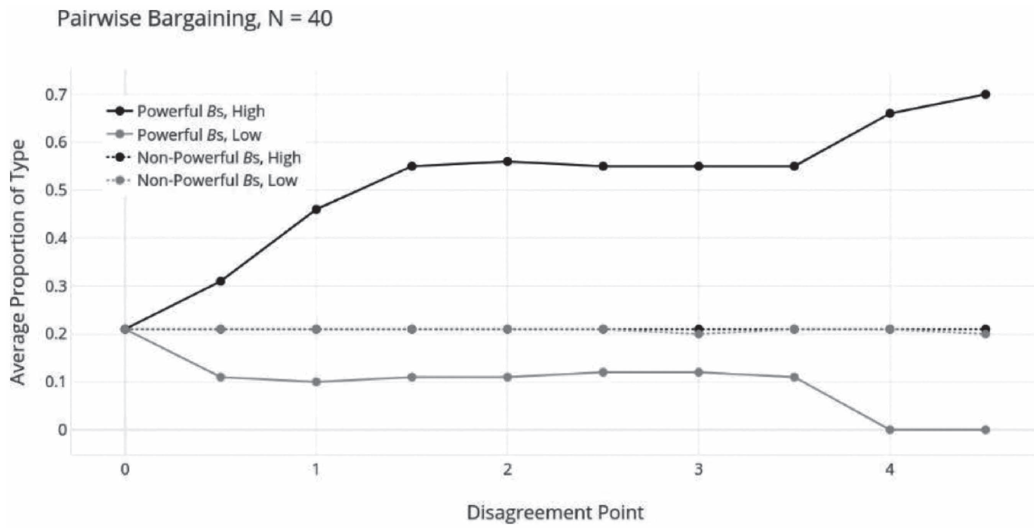

Figure 4. Proportion of outcomes for each type of individual, powerful and non-powerful, in population B. $T = 0$

bargaining. When it comes to traditional (monogamous, heterosexual) household formation, individuals often interact and bargain with multiple members of their out-group, but they eventually pair with a permanent (or semi-permanent) partner.

To address this worry, we present a model like the one in Section 2.1, except that pairing is only random for the first $T \in \{0, 25, 50, 100, 200\}$ time steps. After T time steps, individuals are permanently paired and only interact with their chosen partner for all subsequent rounds. As in Section 2.4, it is possible again under this alteration for each pair of individuals across groups to develop their own, separate conventions. Again, we measure the proportions of low, medium, and high demands at the end of simulation for all individuals to account for this. When pairing happens at the outset ($T = 0$) so that individuals have no opportunity to learn from general interactions with out-group members, only the powerful individuals themselves are afforded a bargaining advantage. This is shown clearly in Figure 4. Since every pairwise bargaining convention develops separately, there is no correlation between members of each group with respect to their bargaining outcomes—i.e., the fact that one member of B demands high does not bear on what another member of B demands, and thus we do not see power by association. Note further that non-powerful members of B are in the same position as each of the members of A and across runs will receive the same average payoff.

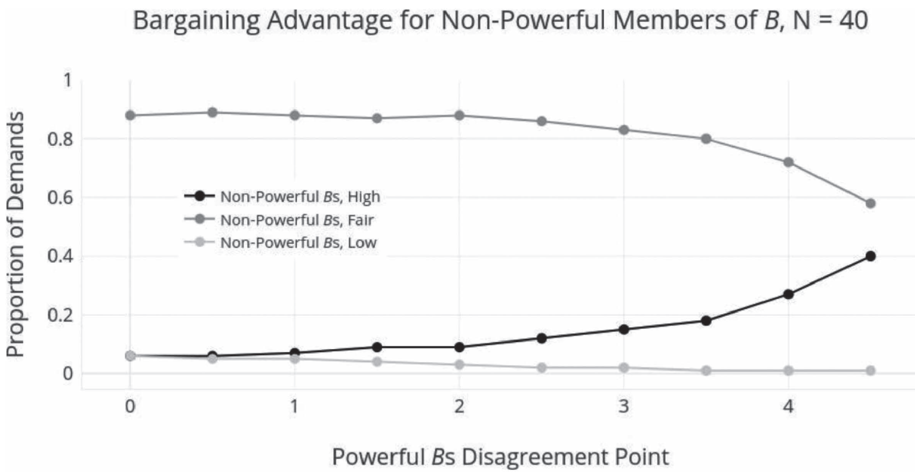

Figure 5. Bargaining outcomes for only non-powerful members of *B* in a population where $N = 40$ as the disagreement point for powerful members of *B* increases, $m = 10$, $T = 200$.

When some period of generalised learning occurs ($T > 0$), however, the phenomena we have been describing reappears. As the amount of time that individuals freely interact increases, an advantage is conferred to non-powerful members of *B*, since more *A*s experience interactions with their powerful *B* counterparts. In Figure 5, we look at an intermediate time-frame for interaction, $T = 200$, and consider just those outcomes for the non-powerful *B*s.[26] There is a clear advantage from their (temporary) association with powerful *B*s. If the disagreement point of powerful members of *B* did *not* confer a bargaining advantage to non-powerful members of *B*, each of these lines would be horizontal, as they are when $T = 0$ (Figure 4).[27]

Unsurprisingly, powerful *B*s gain a more dramatic advantage. For example, when individuals interact randomly for 100 rounds and then are partnered, and when the disagreement point is 4.5, almost all (90%) of the powerful individuals demand High of their out-group partners by the end of the simulation. Only 31% of *non-powerful* individuals end up demanding High. Even so, this is twice the number that would be expected with no power differential between groups.

26. Since individuals are chosen at random for the first 200 time-steps, and there are 40 individuals in the population, this means that each member of each sub-population has, on average, 10 interactions before being paired permanently with a member of their out-group.

27. Note, for these models, we see less fairness when *T* is low, even when there are no powerful individuals. This occurs because, in these sort of bargaining models, large groups tend to be more likely to settle on fair outcomes.

2.6. *Salience*

In the models presented thus far, the differential *prevalence* of powerful individuals creates a disadvantage for one social group. But such a disadvantage can also result when, despite equal numbers of powerful members, powerful individuals of one type are more salient. This might correspond to situations where, for cultural reasons, the power and success of one group are emphasised. Consider, for instance, the prevalence of depictions of white people as wealthy and Black people as poor in US media. We add this kind of salience to our model by supposing that powerful individuals have more influence over the memories and experiences of the other population, without necessarily being subject to extra influence themselves. In particular, we assume that they are chosen for interaction twice as often as others, but that they only update their memories as a result of half these interactions—i.e., sometimes their behaviours are seen by out-group members who are not seen themselves.

In Figure 6, we show results where A and B have equal numbers of powerful individuals (4 each, out of a total population of 40), but where the powerful members of B interact twice as often as others. The qualitative results are much as before: an entire social group ends up disadvantaged. This helps to add some subtlety to how power differentials may affect the dynamics of bargaining. In particular, merely affording one group power does not necessarily entail that group will gain a bargaining advantage. It must be the case that powerful individuals are influencing the expectations and behaviours of the out-group.

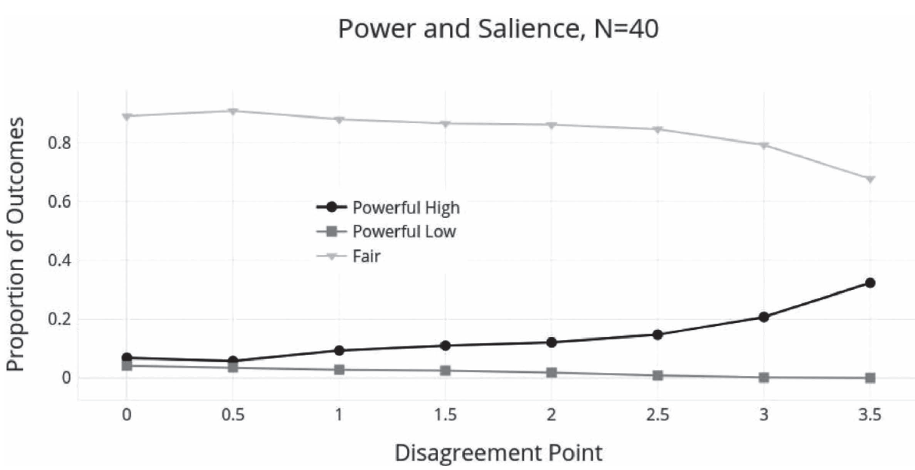

Figure 6. Bargaining outcomes with salience in a population where $N = 40$ as the disagreement point for a few powerful individuals increases, $m = 10$.

3. Conclusion

From a rational-choice perspective, when players are in symmetric positions, we typically predict bargaining to result in symmetric outcomes. However, this only captures part of the story. In our models, most of the *individual* bargaining interactions—i.e., on a given trial, when two individuals are picked at random—are symmetric. (In the models from Section 2.2, across different interactions, either group may be more powerful.) Nonetheless, bargaining power by association has a discernible effect on the end-results of bargaining situations at a global level. The dynamics of our models suggest that the 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.

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 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.

As noted, we use household bargaining as a key example because it is a well-studied and salient token of the type of phenomena we examine. However, many other real-world scenarios are captured by our model, including ones related to race. In many cases, for instance, Black individuals who are personally powerful are nonetheless subject to discriminatory outcomes by dint of group membership. The well-documented case of Viola Desmond, an elite member of the Black community in Nova Scotia, who was nonetheless violently removed from the (unofficially) 'white-only' section of a theatre in 1946 and subsequently convicted for 'non-payment of theatre tax', provides a good example (Backhouse 1994; Bingham 2013). Despite her personal power, Desmond was barred entry to the more expensive seating area solely because of her social-group membership. Conventional bargaining for spatial resources had already been well-established and reinforced in Canada by this time.[28]

There are connections between the modelling work we have presented and feminist accounts of power and oppression. First, what we have been calling 'power by association' is related to the description of *social power* and *identity power* given by Fricker (2007). Using her terminology, our models show how *agential* power (the sort of power that is exercised by an agent, like a powerful individual in our model) can give rise to *purely structural* power—where social power is thoroughly dispersed across a social group that it can be thought of as

28. The Supreme Court rejected a subsequent writ of *certiorari* to remove the record of convictions. Desmond was posthumously pardoned in 2010, almost a half-century after her death. See discussion in Backhouse (1994; 1999).

lacking a subject (Fricker 2007: 10–11); this is what happens when the population learns a (stable) discriminatory convention.

Second, as noted, a high disagreement point can result from an economic advantage. But the inequitable conventions that arise in our models lead to a further economic advantage for one social group. In future scenarios, then, members of that group should expect to have higher disagreement points. In other words, we might expect a cyclical process where power for some members of a social group leads to an economic advantage for all members, which in turn leads to power for all members (O'Connor 2019). This relates to claims from Okin (1989) that gender inequalities in marriages feed into cyclical processes that perpetuate women's oppression. In addition, both Okin (1989) and Cudd (1994; 2006) focus on the ways that social environments constrain the choices of women. Cudd, in particular, points out that in many cases, women make choices that perpetuate their disadvantage. But, as she argues, these choices are often rational, given the conditions women face. And so even though they have free choice, it is still right to label women as oppressed in such cases. Likewise, in our models, individuals make boundedly rational decisions at every stage. The existence of social categories, though, means that individuals face different social environments. And this means that even when they act rationally, they can contribute to developing and sustaining conventions that disadvantage them.

The models we have presented are simple and highly idealised. 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 just the processes we outline here. Cultural evolution and human learning are vastly complicated processes. Still, we can apply these lessons in a 'how-possibly' way. We see how, contra rational-choice predictions, such patterns of behaviour can emerge. Furthermore, given the robustness of the results, we can use these models to guide future study of the effects of power on the emergence of real bargaining conventions. In other words, we should become more attentive to the chance that this sort of cultural evolutionary process is contributing to inequitable conventions we see in the real world.

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. 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 or other discernible social groupings (e.g., unions and employers). 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 Justin Bruner, whose intellectual work helped inspire this project, and to at least two anonymous referees for their helpful comments. This material is based upon work supported by the National Science Foundation under Grant No. 1535139. Many thanks for their support. Thanks to the UCI Social Dynamics seminar, and especially to Simon Huttegger, Brian Skyrms, and Jeffrey Barrett for comments on this work. Thanks to audiences at ALFAn V and Philogica V (2018) in Villa de Leyva, Colombia, as well as the Canadian Philosophical Association (2018) in Montréal, Québec for helpful discussion. Thanks also to the Schwartz Reisman Institute at the University of Toronto for funding this research, and to Mila - Québec Artificial Intelligence Institute for providing computational resources and funding.

References

Alexander, J. McKenzie (2000). Evolutionary Explanations of Distributive Justice. *Philosophy of Science*, *67*(3), 490–516.

Allen, Amy (2016). Feminist Perspectives on Power. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2016 ed.). Retrieved from https://plato.stanford.edu/archives/fall2016/entries/feminist-power/

Arendt, Hannah (1970). *On Violence*. Houghton Mifflin Harcourt.

Axtell, Robert L., Joshua M. Epstein, and H. Peyton Young (2001). The Emergence of Classes in a Multi-Agent Bargaining Model. In Steven N. Durlauf and H. Peyton Young (Eds.), *Social Dynamics* (191–211). The MIT Press.

Backhouse, Constance (1994). Racial Segregation in Canadian Legal History: Viola Desmond's Challenge, Nova Scotia, 1946. *Dalhousie Law Journal*, *17*(2), 299–362.

Backhouse, Constance (1999). *Colour Coded: A Legal History of Racism in Canada, 1900–1950*. University of Toronto Press.

Baker, Michael and Marie Drolet (2010). A New View of the Male/Female Pay Gap. *Canadian Public Policy*, *36*(4), 429–64.

Bingham, Russell (2013). Viola Desmond. *The Canadian Encyclopedia*. Retrieved from https://www.thecanadianencyclopedia.ca/en/article/viola-desmond

Binmore, Ken (1980). *Nash Bargaining Theory II*. International Center for Economics and Related Disciplines, London School of Economics. (Now published in Ken Binmore and Partha Dasgupta (Eds.), (1987), *The Economics of Bargaining*, Basil Blackwell.)

Binmore, Ken (1998). *Game Theory and the Social Contract Volume 2: Just Playing*. The MIT Press.

Binmore, Ken (2005). *Natural Justice*. Oxford University Press.

Binmore, Ken, Ariel Rubinstein, and Asher Wolinsky (1986). The Nash Bargaining Solution in Economic Modelling. *The RAND Journal of Economics*, *17*(2), 176–88.

Binmore, Ken, Peter Morgan, Avner Shaked, and John Sutton (1991). Do People Exploit Their Bargaining Power? An Experimental Study. *Games and Economic Behavior*, *3*(3), 295–322.

Binmore, Ken, Larry Samuelson, and Peyton Young (2003). Equilibrium Selection in Bargaining Models. *Games and Economic Behavior*, *45*(2), 296–328.

Bowles, Samuel (2004). *Microeconomics: Behavior, Institutions, and Evolution*. Princeton University Press.

Bruner, Justin and Cailin O'Connor (2017). Power, Bargaining, and Collaboration. In Thomas Boyer-Kassem, Conor Mayo-Wilson, and Michael Weisberg (Eds.), *Scientific Collaboration and Collective Knowledge: New Essays* (135–157). Oxford University Press.

Bruner, Justin P. (2019). Minority (Dis)advantage in Population Games. *Synthese*, *196*, 413–27.

Cudd, Ann E. (1994). Oppression by Choice. *Journal of Social Philosophy*, *25*(s1), 22–44.

Cudd, Ann E. (2006). *Analyzing Oppression*. Oxford University Press.

Dowding, Keith (1991). *Rational Choice and Political Power*. Edward Elgar.

Dowding, Keith (1996). *Power*. University of Minnesota Press.

Dowding, Keith (2011). *Encyclopedia of Power*. Sage.

Ellingsen, Tore (1997). The Evolution of Bargaining Behavior. *The Quarterly Journal of Economics*, *112*(2), 581–602.

Foster, Dean and Peyton Young (1990). Stochastic Evolutionary Game Dynamics. *Theoretical Population Biology*, *38*(2), 219–32.

Fricker, Miranda (2007). *Epistemic Injustice: Power & the Ethics of Knowing*. Oxford University Press.

Gale, John, Kenneth G. Binmore, and Larry Samuelson (1995). Learning to Be Imperfect: The Ultimatum Game. *Games and Economic Behavior*, *8*(1), 56–90.

Gallo, Edoardo (2020). Communication Networks in Markets. *European Economic Review*, *129*, 103545.

Güth, Werner (1995). An Evolutionary Approach to Explaining Cooperative Behavior by Reciprocal Incentives. *International Journal of Game Theory*, *24*, 323–44.

Güth, Werner and Menahem E. Yaari (1992). Explaining Reciprocal Behavior in Simple Strategic Games: An Evolutionary Approach. In Ulrich Witt (Ed.), *Explaining Process and Change: Approaches to Evolutionary Economics* (23–34). University of Michigan Press.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze (1982). An Experimental Analysis of Ultimatum Bargaining. *Journal of Economic Behavior & Organization*, *3*(4), 367–88.

Henrich, Joseph and Robert Boyd (2008). Division of Labor, Economic Specialization, and the Evolution of Social Stratification. *Current Anthropology*, *49*(4), 715–24.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, and Herbert Gintis (Eds.) (2004). *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. Oxford University Press.

Horne, Rebecca M., Matthew D. Johnson, Nancy L. Galambos, and Harvey J. Krahn (2018). Time, Money, or Gender? Predictors of the Division of Household Labour Across Life Stages. *Sex Roles*, *78*, 731–43.

Huck, Steffen and Jörg Oechssler (1999). The Indirect Evolutionary Approach to Explaining Fair Allocations. *Games and Economic Behavior*, *28*(1), 13–24.

Kalai, Ehud and Meir Smorodinsky (1975). Other Solutions to Nash's Bargaining Problem. *Econometrica: Journal of the Econometric Society*, *43*(3), 513–18.

Kroeger, Teresa, Tanyell Cooke, and Elise Gould (2016). The Class of 2016: The Labor Market Is Still Far from Ideal for Young Graduates. *Economic Policy Institute*. Retrieved from epi.org/103124.

Lewis, David (1969). *Convention: A Philosophical Study*. Harvard University Press.

Lundberg, Shelly (2008). Gender and Household Decision-Making. In Francesca Bettio and Alina Verashchagina (Eds.), *Frontiers in the Economics of Gender* (116–33). Routledge.

Lundberg, Shelly and Robert A. Pollak (1996). Bargaining and Distribution in Marriage. *The Journal of Economic Perspectives*, *10*(4), 139–58.

Manser, Marilyn and Murray Brown (1980). Marriage and Household Decision- Making: A Bargaining Analysis. *International Economic Review*, *21*(1), 31–44.

Maynard Smith, John and George R. Price (1973). The Logic of Animal Conflict. *Nature*, *246*, 15–18.

McElroy, Marjorie B. (1990). The Empirical Content of Nash-Bargained Household Behavior. *The Journal of Human Resources*, *25*(4), 559–83.

McElroy, Marjorie B. and Mary Jean Horney (1981). Nash-Bargained Household Decisions: Toward a Generalization of the Theory of Demand. *International Economic Review*, *22*(2), 333–49.

Moulin, Hervé J. (2004). *Fair Division and Collective Welfare*. The MIT Press.

Moyser, Melissa (2018). Women and Paid Work – Catalogue no. 89–503-X. In *Women in Canada: A Gender-based Statistical Report* (1–38). Statistics Canada (7 ed.). Retrieved from https://www150.statcan.gc.ca/n1/en/catalogue/89-503-X

Nash, John (1953). Two-Person Cooperative Games. *Econometrica: Journal of the Econometric Society*, *21*(1), 128–40.

Nash, John F. (1950). The Bargaining Problem. *Econometrica: Journal of the Econometric Society*, *18*(2), 155–62.

Nowak, Martin A., Karen M. Page, and Karl Sigmund (2000). Fairness Versus Reason in the Ultimatum Game. *Science*, *289*(5485), 1773–75.

O'Connor, Cailin (2017). The Cultural Red King Effect. *The Journal of Mathematical Sociology*, *41*(3), 155–71.

O'Connor, Cailin (2019). *The Origins of Unfairness: Social Categories and Cultural Evolution*. Oxford University Press.

O'Connor, Cailin and Justin Bruner (2019). Dynamics and Diversity in Epistemic Communities. *Erkenntnis*, *84*, 101–19.

Okin, Susan Moller (1989). *Justice, Gender, and the Family*. Basic Books.

Oosterbeek, Hessel, Randolph Sloof, and Gijs van de Kuilen (2004). Cultural Differences in Ultimatum Game Experiments: Evidence from a Meta-Analysis. *Experimental Economics*, *7*(2), 171–88.

Payscale (2019). The State of the Gender Pay Gap 2019: Women of Color Face Multiple Barriers to Advancement In the Workplace. Retrieved from https://www.payscale.com/data/gender-pay-gap

Poza, David J., José I. Santos, José M. Galán, and Adolfo López-Paredes (2011a). 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 (2011b). New Insights on the Emergence of Classes Model. *Discrete Dynamics in Nature and Society*, *2011*, 1–18.

Ridgeway, Cecilia L. (2011). *Framed by Gender: How Gender Inequality Persists in the Modern World*. Oxford University Press.

Rubin, Hannah and Cailin O'Connor (2018). Discrimination and Collaboration in Science. *Philosophy of Science*, *85*(3), 380–402.

Rubinstein, Ariel (1982). Perfect Equilibrium in a Bargaining Model. *Econometrica: Journal of the Econometric Society*, *50*(1), 97–109.

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

Skyrms, Brian (1994). Sex and Justice. *The Journal of Philosophy*, *91*(6), 305–20.

Skyrms, Brian (2014). *Evolution of the Social Contract* (2nd ed.). Cambridge University Press.

Stewart, Quincy Thomas. Big Bad Racists, Subtle Prejudice and Minority Victims: An Agent-Based Analysis of the Dynamics of Racial Inequality. Annual Meeting of the Population Association of America.

Weber, Max (1978). *Economy and Society: An Outline of Interpretive Sociology*. University of California Press.

Young, H. Peyton (1993a). An Evolutionary Model of Bargaining. *Journal of Economic Theory*, *59*(1), 145–68.

Young, H. Peyton (1993b). The Evolution of Conventions. *Econometrica: Journal of the Econometric Society*, *61*(1), 57–84.