5

Power and the Evolution of Inequity

Why do some sorts of people get more and other sorts of people get less? Given the massive inequities in modern human society (of the second, non-functional type), especially those which rest on irrelevant personal characters like race and gender, this question demands attention. Often, the sorts of answers given appeal to the underlying psychology of human beings. Humans have innate in-group preferences. Humans are prone to stereotyping. Humans develop biases against those in certain social categories, including implicit biases, which lead to discrimination. Humans experience stereotype threat, leading to the poor performance of underprivileged groups when it is most crucial. Humans experience confirmation bias, so that events which confirm their stereotypes are salient, while those that oppose them fade away. Recipients of discrimination reasonably disinvest in skill learning on the belief that they will be undervalued in any case, causing further confirmation of stereotypical beliefs that they are unskilled or uneducated (Ogbu, 1978; D'souza, 1995; Loury, 1995; Stewart, 2010).

These sorts of phenomenon are deeply important to understanding inequity, and nothing in this book is intended to suggest otherwise. What I will show now, though, is that another sort of explanation of the second sort of inequity is available. In particular, as I will argue, a modeling perspective can show us that the conditions necessary to generate pernicious inequity in human societies are extremely minimal. Under these minimal conditions, cultural evolutionary pathways will march robustly toward inequitable systems. These models do not prove that real-world systems of inequity have, in fact, evolved via these simple cultural evolutionary pathways, but they tell us that they could. In particular, they show that

The Origins of Unfairness: Social Categories and Cultural Evolution. Cailin O'Connor, Oxford University Press (2019).
DOI: 10.1093/oso/9780198789970.003.0006

even if many of the most pernicious psychological facts about humans are removed or mitigated, inequitable conventions of the second sort are still expected to emerge.[1,2]

Part of the message here is that if we hope to reduce inequity in our own societies, working to eliminate the effects of psychological factors such as implicit bias may not be enough. The robustness of the emergence of inequity may mean that it is a fact we have to face again and again. This may require a retooling of our thinking about inequity not as something to conquer and be done with, but as something that will call for a persistent fight.

I'll start by introducing the primary paradigm used in this half of the book—the Nash demand game. I'll then move on to a perhaps unexpected topic—the evolution of fairness. As we'll see, previous evolutionary models of the Nash demand game have been used to argue that fair bargaining is what we should expect in human groups. I'll show why, once types are involved, fairness is often not the expected outcome of evolutionary systems. I'll then turn to the role power plays in the evolution of bargaining conventions. It has been long argued, using rational choice-based models, that power can influence bargaining outcomes between individuals in one-shot interactions. I will show how power, hashed out in a few different ways, can also influence the emergence of bargaining norms between groups. I will conclude with a model where power in one round can translate to advantage in later rounds for members of a type. As I will show, this can generate a process where even small initial differences between types can lead to wildly different outcomes.

[1] This sort of explanation bears some similarities to structural explanations of racial inequality. As Stewart (2010) describes such theories, "Race, from the structural perspective, is an integral piece of the social machinery that distributes private rewards and allocates an array of public privileges—it is more than an overt racist attitude/behavior expressed by a faction of dominant group members" (5).

[2] Hoffmann (2006) employs similar models to those presented here in work that he describes as looking at the role of human cognition in economic stratification. However, the only features of his models that represent human cognition are the use of types to condition behavior. Categorization and generalization are entirely general aspects of cognition, and so do not strike me as having much to do with humans' special evolved cognitive mechanisms. This said, the minimal explanations I generate do depend on the psychological tendencies of humans to form and use social categories, but they do not depend on psychological tendencies that we usually think of as the problematic ones.

5.1 The Nash Demand Game

The Nash demand game has been widely used as a model of human interaction when actors divide resources (Young, 1993b; Axtell et al., 2000; Stewart, 2010).[3] It involves two actors who each demand some portion of the resource. If these portions are compatible, the actors receive what they demand. If not, the actors receive a payoff called the "disagreement point", often set at 0, the idea being that their inability to settle on a mutually agreeable division leads to poor outcomes for both actors.[4]

Figure 5.1 shows two versions of this game. While in principle actors in the Nash demand game could demand any portion of a resource (53%, say), in practice it is common to look at simplified games where actors instead choose from a limited number of demands, or "mini-games," as Sigmund et al. (2001) have described them.[5] In (a), actors can make either a Low, Med, or High demand, and these correspond to demanding 3, 5, or 7 of a total resource of 10. In (b), the Med demand is dropped, so

(a)

Player 2

Player 1		Low	Med	High
	Low	3, 3	3, 5	3, 7
	Med	5, 3	5, 5	0, 0
	High	7, 3	0, 0	0, 0

(b)

Player 2

Player 1		Low	High
	Low	3, 3	3, 7
	High	7, 3	0, 0

Figure 5.1 Payoff tables for two simplified Nash demand games

3 This game has been given many labels over the years including, at least, the Nash bargaining game, the bargaining game, divide-the-dollar, divide-the-pie, and divide-the-cake.

4 As in the more purely coordination games, one should not take 0 to have special meaning here. Instead, it is a number that provides a baseline of what actors get for failing to coordinate that can be compared to their coordination payoffs.

5 For more on evolution and mini versions of the Nash demand game see Young (1993b); Skyrms (1994, 1996); Alexander and Skyrms (1999); Alexander (2000); Binmore (2008); Skyrms and Zollman (2010); Gallo (2014). These authors all employ games with a partition of finite demands.

actors can only choose Low or High. Notice that this second game has the same form as the complementary coordination games from Part I of the book.

For each of these games, the Nash equilibria are the strategy pairings where actors perfectly divide the resource (Low vs. High, High vs. Low, and Med vs. Med). If either actor demands more, they exceed the resource and both get nothing, if either actor demands less she simply gets less.

There is something special to note about the Nash demand game with a fair demand, which is that it has both a complementary and a correlative coordination character. Most equilibria of the game involve complementary demands—80% and 20%, or 63% and 37%. The fair demand is correlative, though, in the sense that actors take the same action to succeed. This makes the fair demand special from a population level and an evolutionary standpoint, which will be discussed in further detail shortly.

It is worth taking a minute to note that many economists think the Nash demand game misses crucial elements of some bargaining processes. For example, actors in bargains often take turns making offers until an agreement is reached, as in the highly influential model from Rubinstein (1982). Because I will focus in this book on an evolutionary scenario, the mini-game presented here can be thought of as having the potential to represent bargaining outcomes that result from all sorts of complicated, interpersonal dynamics. The representation is a rough one. Whatever the process that leads to a bargaining outcome, if the outcome is more preferable to one player than to the other, they have reached the High vs. Low equilibrium, and vice versa. If it is an approximately equal division, they have reached the Med vs. Med outcome. If they have pushed each other so hard that bargaining has failed, they reach the disagreement point.

5.2 The Evolution of Fairness and the Evolution of Unfairness

Much philosophical and economic work on the evolution of bargaining starts with the following explanandum: humans treat each other fairly. Suppose that you and Dwayne Johnson are going to share a pie, and you would each like to eat as much pie as possible. How would you split it? Chances are good that your first instinct is to split the pie into two exactly equal portions, and each take one.[6] In general, there seems to be a pull toward equality when it comes to dividing things in human groups. Empirical work backs up this intuitive claim. Humans have stated norms for fair divisions of resources (Yaari and Bar-Hillel, 1984), and in experimental settings involving the Nash demand game players strongly tend toward the 50–50 split (Nydegger and Owen, 1974; Roth and Malouf, 1979; Van Huyck et al., 1997).[7] Economists, for this reason, sometimes build other-regarding preferences for fairness into bargaining models (Fehr and Schmidt, 1999; Camerer, 2003).

To explain this phenomenon, Skyrms (1994) looks at the emergence of bargaining in single groups playing the Nash demand game. As he observes, the most likely outcome in these models (the equilibrium with the largest basin of attraction) is always for every member of the population to play the 50–50 split. One way to understand this result in the language of this book is to observe that the 50–50 split is special because it is the only equilibrium that involves a correlative, and not a complementary, solution to the underlying coordination problem. An entire population demanding 50% of a resource will always manage to coordinate their demands when they meet. On the other hand, a population with any other set of demands will sometimes mis-coordinate, for the same reasons as in any other complementary coordination problem. In a population where some actors make low demands, when they meet other low demanders they will waste the resource. In a population where some actors make high demands, when they meet other high demanders they will get the disagreement point.

Sometimes other outcomes arise. In particular, a "fractious" outcome, where some actors make high demands, and some low, sometimes emerges. In the service of explaining fairness, philosophers have looked for conditions under which these fractious outcomes disappear. Skyrms (1994) points out that correlation between strategies, so that Highs tend to meet Highs, Meds meet Meds, and Lows meet Lows, will push the population to the fair outcome. (This correlation prevents the complementary outcomes where Highs and Lows meet from occurring,

[6] Just kidding, obviously you would give all the pie to Dwayne Johnson.
[7] Except you, Mina Pedersen.

making it worse to demand High or Low.)[8] Alexander and Skyrms (1999) point out that this correlation can be generated by network interactions, where people tend to meet their neighbors again and again for interaction. And Alexander (2007) argues extensively that such repeated neighbor interactions can lead to the emergence of equitable bargaining. Lastly, as Skyrms (2014) points out, communication between neighbors—the simple ability to transfer information—tends to generate fair demands. Furthermore, the fair demand in the Nash demand game will always be the only evolutionarily stable strategy, or ESS of the game (Sugden, 1986), and the only stochastically stable equilibrium as well (Young, 1993b).[9,10]

If there is so much evolutionary pressure for fairness, though, the obvious question is: why inequity? To give one example, women regularly receive less in negotiations, and are punished for high demands (Bowles et al., 2007; Babcock and Laschever, 2009).[11] Why do we see these sorts of patterns, rather than the expected equitable ones? This is where types and type-conditioning come into the story. As I made clear in the first half of the book, populations with types are importantly different from populations without types because there is an extra piece of information available when an actor meets someone of another type. There we focused on the functional role this extra piece of information can play in facilitating coordination. When it comes to bargaining, though, this extra information can play a more sinister role. As we will see, it makes

[8] But see D'Arms et al. (1998) who counter that anti-correlation of strategies can lead to the evolution of unfair norms. Anti-correlation can, in effect, create a population that mimics a perfectly divided one. If two bargaining strategies evolve in the population, and only meet the opposite strategies, they end up in the same situation as actors who only meet different types and who evolve to play different strategies.

[9] Maynard-Smith and Price (1973) introduced this concept of evolutionarily stable strategies to identify strategies that are stable against invasion of mutants. ESSes in games can be identified through payoff tables alone. In this way, although a dynamic concept, ESSes do not depend on the choice of dynamic, and are stable for any dynamics that increase the proportion of strategies with higher payoffs. A strategy x_i is an ESS if $u(x_i, x_j)$ is the payoff of strategy x_i played against x_j and : 1) $u(x_i, x_i) > u(x_i, x_j)$ or 2) $u(x_i, x_i) = u(x_i, x_j)$ and $u(x_i, x_j) > u(x_j, x_j)$ for all $x_j \neq x_i$.

[10] A stochastically stable equilibrium is obtained by finding the limiting behavior of the system as the probability of experimentation goes to zero (Foster and Young, 1990; Young, 1993a,b).

[11] In a key study, Bowles et al. (2005) find that when women negotiate for others, they are more aggressive and demanding. In other words, they conform to non-aggressive normative behavior when bargaining, but in different circumstances are perfectly capable of behaving aggressively—indicating that the difference between men and women is not in ability, but in choice.

unfair solutions to the Nash demand game go from an impossibility to a commonplace. Typing creates new equilibria, not possible in single populations, where one group gets less and the other more. Axtell et al. (2000) label these outcomes as "discriminatory" in two-type models because they involve populations who treat out-group members differently from the in-group, to the detriment of one out-group. It is hard to understate the importance of this simple observation—add types to an evolutionary model of bargaining and now fairness is not the expected outcome. Instead, inequity and discrimination emerge.[12]

A number of previous authors have used this fact to model inequity, class stratification, or racism in two-population models. In economics, Axtell et al. (2000) look at the emergence of inequity in two populations playing the Nash demand game (as do Poza et al. (2011) in a follow-up paper). Bowles and Naidu (2006) and Hwang et al. (2014) look at the conditions under which inequitable conventions are stable in two-population models where actors from both groups try to move toward conventions that benefit themselves. Hoffmann (2006) considers the impacts of tags on populations evolving to play hawk–dove, and uses these models to explain social stratification. Smith and Choi (2007) present a rather complicated agent-based model where actors without types divide into upper and lower classes based on initial resource disparities. In sociology, Stewart (2010) shows how small numbers of discriminators can generate a system of persistent racial inequality in a bargaining model where actors sometimes engage in typing and type-conditioning. All of these models employ the base assumptions that allow inequity to emerge in the models

[12] Regarding gender and bargaining, there does not seem to be much experimental work looking at the effects of gender on play in the Nash bargaining game, though D'Exelle et al. (2017) find that women in rural Uganda make lower demands when playing with men than with women. It is worth noting that experimental evidence on asymmetric bargaining games, though, shows that inequity is common between social groups in such games. In the ultimatum game, one player offers another some portion of a resource. The second player has the opportunity to accept the offer or to reject it, leaving both players with nothing. In the dictator game, the second player must simply accept their offer. Solnick (2001) finds that women receive smaller offers from women and men in the simultaneous move ultimatum game (where actors choose strategies at the same time), and offer more to men. In the sequential ultimatum game, women receive lower offers, and are more likely to accept low offers, than men (Eckel and Grossman, 2001). Eckel and Grossman (1998) also find that women make larger offers in the dictator game than men do. (But see Frey and Bohnet (1995) and Bolton et al. (1998) who find no gender differences in the dictator game.)

here—actors are divided into types, they bargain, and via some dynamical process they adopt actions which benefit themselves.

To see how common unfair outcomes are in the simplest model that adds types to an evolving population, let us look at some evolutionary results. Figure 5.2 shows results from a simulation where actors in a group with two categories learn to bargain.[13] I assume, for now, that there are three strategies—Low, Med, and High—and I vary the values of the Low and High demands. In the figure, shaded patches in each bar represent the sizes of the basins of attraction for the three possible equilibria (which correspond to how often a simulation reaches them)—discrimination by one type, the fair outcome, and discrimination by the other type. When the Low and High demands are more inequitable (for example when Low = 1 and High = 9) the fair outcome has a relatively large basin of attraction, though the two inequitable outcomes are still possible. When Low and High are closer together the probability that the populations end up at a discriminatory bargaining convention becomes quite high.[14]

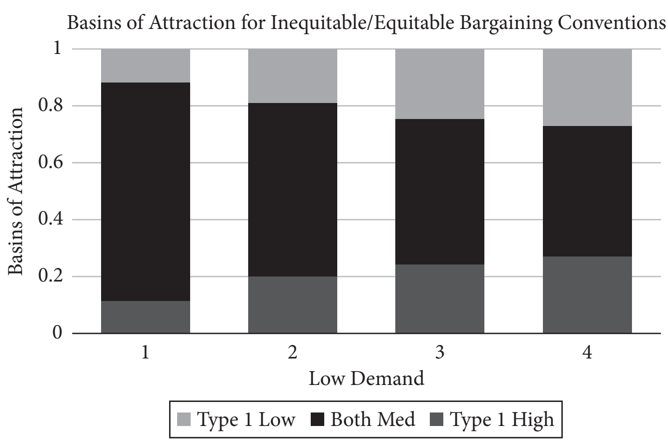

Figure 5.2 Basins of attraction for fair and unfair outcomes between types in the Nash demand game with three strategies

[13] These results use the discrete time-replicator dynamics and a perfectly divided population, but the results will be similar in a two-type mixing population. As elsewhere in the book, reported results are from 10k runs of simulation.

[14] This is somewhat similar to a result from Henrich and Boyd (2008) who find that the more inequitable the equilibrium for a two-strategy complementary coordination game, the less likely it is that an inequitable type-based convention emerges, and the more likely that

When Low = 4, the probability that the population ends up at one of the unfair conventions is 55%, meaning that more than half the time these simulations head to an unfair convention, despite the fact that there is perfect symmetry between the types. There is no reason, other than chance initial conditions, that one type or the other should get more, but the simple fact of types and typing means that they do. Notice that in these simulations there is perfect symmetry between the two types, so it is equally likely that each type ends up discriminating against the other.

Let us look at one more set of results. Consider again a population with two types playing a Nash demand game, but suppose we now have a strategic situation that allows for a more fine-grained division of resources. In particular, we will consider three models with increasingly fine partitions of demands. In the first actors have five available demands corresponding to 1, 3, 5, 7, or 9 of a resource of value 10. The second has 9 available demands (1, 2, 3, 4…), and the third 19 demands (0.5, 1, 1.5, …). The equilibria for each of these models will correspond to the fair outcome and the unfair outcomes where one type takes a low demand and the other type a complementary high demand, such as 0.5 and 9.5, or 3 and 7. This means there is a range of levels of inequity that might evolve, from the most serious to more minor. Figure 5.3 shows the size of the basin of attraction for each of these equilibria in each of the three models. In all three cases, the black patch represents the basin of attraction for the fair norm. Each slightly lighter patch represents the basin of attraction for an unfair norm, with lighter colors corresponding to greater inequity. (There is no legend for this figure since there are 10 possible equilibria for the model with 19 possible strategies.) One thing to note is that the fair equilibrium is always the one with the largest basin of attraction, and the less inequitable equilibria will have larger basins of attraction than the more inequitable ones. In all three models, though, inequitable conventions emerge more than half of the time. Furthermore, as the partitions grow finer, the chance that the population ends up at the fair equilibrium gets smaller and smaller. In the model with 19 possible demands, the population goes to the fair convention only 19% of the

a fractious but equitable outcome emerges. Bowles and Naidu (2006); Hwang et al. (2014) similarly find that inequitable conventions are more likely to be the stochastically stable equilibria of evolutionary models when the conventions themselves are less inequitable.

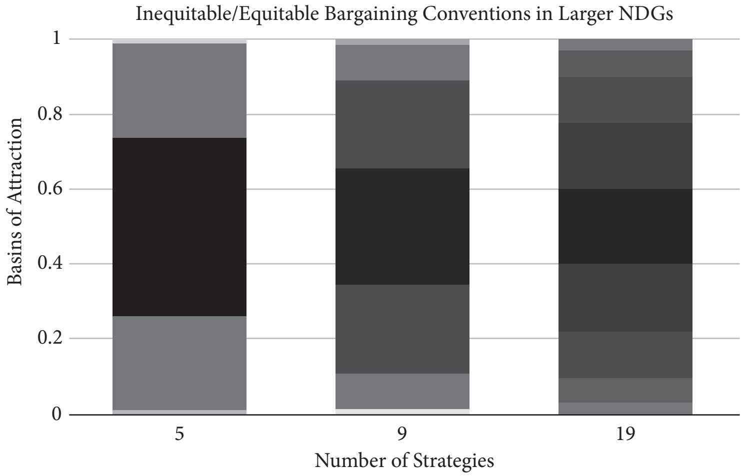

Figure 5.3 Basins of attraction for fair and unfair outcomes between types in the Nash demand game with various numbers of strategies

time. In many real-world cases, of course, fine partitions of resources are possible.[15]

Hopefully now the claim made at the very beginning of this chapter—that the preconditions to generate inequity of the second sort are very minimal—is starting to become clear. In the bargaining models we have just considered, inequitable conventions are expected to emerge between groups despite the possibility of an equally good fair convention, and despite the lack of any sort of asymmetry between groups that might explain why one and not the other should be disadvantaged. The preconditions for this emergence are 1) a strategic scenario well modeled by a bargaining game, 2) actors who learn to repeat strategies that work well for themselves and others, and 3) typing and type-conditioning. Notice that these conditions hold broadly in human groups.

[15] Of course, one might point out that as we increase the number of strategies, many of the possible outcomes are quite close to fair (i.e., one side receives a resource of 4.5 and the other 5.5). We might instead consider the expected payoff difference between the two groups to see whether inequity increases as partitions grow finer. On this measure, inequity is actually less serious for finer partitions in the models in Figure 5.3. (For Figure 5.2 there is no straightforward trend, though the worst inequity emerges when the low demand is 2.) This is to point out that there are different ways to measure inequity in this model, but the main point stands that once we add categories, inequity of some sort or another becomes commonplace.

5.2.1 *Is fair, fair?*

In the next section, we will start to look at models where there is some sort of asymmetry between groups to clarify the conditions under which a social group will be more likely to be disadvantaged by bargaining norms. First, though, a brief discussion about what fairness even means in bargaining models is in order.

What, exactly, does a medium demand consist in? Surely we are not trying to represent just those real-world cases where actors divide a cake (or some other resource) into exactly equal parts. Instead, we are trying to capture an approximately fair division of resources. Of course, what a fair division is might be quite different depending on the background context in which a game is played. As Bicchieri (2005) puts it, "Fairness, reciprocity, trust, and so on, are *local* concepts, in the sense that their interpretation and the expectations and prescriptions that surround them vary with the objects, people, and situations to which they apply" (76). If I buy ingredients and do the extensive work of baking a cake while you laze around, no one would expect me to split it 50–50 with you. It would be fair for me to keep the entire thing, and generous to give you a piece. A 50–50 split in this case would be a division falling heavily in your favor. What I am trying to highlight here is that what a Med vs. Med split corresponds to in a society will itself often be a culturally evolved notion. In our society, we think that people should possess those things that they buy or build (or bake). But we might have had other norms for what is fair, and many societies do (Henrich et al., 2001).[16] Experimental work on ultimatum games—a type of asymmetric bargaining game—show a surprising amount of variation cross-culturally in what is taken to be a fair split. In addition, framing effects will create within-culture differences in beliefs about fairness (Bicchieri, 2005). (For example, the addition of factors that imply one player is deserving will tend to skew percepts of fairness in their favor (Hoffman and Spitzer, 1985; Frey and Bohnet, 1995; Falk et al., 2008).) Note that these evolved standards play an important role. Without them, the sort of special coordination that the fair outcome provides is not possible. In order for fairness to be special, everyone has to agree on what fairness means.

[16] Fiske (1992) describes cases where authority rankings create perceptions of fairness for behaviors (like simply taking goods from group members) that seem objectively one-sided.

Now, though, suppose that a society can have norms of fairness where social categories come into play. Perhaps in a society what everyone (implicitly or explicitly) agrees to be "fair," for example, is that immigrants get less and natives get more in situations of bargaining. In such a society, a Med demand in a representative model could correspond to one group literally demanding 60% and the other 40% of a resource. On this interpretation, the extensive results supporting fair divisions in the Nash demand game described above do not necessarily tell us that fairness should be expected in human cultures at all. If asymmetries are exploited to redefine "fair," the payoff benefits of fairness are recovered to some degree in the new scenario. Actors still manage to coordinate in a correlative way when they meet, but the extra information provided by social categories allows for this coordination to involve asymmetric outcomes.

Sen (1987) pursues a similar argument in claiming that household bargaining models from economics ignore what he calls "perceptions" of dessert. He points out that divisions may be judged fair between parties as a result of social conventions that, objectively, are not.[17] Of course, one might rightly ask what objective measure of fairness is possible if "fair" itself is an emergent, conventional social norm. Sen (1987) identifies unfairness between genders with variable outcomes on measures like mortality, morbidity, health, income, etc., as well as "well-being," which he describes as, ". . . a person's functionings and capabilities: what he or she is able to do or be" (8). Where these diverge as a result of household divisions of goods and labor, he labels the division as inequitable. This, of course, requires assuming that our stated (if not always practiced) social norm that factors like gender and race should not determine dessert is a good one.

This is all to say that there are two different ways of thinking about the role types and typing play in inequitable norms. Either we can think of them as creating an asymmetry that allows inequitable divisions to arise (as in the models discussed here), or as creating an asymmetry that allows for a culturally evolved notion of fairness to emerge that treats types differently. I will not model this second possibility explicitly.

[17] Relatedly, he points out that in many studies women do a disproportionate amount of work contributing to a household, but are perceived as contributing relatively little.

5.3 Power

Nash (1950) was the one who first introduced a bargaining problem similar to the Nash demand game, along with his famous solution to it (which I will say more about in a minute). Just a few years later he produced another paper thinking about his bargaining model in a slightly different way (Nash, 1953). Suppose that before engaging in a bargaining scenario, two actors each have an opportunity to make threats about what will happen to the other should bargaining break down. If you and Dwayne Johnson are dividing your pie, say, you might threaten him with your annoyed glare should bargaining go south and he might threaten you with a crushed spine. These threats can then be taken as the disagreement points of the model, which, under this interpretation, are relabelled as *threatpoints*.

In this version of the game, power differentials between bargaining agents become salient. Suppose one player is able to issue a more credible threat than the other, for whatever reason. Because the less powerful player now stands to lose more should bargaining fail, the more powerful player derives an advantage.

The Nash bargaining solution starts with four axioms that he argues any solution to the bargaining problem should hold to.[18] From there Nash derives a unique division, which is that players should maximize $(v_1 - t_1)(v_2 - t_2)$, or the product of each player's expected payoff (v_i) minus their disagreement (threat) point (t_i). On this formulation, the player with the higher disagreement point, or ability to make a more credible threat, will get a higher value, holding other things fixed. One question that arises given the framework we are working with is the following: how does power influence not individual bargaining outcomes in one-shot games, but the emergence of bargaining conventions between social groups? This is the question that I will focus on for the rest of the chapter.

In order to address this question another question must first be answered and that is: what is power? There are many ways in which

[18] The details of these axioms are beyond the scope of this discussion. They are *Pareto efficiency*, essentially that there should be no resource gone to waste, *symmetry*, that if players have completely symmetric positions they should get equal amounts of the resource, *invariance to affine transformations*, that an affine transformation of a player's payoffs should not influence the outcome, and *independence of irrelevant alternatives*, basically that the removal of options that are not chosen should not change the expected outcome.

one person can have power over another person, and many ways in which members of one social group can have more power than members of another social group. For our purposes here, I will not give a conceptual analysis of power, or try to otherwise pin down this notion, but will instead look at a number of different ways of operationalizing power within a cultural evolutionary framework. As I will argue, these operationalizations fit well with at least some properties of social groups that we associate with power. And, as I will show, there are a number of different ways in which these operationalizations can impact the chances that certain social groups are disadvantaged with respect to bargaining norms. This part of the book draws heavily on joint work with Justin Bruner (Bruner and O'Connor, 2015).

The models we are about to consider are particularly important for the following reason: in almost every case, the social situation for members of two social categories will be asymmetric. When it comes to gender, we have already seen one important reason why—division of labor by gender means that men and women are always in asymmetric positions with respect to social roles. These small differences of inequity between genders, as we will now see, can feed into evolutionary processes that generate more serious inequity. In general asymmetry rather than symmetry is the rule in the social realm. If we have one group, there is only one way for a second group to be identical to it, and uncountable ways for it to be different. As we will see, these differences increase the chances that inequitable conventions will emerge.

5.3.1 Evolution with threats

Suppose that members of one social group tend to be able to issue credible threats to members of another social group to be carried out should bargaining between them fail. Scenarios that fit this assumption might include household bargaining between genders where there is a more serious threat of domestic violence for women owing to size dimorphism, or bargaining between classes when one class holds greater economic or political power than another, or bargaining between members of different races when there is a similar economic or political divide. (For an example where both race and class play such roles, consider the negotiations taking place between landowners and sharecroppers in the American South post-Civil War.) In Bruner and O'Connor (2015), where Justin

Bruner and I first explored evolutionary bargaining models with power differences between groups, we use the models to represent bargaining across the academic hierarchy, where members of one group (professors) can make choices that more seriously impact the careers of the other group (grad students). Notice that although the source of "power" in these examples is different—physical strength versus economic advantage versus asymmetric working conditions—in each case the difference creates potential for one side to lower the payoffs of the other.

Figure 5.4 shows an underlying game that might capture such a situation. We see a Nash demand game with Low, Med, and High demands, but now instead of both actors receiving the same disagreement point of 0, we have actors receiving threatpoints of t and T where we assume (without loss of generality) that $t \leq T$. In other words, if bargaining should go south, one player can lower the payoff of the other more effectively than the converse.

In an evolutionary model where $t = T = 0$, as we know, inequity can emerge, but it is equally likely that each type ends up at an advantaged equilibrium. As we vary the values of t and T, though, we no longer see this sort of symmetry. Let $t = 0$ and let T vary between 0 and 3.5. Figure 5.5 shows results for evolution in such a model. As before, there are three possible outcomes to this model—two discriminatory outcomes where one side demands High and the other Low, and one fair outcome. What we see now, though, is that as T increases, or as the difference between the threatpoints for the two groups becomes wider, three things happen. First, the likelihood that the fair outcomes is reached decreases. Second, the likelihood that the powerful group demands High goes up. And last, the likelihood that the less powerful group demands High goes down. When $T = 3.5$, 61% of simulations go to the outcome where the powerful type discriminates, and only 4% of simulations go to the outcome where the less powerful type discriminates.

		Player 2		
		Low	Med	High
	Low	4, 4	4, 5	4, 6
Player 1	Med	5, 4	5, 5	t, T
	High	6, 4	t, T	t, T

Figure 5.4 Payoff table for a three-strategy Nash demand game with threatpoints

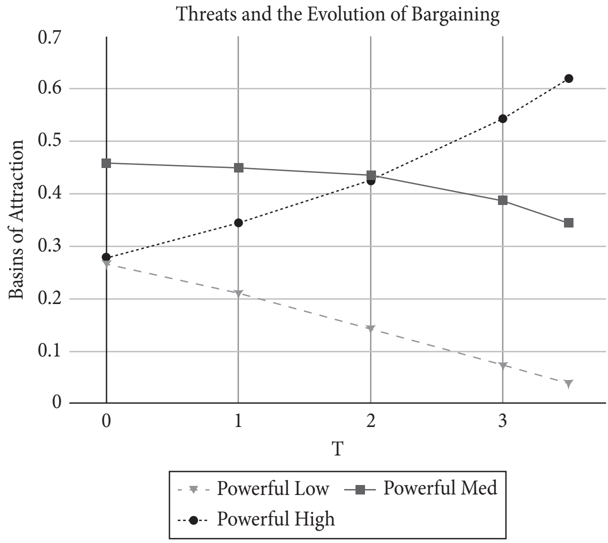

Figure 5.5 Basins of attraction for a model where one type has a higher disagreement point in the Nash demand game

From the Nash perspective, a higher threatpoint can translate into a bargaining advantage between individuals. Why, though, do high threatpoints translate to advantage in the emergence of bargaining conventions between social groups? The explanation for this phenomenon has to do with the incentives actors have for learning various strategies. For actors with a high threatpoint, the difference between receiving Low and getting the threatpoint is relatively small, meaning that there is relatively little incentive to learn to demand Low. On the other hand, demanding High is not as risky for them, and still generates the same rewards, so there is a greater incentive to learn to play High. This asymmetry creates a selective environment where the powerful group then tends to move toward demanding High. In this case, the details of the game matter little. A similar effect will be observed for Nash demand games with different demands, or finer partitions, for the very same reasons.[19]

[19] For example, see Bruner and O'Connor (2015). In LaCroix and O'Connor (2017) we show the robustness of this effect in models with different dynamics, finite populations, and cases where disagreement points are heterogeneous across the groups.

Note that there is another sort of threat that members of one social group might be able to wield that shifts payoffs in a different way, but still advantages the powerful group. Suppose that instead of issuing a threat contingent on bargaining breakdown, actors of one type threaten to punish their opponents for making high demands simpliciter. (Bowles et al. (2007), for instance, find that women are directly punished for making high bargaining demands.) Suppose that a powerful actor can decrease the payoff of their opponent by 2 should the opponent choose High. This induces a payoff table like the one in Figure 5.6. Notice that unlike changes to the threatpoint of the model, this sort of threat disrupts one of the Nash equilibria of the game. Now the strategy pairing where player 1 chooses High and player 2 Low is no longer an equilibrium. In an evolutionary model, this strategy pairing will no longer emerge as a stable population convention either. In this sort of threat scenario, then, one side can entirely eliminate the possibility of ending up at the disadvantaged outcome via a threat.

Notice that these models are applicable only to cases where, for some reason, the members of one social group are actually more able to change the payoffs of the other group. This is importantly different from the role that threats play in standard game theoretic models, as I will now elaborate.

In many game theoretic analyses, the way threatpoints impact outcomes is by changing the beliefs and expected payoff calculations of the players. Importantly, threats impact bargaining outcomes without ever having to be carried out! Of course, it is important that they be credible. Threats are not expected to have an impact on outcomes if the other player does not believe them. (If you tell Dwayne Johnson you'll crush his spine, for example, you will not end up with more pie.) In fact many economists have pointed out that carrying out threats is often costly (in terms of time,

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

Figure 5.6 Payoff table for a Nash demand game where one player carriers out a threat in response to a High demand

energy, social standing, etc.), and so once bargaining is over one is not incentivized to make good on a threat. This means that threats should only work if one is precommitted to them. Nash (1953) says that ". . . it is essential for the success of the threat that A be *compelled* to carry out his threat T if B fails to comply. Otherwise it will have little meaning. For, in general, to execute the threat will not be something A would want to do, just of itself" (130).[20]

This seems to lead to a problem for the inclusion of threats in evolutionary models. Unlike models of rational choice, threatpoints do nothing in an evolutionary scenario unless they are actually carried out. They must, in fact, lower the payoffs of some actors to impact how learning happens in cultural groups. The thing to consider, though, in assessing whether evolutionary threatpoint models are good ones is not whether it is rational to punish, but whether actors who are physically stronger, more economically powerful, or more politically powerful *do* in fact carry out threats in the real world.

The evidence suggests that, yes, humans carry out threats and punish one another. This is despite sometimes clear irrationality from a short-term payoff point of view. Evidence for this claim can be derived from empirical work. Domestic violence, for example, is almost always directed toward women and it has not been lost on researchers in sociology and economics that this sort of violence is often used to improve men's bargaining power.[21] Goode (1971) argues that men use physical threats to gain resources for themselves in the household. Eswaran and Malhotra (2011) use empirical data to argue that men in India use domestic

[20] Similarly Schelling (1960) says that "[t]he distinctive character of a threat is that one asserts that he will do, in a contingency, what he would manifestly prefer not to do if the contingency occurred" (123).

[21] Blood and Wolfe (1960), in their influential study of the family, argue that "[t]he dominance of men in marriage is often attributed to physical strength. While men do have more muscle power, any factor which operates purely in a biological fashion would be expected to have a universal effect. If superior musculature were the only reason for male dominance, we would expect men to dominate everywhere and in all times. The fact that they do not suggests that other factors must, as least, contribute to the picture" (Blood and Wolfe, 1960, 15). These authors reason from the incomplete dominance of men in the household to an argument that biology cannot be causally responsible for this dominance. Like the social-structural theorists discussed in the last chapter, they fail to recognize that cultural evolutionary processes need not be deterministic. The models here show how physical dominance can lead to preferential bargaining outcomes, even though this outcome is not determined.

violence to prevent women's autonomy for the purposes of household bargaining. (Rao (1998) uses modeling work to show how domestic violence might increase bargaining power.)

Experimental work also provides strong evidence that humans do, in fact, carry out threats and punishments in bargaining-type scenarios, even when doing so is costly to themselves. Güth et al. (1982), in a seminal study, had experimental subjects play the ultimatum game. This game is sometimes interpreted as a type of asymmetric bargaining situation because actors divide a resource. One actor offers some portion of a resource to the second, and the second has the option to either accept the offer or reject it. Rejection means neither player gets anything. Unlike the Nash demand game, the sequential nature of the ultimatum game means that there is a single *subgame perfect equilibrium*. (This concept is slightly different from the Nash equilibrium concept, and applies to sequential games. It assumes that actors play Nash equilibria of every *subgame*, or part of the sequential game.) Upon receiving an offer, the second player will do their best to accept it, no matter what it is. Knowing this means that the first player will make the smallest possible offer. The unique equilibrium of the game is then that the first player takes almost the entire resource.

What Güth et al. (1982) found was that experimental subjects behaved nothing like what this concept would predict. Instead, most offers were significant. Importantly, second actors regularly rejected offers that were too low. As the authors point out, "Subjects do not hesitate to punish if their opponent asks for 'too much'" (384). Here actors forgo taking payment in order to decrease the payoff of the other actor, or accept personal costs to punish. Subsequent experiments find cross-cultural variation in the ultimatum bargaining game, but confirm that actors punish norm violators, even when it is costly for them to do so (Güth and Tietz, 1990; Henrich et al., 2006). Henrich et al. (2006) find that, cross-culturally, humans are willing to pay to punish those who offer too little in the ultimatum game even if they are third-party bystanders.[22]

[22] Costly punishment has been observed under many other experimental paradigms as well, such as public goods games (Fehr and Gächter, 1999). Costly punishment can evolve for the ultimatum game under the replicator dynamics (Gale et al., 1995; Harms, 1997; Skyrms, 2014). Page et al. (2000) derive similar results for actors learning on a network.

If humans are willing to punish, even at cost to themselves, threatpoint models can tell us something about the emergence of bargaining behavior across types. In particular, they suggest that when members of one social group can wield a credible threat, and at least sometimes actually carry it out, they gain an advantage with respect to emerging bargaining conventions.

5.3.2 *Threatpoints without threat*

While I have just defended the real possibility that powerful social groups can gain bargaining advantages via threats and punishment, importantly in many real-world scenarios power differentials can translate directly to advantage, no threat needed. This is because differences in the economic, political, or social spheres of those in different social categories can often impact disagreement points directly.

A classic example comes from the economics of the family, and will be discussed in further detail in Chapter 8. Suppose that men are more economically advantaged than women within a society. When household bargaining breaks down, men need not make threats against women for the two groups to have very different disagreement points. The facts of the economic situation mean that women will tend to be at risk of poverty, and its attendant discomforts, should the household dissolve, while men will not.[23] This creates a scenario well represented by the threatpoint models just described. As discussed, in a situation like this, men should be expected to end up at a better outcome by dint of their higher disagreement points.

This sort of asymmetry between types is extremely prevalent in the real world. These scenarios arise whenever two groups engage in bargaining and one side generally has a better fall-back position. Especially relevant is the impact that different economic situations of social groups can have on bargaining norms. Whenever members of one social group tend to have more, this means that they are also likely to get more as a result of cultural evolution. This observation helps us start to understand one of the claims I made in the introduction to this chapter—that initial differences in power between types can compound into greater differences. An

[23] Sen (1987) also focuses on pregnancy as a biological factor that creates different disagreement points for women and men upon breakdown of a household, with the woman's disagreement point lowered by the physical facts of pregnancy.

economic advantage today can yield an economic advantage tomorrow via the emergence of bargaining conventions. As I will argue later in the chapter, this can create a sort of runaway process that leads to serious, persistent inequity between groups.

One further thing to note about this sort of structural power asymmetry, briefly mentioned earlier, is that it can arise via gendered (or class or race) division of labor. Suppose that actors in a society divide labor by type, and this creates benefits for both types, but in an inegalitarian way. Perhaps women in a society control crop production, and so have more direct access to the food supply. If women derive greater economic benefits as a result of this conventional division, they are expected to end up at preferred bargaining conventions in other arenas as well. These sorts of asymmetries may even emerge in a previously egalitarian regime and provide new power imbalances as technology changes. Imagine, for example, that ploughs are made much more efficient. This might mean that men in a plough-based agricultural society gain a sudden boost in disagreement points.

Before continuing, as a sort of robustness check, let us discuss models from Young (1993b). Young's framework, remember, involves actors from two different classes who randomly interact in a bargaining engagement. Actors always choose the strategy that yields the best expected payoff in response to a sampling of past play. This, remember, leads to the emergence of bargaining conventions just like those from replicator dynamics models, but the process of change appeals only to the bounded rationality of agents, and not to cultural imitation or individual learning. In this quite different evolutionary model, high disagreement points also advantage social groups, but in a way that is closer to Nash's original findings. Young finds that the stochastically stable equilibrium of these models is for actors to play the Nash bargaining solution, so that an advantage in disagreement point for one type will translate into an advantage in the unique, expected convention that populations evolve to play. Binmore et al. (2003) expand these results for different versions of best-response dynamics, and for some related coordination games. These results can increase the weight we give to claims about the cultural evolutionary impact of disagreement (threat) points on the emergence of bargaining conventions. Even under quite different assumptions, the benefits of high disagreement points for a bargaining social group are robust.

In LaCroix and O'Connor (2017) we use simulations of this sort to show how even one or two individuals with high disagreement points can improve the chances that one group ends up with a bargaining advantage. We also look at populations with distributions of disagreement points, and show that ones with higher means tend to be advantaged. These results, which introduce heterogeny into the two populations, demonstrate an extra point—when it comes to the emergence of bargaining norms, individual power matters less than average power across a group. An economically advantaged woman may still engage in extra household labor, because she is part of the social category "women" and this generally low-powered category has ended up conventionally disadvantaged.

5.3.3 *Outside options*

As I said at the beginning of this section, I will operationalize power in a few different ways to investigate the impact of power on bargaining conventions. To this point, power has been operationalized as a threat that impacts either the threatpoints, or the payoffs for high demands of actors. It has also been interpreted as a disagreement point that emerges from the asymmetric situations of social groups. Now I will consider another mechanism whereby power differences, especially those created by economic or social asymmetries, can impact bargaining conventions.

An *outside option* in a game allows actors to essentially opt out of a strategic scenario. In a bargaining game, an outside option means that rather than bargaining, an actor can instead engage in some sort of solo work that generates a predetermined payoff. In Bruner and O'Connor (2015) we show that when members of groups have different outside options, this impacts the basins of attraction for bargaining outcomes between them. Consider again household bargaining in a society where men make more money than women. In this case, men may be able to survive perfectly well without ever forming a household. Instead, they can use their solo income to support themselves. Women, on the other hand, might not have that option. In the case of landowners and sharecroppers, those who own the land may be able to use it in many different ways that do not involve a bargain, while sharecroppers depend on a bargain being struck for any sort of livelihood.

Figure 5.7 shows a Nash demand game with three demands—Low, Med, and High—and an outside option. In this case, player 1 gets X and

Player 2

Player 1		Low	Med	High	Outside
	Low	4, 4	4, 5	4, 6	0, Y
	Med	5, 4	5, 5	0, 0	0, Y
	High	6, 4	0, 0	0, 0	0, Y
	Outside	X, 0	X, 0	X, 0	X, Y

Figure 5.7 Nash demand game with different outside options for each player

player 2 Y for solo work. For simplicity's sake assume that $Y \leq X$.[24] This figure should help make clear the difference between asymmetries in outside options and disagreement points. Disagreement points are reached only when players bargain and fail, outside options are reached only when players choose not to bargain. Of course, in many scenarios actors with high disagreement points will also have high outside options and vice versa. This will occur if, for instance, one class of actor has better access to jobs, and so has both a better non-bargaining outside option, and a better fall-back position if bargaining fails.

Changing the values of the outside options in this game will alter the Nash equilibria. Why? If an actor expects to receive a low bargaining demand, and her outside option provides a better payoff, she can unilaterally switch strategies and do better. If an outside option is higher than the Low bargaining demand, the equilibrium where the actor with that option receives Low disappears. If an outside option is higher than the Med demand, the fair equilibrium disappears.

For simplicity's sake, assume $Y = 0$. Let X vary between 0 and 6. When $X \leq 4$, all three bargaining outcomes are still Nash equilibria. When $4 < X \leq 5$, only the bargaining outcomes where player 1 receives Med or High remain Nash equilibria. When $5 < X \leq 6$, only the bargaining outcome where player 1 demands High is a Nash equilibrium. The elimination of these equilibria means that the corresponding outcomes no longer emerge in evolutionary processes. Importantly, this means that the type with the higher outside option is expected to end up at an advantaged

[24] Wagner (2012) uses this game to model actors who first play a stag hunt and then a Nash demand game to divide the fruits of their labor. On this interpretation, the outside option is hare-hunting—individual work that yields some decent payoff. The other options represent attempts to work together to yield a higher joint payoff, but where this adds the requirement that actors then must bargain either explicitly or implicitly to divide their group yield. More on this interpretation in Chapter 7.

bargaining convention by dint of having that outside option. If we return to the sharecropper example, we should never expect landowners to accept a portion of crop yield that pays them less than, say, letting goats graze on the land. In the case of household bargaining, if men can have perfectly successful solo households and women cannot, we should expect conventions to emerge where men do no more household labor than they would on their own.

5.4 Compounding Power

In the last section we saw that power differences between types can translate to bargaining advantage over cultural evolutionary time. This observation begins a narrative that will be expanded in this section. One way to gain economic power in the first place is through successful bargaining. This means we can observe circular patterns where economic advantage lead to better bargaining outcomes, leads to further economic advantage. Sen (1987) focuses on a similar phenomenon, which he calls "feedback transmission" in the context of rational choice models. As he says, "The asymmetries of immediate benefits sustain future asymmetries of future bases of sexual divisions" (27). His compounding of power and bargaining advantage happens when social conventions lead to improved bargaining outcomes for individuals in the household, though he does not analyze the emergence of these social conventions in the first place.

We'll explore the possibility of this sort of feedback loop using a toy model—one that tells a how-possibly story about the accumulation of power for one social group. Suppose that two types play a Nash demand game where they have different levels of power operationalized as varying disagreement points. Assume that there are multiple arenas of bargaining where types develop conventions of interaction. And, in particular, assume that in each of these arenas, the disagreement points for each type are determined by their success in previous bargaining interactions.

For our model, let's use the game displayed in Figure 5.4. Demands are 4, 5, and 6, and the disagreement points for the two sides are T and t. To instantiate the assumptions just described, about past bargaining success influencing current power, we'll start simulations with $T = t = 2$. Then after conventions develop, if one type is demanding High, we'll increase their disagreement point, and decrease the disagreement point for the other type.[25] If they make equal demands, the disagreement points will remain the same. The question is: does this carry-over of bargaining outcomes to future disagreement points mean that one side tends to keep power for themselves?

Suppose we run this model without the disagreement-point adjustments just described. In this case, over the course of many arenas of bargaining the two sides will each be equally likely to end up demanding High (in about 25% of simulations each), though the fair outcome will be the most likely one (about 50% of simulations). If we add the adjustment, things change. One side tends to gain an advantage fairly early on in the simulation and maintain this advantage. Each side is equally likely to gain the advantage, but once they do, they'll end up demanding High for a large percentage of simulations (around 70%), making fair demands less often (around 30%), and essentially never demanding Low (less than 1% of simulation outcomes). In other words, we see that a process by which current power feeds into the emergence of bargaining conventions can preserve power for one group over a long period of time and across bargaining arenas.

There is one more observation I'd like to make here, addressing cases where there are some innate asymmetries between groups. This is particularly germane to understanding gender and power, where differences in strength between men and women could have acted as an initial asymmetry that allowed men to generally end up with greater power across societies and over time. Suppose that in many different cultures genders developed bargaining conventions across arenas and over time, and suppose that previous success in bargaining impacted later interactions as above. But suppose that one side tended to have a slight power advantage to start with. To capture this possibility, I'll run the models described above, but many times, seeing each time which side tends to

[25] I do this by taking the difference between 4 (the Low demand) and the current higher disagreement point. I then increment/decrement the disagreement points by 10% of this difference. So say men currently have $T = 2.2$ and women have $t = 1.8$. If men end up demanding High, I calculate $4 - 2.2 = 1.8$, and change the disagreement points by .18 so that they are now $T = 2.38$ and $t = 1.62$. This method is somewhat arbitrary. I use it because it keeps the disagreement points from ever exceeding the Low demand or going below 0. Note that this keeps the powerful side from accumulating too much power since the outcome where they demand Low always remains an equilibrium of the model. On the other hand, as the disagreement points become more disparate, they change more slowly, meaning that the less powerful side will have a harder time recovering lost ground.

end up empowered. And I'll also vary the initial disagreement points in the model from equal (2 and 2) to quite unequal (3 and 1).

As we see in Figure 5.8, introducing an initial asymmetry to the model makes it more likely that one side gains and keeps power over time. The greater the initial asymmetry in disagreement points, the more likely it is that the more powerful side remains that way.

This set of models shows how random initial events that advantage one group over another can ultimately lead to a sustained power difference between groups, and to enduring economic advantage for the powerful group. One thing this result tells us is that when we see persistent inequity between groups, the explanation need not always appeal to anything in particular about those groups. Random chance can advantage one group over another, and the dynamics of bargaining conventions can sustain that advantage indefinitely.

This model relates to sociological work on the concept of "cumulative advantage," or the idea that "current levels of accumulation have a direct causal relationship on future levels of accumulation" (DiPrete and Eirich, 2006, 272). For example, Gould (2002) and Lynn et al. (2009) provide models where displays of deference to other individuals create status hierarchies which are not based on merit, because current status deter-

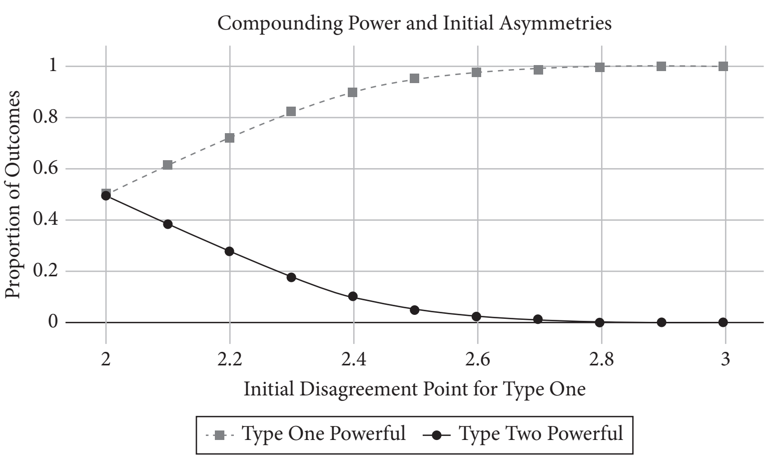

Figure 5.8 Proportions of outcomes where an initially powerful group maintains power for a Nash demand game where disagreement points are determined by past success

mines level of deference from others. Merton et al. (1968); Merton (1988) outlines how academic communities that use past success as a predictor of future productivity will develop hierarchies only loosely based on merit, and significantly impacted by chance initial conditions. Some of this work addresses status and wealth differences between those in different social groups (DiPrete and Eirich, 2006). Tilly (1998) focuses on social categories, and illustrates detailed causal processes by which members of one group can use power to compound advantages in terms of resources, much like processes in the models just described. Okin (1989) makes clear how gender inequities in the context of marriage can lead to cyclical processes which prevent women from gaining power. (See also Cudd (1994).) Note that these approaches focus on how different conditions caused by social group membership lead to inequitable outcomes for individuals, but, unlike the model here, have not typically focused on how the existence of types simpliciter is a causal factor in the emergence of inequity.

Notice that these observations about the compounding of power have meta-implications regarding inequality and innate ability. Sometimes gender and racial inequality are taken as evidence that women or members of some race must be less innately talented than men or members of another race. But if we expect categorization to lead to inequity, and we expect inequity to generate further inequity, we need not appeal to innate ability in explaining these inequities in the first place. They are easy to get.[26]

• • •

In this chapter we began to explore the dynamics of inequity of the second, more pernicious sort. There are a handful of takeaways that the models discussed underscore. First, although previous evolutionary models of bargaining have focused on explaining fairness, once types

[26] For example, in *The Origin of Species* Darwin argues that "[t]he chief distinction in the intellectual powers of the two sexes is shewn by man's attaining to a higher eminence, in whatever he takes up, than can woman—whether requiring deep thought, reason, or imagination, or merely the use of the senses and hands. If two lists were made of the most eminent men and women in poetry, painting, sculpture, music (inclusive both of composition and performance), history, science, and philosophy, with half-a-dozen names under each subject, the two lists would not bear comparison" (361). In other words, Darwin takes the "eminence" of men to be evidence that women are inferior to men. These models indicate that social position provides little evidence with regard to the qualities or abilities of those in power.

are introduced to an evolutionary model, unfairness is just as likely an outcome. This sort of inequity requires very little to emerge beyond the simple fact of typing in a group. Furthermore, we have seen how power for a social group can lead to advantage when it comes to bargaining conventions. Groups with power are more likely to end up with resources, groups with less get less. And lastly, these sorts of processes can lead power and resources to concentrate for the same groups meaning that the same sorts of inequity persist over time.

In the next chapter we'll look at a different sort of causal variable which can impact inequity—asymmetries in the learning environments of different types. Before moving on though, a word about the epistemic role of the models presented here and in the next chapter. All of these models instantiate causal variables that mimic those seen in the real world. For this reason, it is plausible that the processes observed in these models can tell us something about what really happens when groups learn to divide resources. They are how-potentially models. At the same time, they are how-minimally models. Even if their processes don't fully match the real processes that generate divisional inequity, they give us counterfactual information about how easy it is to generate this sort of inequity that, as I will elaborate in Chapter 9, may be helpful in thinking about intervention.