3

Cultural Evolution with Social Categories

Cultures that use the plough for agriculture are more patriarchal than those that use the hoe, but that is not the end of the story.

In communities descended from those that used plough-based agriculture, women tend to work outside the home less, and there is greater gender inequality (Alesina et al., 2013). In other words, these cultures have developed norms of gender inequality, which have staying power over cultural evolutionary time, even under new circumstances where there is no current ecological reason for men to be more powerful.

In the last chapter, I discussed type-conditioning, and described how type divisions can facilitate efficient outcomes in complementary coordination games that are unavailable to homogenous groups. In this chapter, I move on to explicitly evolutionary models. I follow authors like Young (1993a), Skyrms (1996), and Bicchieri (2005) in thinking of these models as capable of representing the emergence of conventions between actors in different social groups. What we will see is that from a starting point of uncoordinated behavior, groups move toward states where everyone follows unified patterns in a way that tends to lead to successful outcomes. Once groups have arrived at these coordinated behaviors they tend to remain there, so that the patterns of behavior persist over time.[1]

[1] Of course, there are mismatches between model and world. Actors in these models will uniformly adopt the same sorts of behavior, whereas real-world conventions and norms are followed to greater or lesser degrees, and almost never universally. These real patterns, though, involve what Young (2015) calls *compression*. The vast variety of human choices are compressed so that normative or conventional behaviors happen much more than would be expected if these patterns had not evolved culturally. The models capture this idea—that from an uncoordinated state emerges relative coordination.

The Origins of Unfairness: Social Categories and Cultural Evolution. Cailin O'Connor, Oxford University Press (2019). © Oxford University Press.
DOI: 10.1093/oso/9780198789970.003.0004

As we will see, when a group develops norms and conventions, the addition of social categories radically changes the evolutionary process. Entirely different sets of norms are expected to emerge in a group with social categories from those emerging in a homogenous one. In particular, groups with categories reach the inequitable, but efficient, outcomes described in the last chapter. Thus we manage to answer the explanandum for social scientists regarding gendered division of labor. Why do all cultures divide labor by gender? In situations where groups encounter complementary coordination problems, gender improves efficiency by dint of breaking symmetry. This efficiency drives cultural evolutionary processes to select norms where gender plays this role, whether or not these norms are equitable.

I will begin by describing a different set of models entirely. In economics, the question of why gendered division of labor is so ubiquitous has also been addressed using models where actors must take complementary roles to succeed. As I argue, these models capture important insights about the strategic situation that generates gendered division of labor, but miss some crucial aspects of the process by which groups end up divided by gender. I will then describe and discuss the strategy used here to model cultural evolution, including some non-trivial modeling choices. We'll move on to look at what happens to groups in cultural evolutionary models who play complementary coordination games, noting, in particular, the differences between outcomes in groups with social categories. Although one might question some of the modeling choices used to generate these results, as we will see, the results are supported by a host of related models from across the social sciences. In other words, they are highly robust, supporting the claim that we should expect real cultural evolutionary processes to take advantage of social categories to solve coordination problems. I finish by showing how radically the addition of social categories changes the evolutionary process even when complementary coordination is not occurring—for models of correlative coordination and prosociality.

3.1 Rational Choice and Division of Labor

Economists, starting with Becker (1981), employ models of the marriage market to explain widespread household division of labor. As mentioned, these models are in some ways similar to the ones in this book, but are based in rational choice (rather than cultural evolutionary processes), and put weight on the idea that individuals choose to specialize in certain types of labor to impress potential mates. Here I will not survey this literature, but will discuss some representative work from it.[2]

Hadfield (1999) presents a model of a population with two types where these each form households with only the other type. It is assumed that actors invest in skills before marriage, and that they may choose between two types of skill sets—representing those appropriate for home and market labor. This investment translates into payoffs after marriage where those with strong skills in either area will generate better outcomes for the household. Lastly, actors choose partners based on whether they have complementary skills. As she shows, the only plausible equilibria of this model are those where all males specialize in one of the two forms of labor, and all females in the other form. This is because actors need to make choices about skill specialization well ahead of when the actual coordination interaction happens. The benefits of skill learning mean that households do best when each individual fully specializes. Gender acts as a symmetry breaker which allows all individuals in the population to successfully commit to a skill that they know will pair well with the skill their future partner has learned. In other words, benefits of coordination, along with a desire to cultivate skills that will attract a mate, can lead to complete division of labor even if the types are modeled as otherwise identical.

These models, much like those here, show that when specialization is beneficial, and when types meet for interaction, society-wide patterns of division of labor are stable equilibria. The difference from evolutionary models is that the story of how actors come to reach these equilibria is not particularly filled in, and when we try to fill it in using the assumptions of the model, things look implausible. In order to reach such an equilibrium, on a rational choice model, actors must decide whether to specialize based on expected returns. They must each predict what the other type will be learning, and so what the other type will prefer on the marriage market, in order to decide what they themselves should learn to do. Then they must

[2] For a non-exhaustive list of other work in this area see Francois et al. (1996); Danziger and Katz (1996); Echevarria and Merlo (1999); Peters and Siow (2002); Baker and Jacobsen (2007); Nosaka (2007).

choose a partner based on specializations in a way that maximizes output. And furthermore, these calculations must be performed generation after generation anew. Notably there is no reason, in these models, that from one generation to the next the genders should not switch which role they perform.

Evolutionary models put a lower burden of rationality on actors by supposing that social learning and individual learning are doing the work to drive populations toward division of labor equilibria. Furthermore, the assumption that individuals are choosing spouses primarily based on their marketable skills, and developing their own skills in anticipation of this is dropped. This is to say that the preconditions in these models for the emergence of division of labor by type are similar, but the processes by which this division of labor is attained are very different. Evolutionary models offer a more natural representation for how this division arises, as it is clear that social learning is an important factor when it comes to gender socialization and division of labor. They also explain why we see stable patterns for the division of labor emerging in a society and persisting across generations. (In the case of the plough, this is exactly what we see—social transmission maintains a norm over the course of several generations.) And lastly, as we will see in the next chapter, evolutionary models can tell us something about why some patterns of division of labor seem to be highly conventional (basket-making) others not very conventional (big-game hunting), and others in between.

3.2 Cultural Evolution and Dynamics

As mentioned, the goal of this chapter is to look at what strategies evolve in populations with types playing complementary coordination (and some other) games. To answer this question with respect to human societies we must ask: how do cultural evolutionary processes happen? And: how should we model them? I will start by describing the approach that evolutionary game theory takes in general, and then fill in the details of how this approach will be used here to represent cultural change.

3.2.1 The evolutionary game theoretic approach

The game theoretic approach to understanding and predicting behavior, described in Chapter 1, can be contrasted to the evolutionary game theoretic approach. In evolutionary game theory it is assumed that actors slowly develop strategies over time, usually in the context of an interacting population or group.[3]

This methodology was developed in biology, starting with Maynard-Smith and Price (1973), to represent the evolution of strategic behavior in animals.[4] It was subsequently adopted by social scientists, and eventually philosophers, to model the cultural evolution of human behavior. Actors in biological models leave behind more offspring when they are strategically successful. Actors in cultural models update strategies based on whether they do or do not work in practice. There are various ways this is done. One common method is to assume that actors imitate successful or influential peers. (More on this shortly.) Another way is to assume actors simply repeat what has worked for them in the past. It should be noted that cultural evolution, here, simply refers to change over time in behaviors of human actors. We do not need to invoke memes, or any cultural analog of genes, nor do we need to answer thorny questions about information transfer to suppose that human cultures change in regular ways based on the successes of various behavioral choices.

In evolutionary game theory, all these types of change are modeled by applying what are called *dynamics* to games. Dynamics are rules for how strategies are updated based on their past success (and sometimes other factors). Below I will describe the *replicator dynamics*—the most commonly used model of change in evolutionary game theory, and the main process I will use to model cultural change.

It is important to keep distinct game theoretic and evolutionary game theoretic models. Of course, evolutionary game theory is grounded in, and deeply related to, game theory. This connection can at times be misleading, however. Games, in the two contexts, are formally similar, but often need to be interpreted in different ways. In particular, under the game theoretic interpretation, as discussed, payoff in a game represents utility. When games are employed in an evolutionary model, however, payoffs simply specify, based on the dynamics, how evolution happens. For example, payoffs determine whose strategies are most imitated by

[3] For a more in-depth treatment of this paradigm, see Weibull (1997), Gintis (2009), or Sandholm (2010).

[4] Well before this development game theorists were starting to think in terms of dynamic solutions to games, rather than static ones (Brown, 1951; Robinson, 1951).

group members, or who has what number of children, or, for some of the models in this book, they determine how quickly actors learn.

The disanalogy means that there are two ways to interpret evolutionary models. The first continues to associate payoff with success, or preference, or happiness, and further assumes that cultural evolution also tracks these things. We learn to do things that made us, and others, happy or well in the past. The second sort of interpretation pulls apart evolution and preference. For instance, the sorts of behaviors that make one a successful business person lead to prominence in an institution and increased social imitation as a result. But this prominence need not be associated with happiness or pleasure for a "successful" actor. In this case, utility and evolvability pull apart, and an evolutionary model should generally be thought of as tracking the latter.[5] Throughout the book, I generally use the first interpretation for cultural evolutionary models—that actors are learning to do things that benefit them and/or that they prefer.

Predictions and explanations from evolutionary game theoretic models often differ dramatically from those provided by standard game theory. Although evolutionary models are used for myriad explanatory purposes, there are a few particularly salient roles they have played in the literature. First, they have often been used to explain seemingly irrational human behaviors. Many theorists have used evolutionary game theoretic models to explain the prevalence of altruism in the biological and social realms, for example, despite the irrationality of altruism.[6] Another way evolutionary models have been successfully employed is to solve what are called 'equilibrium selection problems'. Sometimes rational choice models predict a set of possible equilibrium behaviors, while evolutionary models show that only one of these is likely to arise in an evolutionary context. Young (1993b), for example, predicts the emergence of fair bargaining equilibria, rather than inequitable ones, in an evolutionary model. Evolutionary models can also give deep insights into the conditions

[5] An excellent real-world example comes from the advent of agriculture. Early agriculturists seem to have been generally miserable. They suffered from poor nutrition, worked harder, and were less healthy than the hunter/gatherers that directly preceded them. Agriculture spread rapidly through human groups nonetheless in part because it increased the number of children humans could raise (Armelagos and Cohen, 1984; Armelagos et al., 1991; Steckel and Rose, 2002).

[6] This literature is truly massive, but for a few examples, see Nowak (2006); Alexander (2007).

under which behaviors can emerge in a way that is often not possible using rational choice models Alexander (2007), for example, shows how social interaction with neighbors can promote the emergence of prosocial behaviors in models (like the stag hunt) with multiple equilibria.

The evolutionary models in this book will especially play the last sort of epistemic role. As we will see in Chapter 4, knowing something about how likely various equilibria are to arise in coordination games will help inform gendered division of labor. In Part II of the book, the evolutionary framework will help us analyze the conditions under which equitable and inequitable outcomes are more or less likely to arise. In Chapter 9, the evolutionary perspective will be crucial to understanding how to intervene on inequitable conventions.

3.2.2 *Modeling cultural evolution*

As I elaborated in the Introduction, cultural change happens via a number of processes—social learning from peers, parents, and other teachers, individual learning as a result of experience in the social sphere, and rational choice, at very least. Furthermore, each of these processes may happen in different ways. Individual learning, for example, is itself a process that is updated over the course of a human lifetime. There are many versions of each sort of social imitation, and these, too, may be updated by agents who learn to learn better.

I will mostly abstract away from these complex processes by using a few simple models of evolutionary change to represent cultural evolution. Most of the models I will discuss involve change via the replicator dynamics, mentioned above. These dynamics assume that within a population better strategies will proliferate and poorer strategies will die out. Furthermore, strategies proliferate, under these dynamics, in proportion to how much they benefit those who employ them. This type of change can be thought of as underlying many evolutionary processes—behaviors that do well continue to exist and expand in number, behaviors that do not fail. Weibull (1997) shows that the replicator dynamics can be explicitly used to model cultural change via differential imitation of successful group members, a process that regularly occurs in human societies (Lancy, 1996; Fiske, 1999; Henrich and Gil-White, 2001; Henrich and Henrich, 2007; Richerson and Boyd, 2008). And Börgers and Sarin (1997) and Hopkins (2002) show that the replicator dynamics can act as a successful model of individual reinforcement learning, which has also been observed to guide the dynamics of human behavioral change (Thorndike, 1898;

Herrnstein, 1970; Roth and Erev, 1995).[7] Many previous authors have used the replicator dynamics to represent cultural change, often with explanatory success. The Appendix describes the equations that govern these dynamics.

The replicator dynamics, in their basic form, represent a purely *adaptive* process. The reason populations shift in their composition is that some traits are more successful than others, and there is a process by which success translates into prevalence. This is obviously an idealization away from real cultural processes which are sometimes driven by non-adaptive processes such as prestige bias, where actors copy prestigious group members regardless of the success of their behaviors, conformity bias, where actors simply copy popular behaviors, and similarity bias, where actors copy those they judge similar to themselves (Henrich and McElreath, 2003; Boyd and Richerson, 2004). In addition, these dynamics will miss "attraction" effects, where because of psychological biases or ecological conditions populations will tend to progressively move toward particular cultural variants regardless of adaptiveness (Sperber and Sperber, 1996; Claidière and Sperber, 2007; Claidière et al., 2014). Furthermore, because the replicator dynamics are purely adaptive, their simplest form does not include any representation of stochastic effects on cultural evolution. Of course, real populations are messy and random. Perfectly good behavioral variants may die out for no good reason, and poorer variants may become prevalent along the same lines.

Should all these idealizations be concerning? The modeling strategy here is to represent a generic selection process in a large population. This should capture the underlying, general direction of many of the processes of learning/cultural evolution mentioned above, while avoiding the sticky problem of how to accurately represent these various, interacting processes. It seems quite likely that for some cases, this generic process will not be a good representation of the cultural evolutionary path of a real population as a result. In other cases, it will be a good representation, and the simplicity of the models will allow us to gain explanatory clarity when it comes to the emergence of conventions related to gender and other social categories. Furthermore, even in cases where

[7] For both these dynamics, the replicator dynamics are the *mean field dynamics*, meaning that they track expected change in these models if the stochastic elements of change are averaged out.

the details of a real population do not match the models well, the models can help us understand the underlying selective pressures that such populations undergo, if simultaneously ignoring other sorts of change. And lastly, in order to ameliorate worries about the modeling choices, I employ "robustness checks" throughout the book. This is a technique that involves altering various factors of models to ensure that the key insights still emerge even under different assumptions. (More on this shortly.)

3.2.3 Learning from those like us

There is one more issue to tackle regarding modeling choices. In the models I will present shortly, one assumption is that everyone imitates those in their own social category, rather than other social categories.[8] As we will see, this assumption is not necessary to generate the main results, but it is still worth asking, is this realistic? To answer this question, I will describe the last set of empirical results related to gender—those on same-gendered cultural transmission.

When it comes to gender it is clear that separate modes of learning are happening. If they weren't, we wouldn't see such significant behavioral differences between men and women. (We know these differences cannot be completely innate, since there is significant cross-cultural variation.) Children are socialized into gendered behavior very early. Usually by age 2 or 3, they can say which gender they belong to, and shortly thereafter begin the long process of learning behaviors appropriate to their gender (Thrall, 1978; Bem, 1983; Basow, 1992; Wood and Eagly, 2002; Lippa, 2005; Kinzler et al., 2010; Kamei, 2010). Both adults and peers involve themselves in gender socialization, with adults directly teaching children proper gendered behavior, children using same gender models for imitative learning, and peers encouraging proper gendered behavior and discouraging inappropriate behavior. (See Wood and Eagly (2012) for an excellent review.) Henrich (2015) reviews a broad range of studies that have found social imitation of same gender types for many behaviors. Losin et al. (2012) find that own-gender imitation, but not other-gender imitation, of irrelevant gestures activates reward circuitry in the brain.

[8] Relatedly, Tilly (1998) identifies emulation, by which he means the copying of established behaviors, as a key mechanism leading to the emergence of inequity between social categories.

Henrich and Henrich (2007) point out that selective processes should favor copying those of the same gender and same ethnicity, for the very reason that doing so improves uptake of appropriate social roles and behaviors.

Proper gendered behavior, in all the arenas discussed, is subsequently enforced via gossip, ostracism, open criticism, and sometimes more serious forms of punishment (West and Zimmerman, 1987; Butler, 1988; Lorber, 1994). Fagot (1977) finds, for example, that among school-aged children peer criticism is harsh for cross-gender behavior. Glick and Fiske (2001) find that experimental subjects hold approving beliefs about women who conform to gender stereotypes and hostile beliefs about those who do not. The adoption of signals of gender membership in humans is also normatively regulated in many cultures. Men and women who adopt the wrong signals are subjected to social sanction (Garfinkel, 1967; West and Zimmerman, 1987).[9] (See Wood and Eagly (2012) for an extensive overview of the literature on social sanctions and gender.) All these sorts of social punishment should enforce same-gender learning, and increase the relevance of social models that involve same-gender imitation.

There is a worry here, which is that same-gender imitation, and other mechanisms for enforcing proper-gendered learning, are surely at least in part a response to the existence of gender roles and norms. In other words, while the models assume this sort of learning in order to get gendered division of labor, perhaps that is putting the cart before the horse. As I will point out at the end of Chapter 4, there is actually a bundle of social features that must be in place to get gendered division of labor (including own-gender learning). A full account of how groups manage to get all these features in place at once is beyond this book, but I will say a bit more there about this chicken and egg issue.

[9] To share a personal anecdote—when I took my twin girls for their three-month doctor appointment, I dressed one of them in a blue-striped onesie. The doctor's receptionist exclaimed, "He's so cute! Oh, but look, on his chart it says he's a *girl*." When I explained that my daughter was a girl she apologized for my mistake, saying, "When dads dress them, they just put anything on." In this case, she took the gender signaling so seriously that she trusted it over a medical chart. And upon discovering that I had failed to perform the correct gender signaling act, she felt the need to diffuse my expected embarrassment by trying to foist my lapse in normative duties onto my husband!

3.3 Evolving to Solve Complementary Coordination Problems

Lewis (1969) describes situations where precedent and repeated interaction lead to joint expectations between actors for solving coordination problems. Schelling (1960), meanwhile, points out that with respect to such solutions, "[t]he fundamental psychic and intellectual process is that of participating in the creation of *traditions*" (106, emphasis his). Although both thinkers were focused on a rational choice perspective, the processes they describe are essentially dynamic.

What happens when homogenous, two-type mixing, and perfectly divided populations evolve while the actors in them play complementary coordination games? In each case, the outcome is very different. I will start by discussing evolution in homogenous populations, and then move on to perfectly divided ones. The two-type mixing population results are a sort of combination of the other two, so I will save those for last.

3.3.1 Homogenous groups

Let us imagine a homogenous group where individuals randomly meet one another and engage in coordination problems. Over time people imitate those who seem to be doing well vis a vis coordination, and repeat behaviors that have worked for themselves in the past. Slowly, individuals develop patterns that tend to work for them.

For reference, Figure 3.1 shows the dancing game. As discussed, a homogenous group playing this game does very poorly if all members play either of the available strategies. When a group playing this game is evolved using the replicator dynamics, for this reason, they end up with a mix of strategies. In particular, regardless of what individuals start doing, a group playing this game under the replicator dynamics will end up with half of its actors playing A and half B.

Figure 3.2 shows what is called a *phase diagram* for a single population playing the dancing game and evolving according to the replicator

		Player 2	
		A	B
Player 1	A	0, 0	1, 1
	B	1, 1	0, 0

Figure 3.1 A dancing game

Figure 3.2 The phase diagram for a single population playing the dancing game

dynamics. The line in this diagram represents all the possible states of this population, from every member playing A (on the far left) to every member playing B (on the far right). Each other point on the line represents a unique population proportion—.37 play A and .63 play B, for example. There are three dots on this line, and each represents what is called a *rest point* for the replicator dynamics. At these points, the population will stop evolving. The two rest points on the ends represent all A and all B. At these points, the population does not change because it is assumed that there is no cultural model to copy from of the other sort. (Or in the genetic case that no genes of the other sort exist to spread in the population.) These rest points are unstable, though, in the following sense. If the population is perturbed from them, even a tiny bit, it will be carried away from the rest point. (A "perturbation" of this sort might occur when individuals experiment with other strategies, or err, or in a biological population perturbations might be the result of mutation.) This is what the arrows on the line represent—the direction of change for the population at states represented by that area of the phase diagram. To the right of all A, the population moves toward a state with more Bs. To the left of all B, the population moves toward a state with more As.

The central rest point, represented by a filled-in dot, is stable in the sense that if a population is perturbed, it will return to this state. Imagine, for example, moving a population away from this rest point a bit to the right. The direction of change carries it right back. This particular point is *asymptotically stable*, meaning that for some set of other population states, the dynamics will move the population toward this state and then remain there.[10] For the models we will look at (though not every possible evolutionary game theoretic model) the rest points we will be interested in are the asymptotically stable ones. These represent the outcomes that we expect real evolving populations to end up at.

[10] This particular rest point is actually *globally asymptotically stable*—not only do other population states evolve to this one, but every other possible population state (except the other rest points) evolves to this one.

The take away here is that, under the replicator dynamics, in every case the group in this model will evolve toward the state where half of all actors play A and half play B. (Note that this outcome corresponds to the population level equilibria described in the last chapter.) This sort of population state, where some actors use one strategy and some another, is called a *polymorphism*.[11] For a game with different payoffs, like the leader–follower game, the actual proportion of As and Bs will be different, but in general any complementary coordination game will end up with a mix of the two available strategies. As Henrich and Boyd (2008) point out, if we use this model to represent division of labor, this could correspond to a case where everyone does some of each available job, rather than splitting jobs between actors.

One natural question is whether polymorphic outcomes of this sort really occur in the biological and social world. There are many documented cases of real-world polymorphisms in the biological world.[12] When it comes to coordination, though, human populations usually seem to to find better solutions than polymorphic population states. We tend to make use of the sorts of solutions identified by Aumann and Maynard-Smith, and, of course, of types and type-conditioning to improve coordination

3.3.2 *Perfectly divided groups*

Now imagine a group with both men and women playing a dancing game. They only meet each other for interaction, and over time women imitate the successful actions of other women and men of other men.[13] Slowly they adopt strategies that work for them.

Perhaps unsurprisingly, the efficient population equilibria described in the last chapter are what evolve in this case. One type always learns to play

[11] Alternatively, we can interpret this outcome as one where every actor plays the same strategy, which is to mix between both possible actions, taking each 50% of the time. For our purposes the interpretory difference will not matter.

[12] For example, Sinervo et al. (1996) describe populations of side-blotched lizards where males display one of three throat colors, each associated with a different mating strategy. These strategy/color pairings persist side by side analogously to the three strategies in the evolutionary game theoretic version of rock-paper-scissors. See also Morgan (1980); Kaitala and Getz (1995).

[13] As discussed, social learning often involves own-gender imitation, but this is by no means a hard and fast rule. In section 3.4 I show how a model that makes different structural assumptions, but does not involve own-gender imitation, yields similar results to those here.

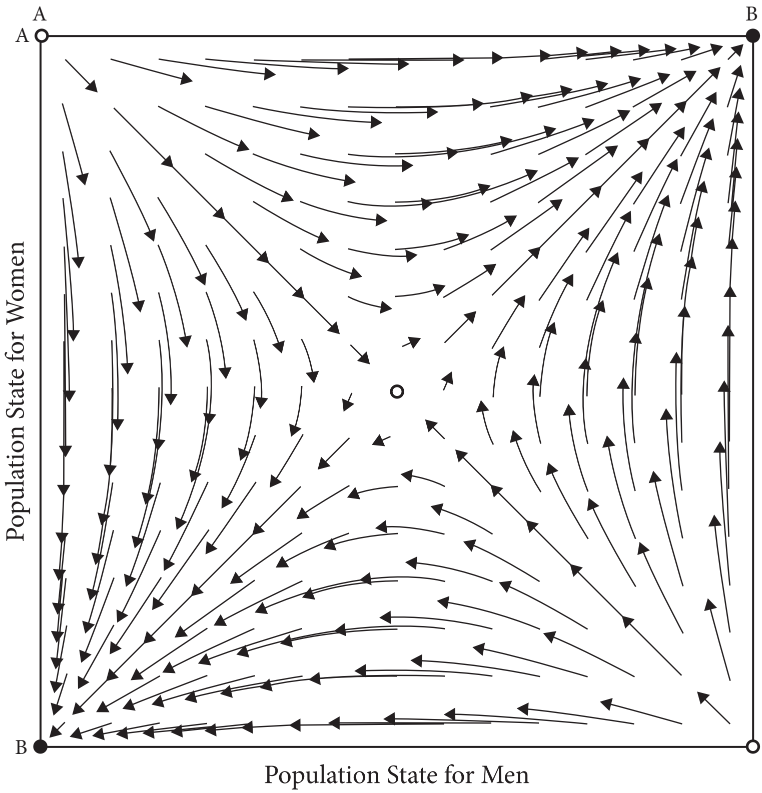

Figure 3.3 A phase diagram for the dancing game evolved with the two-population replicator dynamics

one strategy, and the other the other. Figure 3.3 shows a phase diagram for the dancing game under the *two-population replicator dynamics* (see the Appendix).[14] Obviously this figure is a bit different from the one pictured in Figure 3.2. This space represents *all* the possible population states for each of the two types in such a model. The x-axis represents proportions from 0 to 1 of actors of the first type (men, say) playing A and B. The y-axis represents the same thing for actors of the second type (women)

[14] This figure, and others like it throughout the book, were made using the program Dynamo developed by Sandholm et al. (2012). The dynamics are the continuous time version of the two-population replicator dynamics.

playing A and B. Each point in the diagram, then, represents one joint population state (for example, the first type is .25 A and .75 B, the second type is .55 A and .45 B). For each of these points, the diagram also shows the direction of population change. These arrows play the same role as the ones in the single-population phase diagram, but now instead of just pointing right or left, they can point in any direction in the plane. Look, for example, at the lower-right hand corner. This corner represents the state where both populations all play B. There is an arrow pointing up and to the left, which indicates that from this point, the replicator dynamics carry the populations away from this point, and toward a state where more actors in both groups play A. The lengths of the arrows tell us how quickly the population is evolving at any particular spot.

In this figure one can see that the arrows successively point to the lower left-and upper right-hand corners. In fact, from any starting point, the population will go to one of these two rest points. These represent the states just described. (Men engage in woodworking, and women practical pottery. Men mow the lawn, and women clean the dishes.) Notice here that we expect these outcomes to emerge, even though the two types are completely symmetric. They face the same strategic situation. They have the same preferences and abilities. But nonetheless, we expect them to end up at asymmetric outcomes.

We can identify what are called *basins of attraction* for each of these rest points. The basin of attraction, for a rest point, refers to the collection of states that eventually evolve to that rest point.[15] For the upper right rest point, here, the basin of attraction is the upper right triangle of the diagram, and for the lower left rest point, it is, likewise, the lower left triangle. Note that these basins of attraction are the same size. What this means is that for this model either type is equally likely to end up playing either strategy, assuming we don't know where the population starts.

There are also three unstable rest points, two at the other corners of the phase diagram and one right in the middle. Again, these are not evolutionarily significant because populations that are slightly perturbed from them will tend to move toward the stable rest points. Notice, though, that the rest point in the middle corresponds to a population where half

[15] For the homogenous populations, I did not bother identifying the basins of attraction. Since there is one globally asymptotically stable rest point, the basin of attraction for it is the entire state space (minus the two unstable rest points).

of all individuals take strategy A and half B. In other words, it is exactly the population that was the *only* stable rest point in the homogenous group. This is just to drive home how drastically the introduction of types changes the evolution here—the rest point which was guaranteed to evolve is now guaranteed not to.

As in homogenous groups, changing payoffs of the game will change the location of the interior rest point in these evolving populations, but will not change the general shape of the phase diagram. Figure 3.4 shows the phase diagram for the leader–follower game introduced in Chapter 1. In this figure, the interior rest point has shifted, but the stable rest points and their basins of attraction are the same. It is worth noting that as long as we assume complete symmetry between the two types—same strategies,

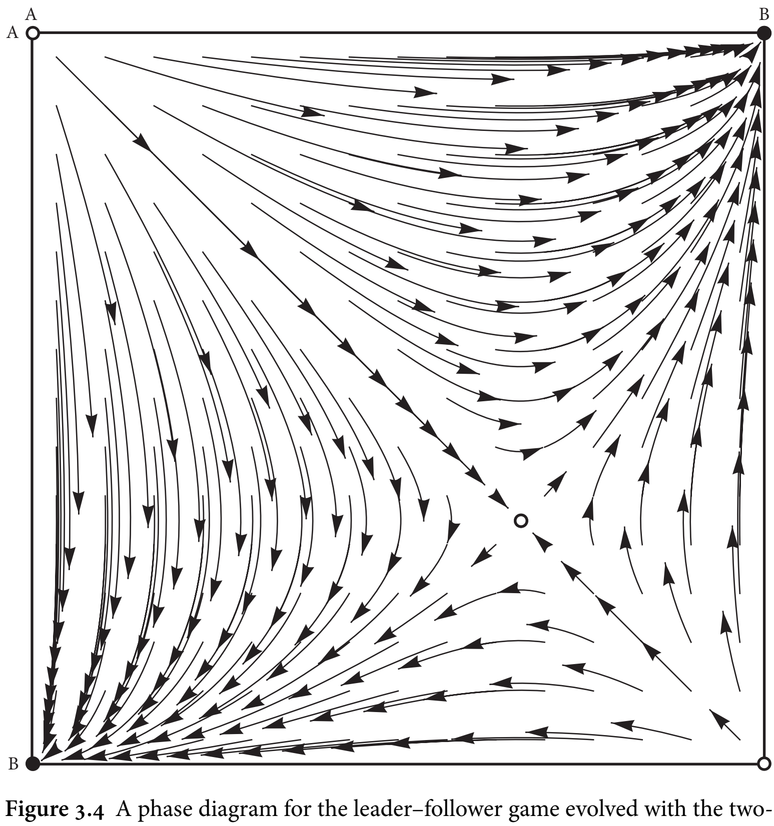

Figure 3.4 A phase diagram for the leader–follower game evolved with the two-population replicator dynamics

same payoffs, same dynamics—we will also find that the types are equally likely to end up at either outcome. In the case of gendered division of labor, this symmetry means strong conventionality as to gender roles. Later in the book I will look at models where these assumptions are dropped.

3.3.3 *Two-type mixing groups*

Last imagine a population where individuals have to decide whether to open doors for each other, or whether to walk through. The population is divided into women and men. As they interact women imitate other women who seem to be doing well, and men other men.

This is a slightly different evolutionary situation from either of the previous ones described. In either homogenous or perfectly divided models, actors meet only one other type. In homogenous groups, actors only meet their own type. In perfectly divided ones, they meet only the other type. This means that actors have only a single strategy that specifies what they do when they meet another agent. In two-type mixing groups, actors meet both their own type *and* the other type. This means that their strategies have to specify what happens for each interaction, meaning each of the types has four possible strategies for a two-strategy game. For simplicity's sake, I'll list a strategy as a pair <X,Y>, where X is what you do against your own type and Y against the other. The four strategies for a two-strategy game are then <A,A>, <A,B>, <B,A>, and <B,B>, representing something like "open the door for other women, but let men open the door for me".

In these groups, when you evolve the four strategies, outcomes are (more or less) a combination of the outcomes described in the last two sections.[16] Actors always evolve to play the outcome that is a rest point for the undivided population when interacting with their own type, and learn to play one of the two perfectly divided outcomes with the other type. (These, again, are the population level equilibria described in the last chapter.) For example, women and men will all sometimes hold the door for members of their own gender, but men will always hold the door for women, or vice versa. Again, in these models these two outcomes

[16] See the Appendix for the dynamics governing this process.

against the other group will happen equally often as long as the game is symmetric. Unfortunately, it is not possible to show a phase diagram for this evolving population because it has too many strategies.

The main take away should be that in the class of games we are largely concerned with here, all three of the sorts of groups we are looking at evolve in dependable ways. Homogenous groups evolve to a mixture of both strategies (with different proportions based on payoffs), perfectly divided groups evolve so that each type takes one strategy, and in two-type mixing populations each type evolves to play the polymorphic equilibrium against their own type, and a coordination equilibrium with the other type. So what we see is a case where social categories change the almost inevitable evolutionary outcome to a completely different (and more successful) one. On a cultural evolutionary picture, we expect groups of people to, in fact, end up taking advantage of social categories to improve coordination.

A more general note is that the simple addition of categories to these models dramatically alters cultural evolution. This observation tells us first that as a methodological point, cultural evolutionary predictions of behavior in human groups should take categories into account. It also provides deep explanation of the fact that social categories are so often associated with different social roles, and often attendant inequity (Tilly, 1998; Ridgeway, 2011).

3.4 Bounded Rationality, Evolution, and Robustness

The main claim on the table is that we should expect real human groups to culturally evolve such that social categories end up being important for coordination. In particular, as we saw in the last chapter, this will mean that some types should end up relatively disadvantaged by the processes that create more efficient groups. Women will end up with less, for example, because they, as a group, do not control food production. But to this point, these claims are grounded in one set of highly simplified, idealized models. How do we grow a bit more confident in their ability to tell us about real groups?

Robustness analysis in modeling involves changing various features of a model to ensure that those features are not responsible for a result attributed to something else (Weisberg, 2006).[17] In other words, robustness can be checked by reinstantiating causal variables in a different structure and seeing whether they lead to the same outcome. In our case, the variables responsible for the outcome (groups developing sometimes inequitable conventions to solve coordination problems) are that 1) groups encounter complementary coordination problems, 2) groups learn to do what benefits them, and 3) groups have two types. Let's explore some models from the literature with these features to test the robustness of the emergence of inequity.

Young (1993b) presents some of the first results showing that class inequities can emerge endogenously in populations with types where actors play bargaining games, Axtell et al. (2000) reproduce these results in a computational framework.[18] Although we will not introduce bargaining games until Part II, all we need to know here is that they have a partially complementary coordination structure. Axtell et al. (2000) model finite populations (sized 10–100) where agents are either in a homogenous group or of two types (landowners and sharecroppers, say), and where they repeatedly interact with both their own and the opposite type. (This corresponds to the two-type mixing model.) These authors use a dynamics that is very different from those employed here, where agents update their behavior based on a string of past interactions. Each round, several agents are paired to play the game. Each agent remembers some short set of interactions they have had, say the last five. Based on this list, they play whatever strategy will yield the best payoff assuming that their recent interactions reflect the larger population. In other words, agents *best respond* to their memory. When the population has two types, actors remember and best respond to their interactions with the two types separately. With some small probability agents play another strategy instead of their best response. Note that this sort of dynamic involves actors that display *bounded rationality*. The actors do not have infinite

[17] Some philosophers of science think that robustness analysis is not effective for scientific confirmation (see Cartwright (1991); Orzack and Sober (1993); Odenbaugh and Alexandrova (2011)), while others argue that robust results are better confirmed (see Levins (1966); Weisberg (2006); Wimsatt (2012); Heesen et al. (2017)).

[18] As they also point out, short- and medium-run behavior in these models is very important in understanding real-world behavior. Mohseni (2017) demonstrates why short and medium-term behavior is important in this sort of model. Their results have been replicated computationally by López-Paredes et al. (2004); Phan et al. (2005); Poza et al. (2011).

memories, neither do they have higher-order expectations about what their opponents will do given their opponents' beliefs. On the other hand, they do not blindly copy cultural variants, but instead make some calculation as to what will benefit them. In groups with types, Axtell et al. (2000) find that, as in our replicator dynamics models, inequity can emerge that would not be possible for a homogenous group. Again, this is because the types break symmetry.

These models demonstrate a sort of "throw everything at it" robustness of the results described in this chapter. The models in Axtell et al. (2000) have finite groups, randomness, and a boundedly rational choice rule for strategy selection, as opposed to replicator dynamic models with an infinite population, non-random dynamics, and a selection-based strategy update. Furthermore, under the most common cultural interpretation of the replicator dynamics (imitation of successful group-mates), actors copy the successes of their own type, while in the Axtell et al. models they simply respond to what they encounter in the environment. (This is the sense in which own-gender imitation is not necessary to the main results of this chapter.) We might even say that the Axtell et al. models do not involve type-conditioning. Agents do have separate memories for the separate types, so they recognize type distinctions, but they use one general rule (best response) for deciding behavior in both cases.[19] Nonetheless, largely similar outcomes are seen. This can increase our confidence that these are the sorts of outcomes we expect in evolving populations with types playing complementary coordination games.

There are other results that support the robustness of these evolutionary results. Henrich and Boyd (2008), in an exploration of the emergence of social stratification and division of labor, consider models that are very similar to those discussed here, but that add several features.[20] In particular, they consider conditions where learning does not occur

[19] The Axtell et al. models can be thought of as representing the emergence of what Millikan (2005) calls "counterpart reproduction," where actors learn complementary roles using their opponents as templates, rather than using social imitation of others. She uses the example of reproducing a handshake to illustrate this sort of reproduction. I think this is an ineffective example of what she is trying to pick out, because actors might learn to shake hands by exactly replicating their partner's actions. In complementary coordination problems where actors learn a complementary role from their opponent, direct social imitation cannot lead to the perpetuation of the behavior.

[20] They look at models where two populations evolve under the replicator dynamics and play a complementary coordination game where one role is preferable to both types.

entirely from members of one's own type. Instead, actors learn from their own group with some probability, and from the other type the rest of the time. Importantly, they find that under a wide range of conditions the same outcome still emerges where the population uses type membership to divide labor. When mixing is very high, so that the groups are more like a homogenous group, this does not occur, but for low to moderate levels of mixing (depending on a few other factors), the benefits of typing and type-conditioning drive the population toward the relatively efficient, inequitable equilibrium.[21, 22]

Additionally, Bowles (2004); Bowles and Naidu (2006); Hwang et al. (2014) look at game theoretic models that, like those presented by Young and co-authors, involve finite populations and best response dynamics, though they assume that actors observe the current state of the other population and best respond accordingly, rather than responding to some small set of memories. In addition, their actors tend to err toward conventions that will benefit them. In models where actors play a coordination game with an equitable and an inequitable equilibrium, these authors find that the stochastically stable equilibrium between groups will be one of these two Nash equilibria depending on other parameters of the model. (This equilibrium concept looks at where an evolving population spends its time as the likelihood of mutation goes to zero. See Foster and Young (1990).) They find similar results when actors are networked. In other words, again, their groups use types to improve coordination, sometimes to the detriment of one type. Hoffmann (2006) likewise presents agent-based models where actors play hawk–dove and have a number of social

[21] As I have argued for the case of gender, in many cases in-group learning is a good assumption, but surely this learning is not a solid rule. These results from Henrich and Boyd (2008) indicate that even in cases where actors sometimes learn from other types, we can see the sorts of outcomes described in this chapter.

[22] Nakahashi and Feldman (2014) provide another model of division of labor. While their model does not involve game theory, the set-up is one where actors can take advantage of two different sorts of resources via two different skill sets, and where they have some common interest as a result of resource sharing. They must decide how much effort to put into learning these skills, and then into exploiting the resources. Although their model does not assume a complementary coordination problem exactly (because actors can gather resources for themselves only and because they share as a group) there are suggestive similarities to the models here. They find that gendered division of labor emerges only when the genders have different innate skills, but this is because their models involve groups that share resources together, rather than situations where pairs of individuals have to figure out how to coordinate behavior.

identity markers which they use to generalize their expectations about opponent play. His actors best respond to their learned expectations. He finds that they use their identities to facilitate complementary coordination, as in the models here.

So in models with infinite and finite populations, in models on networks, in models with cultural imitation, and with boundedly rational response, in models where actors learn from their own and from other types, we see groups evolving so that social categories facilitate complementary coordination. This reflects a deep truth about efficiency outlined in the last chapter—social categories do something that homogenous populations do not. We shouldn't be entirely surprised that many sorts of cultural evolutionary processes are responsive to this efficiency.

3.5 Social Categories and Correlative Coordination

As we have seen, the addition of categories to a social evolutionary process drastically changes outcomes. In the case of complementary coordination, it does so in a way that allows for inequity, where previously only equitable outcomes were expected. But this is not the only way that social categories shape cultural evolution. In this section, we'll take a short detour to see how conventions and norms for correlative coordination turn out differently in groups with types, sometimes to the detriment of one type.[23] (Since this strays beyond the main narrative of the book, and introduces some games that will not be relevant elsewhere, readers can safely skip this section.)

3.5.1 Correlative coordination games

Suppose we have a homogenous group learning which side of the road to drive on. Unsurprisingly, what happens is that everyone evolves to do the same thing—drive on the right or the left. If we have a perfect division—men meeting only women and women only men, for example—again cultural evolution will drive everyone to solve the coordination problem by all doing the same thing.

[23] Thanks to Justin Bruner for inspiration here. He explores the evolution of the stag hunt in two-type mixing models in a previous version of Bruner (2017).

In these cases, we should not expect inequity unless there is a situation with something like a Bach–Stravinsky game (remember, this is where everyone wants to listen to music together, but some people prefer Bach and others Stravinsky) and where for some reason members of one social category generally have different preferences or needs from the other. Suppose that everyone faces a coordination problem in deciding how much parental leave to grant. The optimal level for men and women might differ owing to the physical demands of childbirth, and so a coordination outcome that is symmetric and equal might nonetheless yield higher payoffs for one type.

What happens in a two-type mixing population evolving to play correlative coordination games? Here things get interesting. Because there are two types, it is no longer the case that everyone will necessarily evolve to do only one thing. Instead, actors develop three separate conventions. The first governs what happens in the in-group of the first type, the second governs what happens in the in-group of the second type, and the third governs what happens between groups. For an example, in a Bach–Stravinsky game, it could be that women listen to Bach together, men listen to Stravinsky, and when men and women meet they listen to Bach.

Of particular interest here are games where one outcome is generally preferable to the other. Consider Figure 3.5, which shows the game presented in Figure 1.2. In this game, the B vs. B outcome is preferable to both players. But the other outcome will still be an end result of some evolutionary processes. While this may sound unintuitive, the worse equilibrium is still an equilibrium. (In fact, societies end up at suboptimal social equilibria all the time. Bicchieri (2005) gives a number of examples.) In two-type mixing populations, outcomes can occur where only one group learns the more beneficial equilibrium. Or where the poorer equilibrium arises only between groups. As a result, one group will have higher payoffs on average than the other. Although these outcomes

		Player 2	
		A	B
Player 1	A	1, 1	0, 0
	B	0, 0	2, 2

Figure 3.5 Correlative coordination game with a preferable outcome for both players

are not as compelling as representations of discrimination as those we have been discussing, they still represent a case where the existence of social categories can lead to a sort of inequity, or at least different ultimate levels of payoff for different types.

3.5.1.1 THE STAG HUNT

The motivating story behind the stag hunt comes from Rousseau (1984, orig. 1754). Two hunters each have the choice to pursue a stag, or to hunt small game—hare. If both hunt stag, they are able to collectively take down the stag, and benefit accordingly. If either hunts hare, she catches the hare and gets a small payoff. If one hunts stag, and the other hunts hare, the stag-hunter is unable to succeed on her own and gets nothing. This game is represented in Figure 3.6. The payoff here is 2 for hare, 3 for successful stag hunting, and 0 for unsuccessful stag hunting.[24]

The stag hunt is usually thought of as a model of cooperation under risk, but it technically also fits into our definition of a correlative coordination game.[25] For this model, the two Nash equilibria are the two outcomes where actors correlate their behavior—Stag vs. Stag and Hare vs. Hare. Actors in this game have preferences about what happens when they do not coordinate—Hare always yields a dependable payoff—but coordination is expected. Unsurprisingly, when a homogenous group evolves to play this game, everyone plays Hare or else they all play Stag.

When two types evolve to play this game, sometimes only one type will learn to play stag, or else everyone will play stag only with their out-group. As Justin Bruner has pointed out, when everyone cooperates with

		Player 2	
		Stag	Hare
Player 1	Stag	3, 3	0, 2
	Hare	2, 0	2, 2

Figure 3.6 A stag hunt

[24] This game has been employed by philosophers as a model of cooperation in human groups (Skyrms, 2004; Alexander, 2007; Skyrms, 2014).

[25] We can see that if we take the analysis of what counts as a coordination game—the payoffs for the complementary coordination choices are higher than the alternative payoffs given the other actor's choice—it will apply here as well. 3 is greater than 2, and 2 is greater than 0.

their in-group, but not the out-group, the convention looks like a kind of discrimination against the out-group. These sorts of outcomes have combined basins of attraction that make up about 38 percent of the state space in these models—a significant portion of outcomes. In outcomes where types cooperate between each other, but one type learns not to cooperate with their in-group, the non-cooperative group will be worse off. These outcomes have basins of attraction that make up another 9% of the state space. Again, in this case, the mere presence of types creates unequal possibilities that wouldn't exist otherwise.

• • •

We have seen that groups with social categories evolve and develop conventions in very different ways from homogenous groups. In particular, evolution takes advantage of social categories to improve coordination, even when the resulting conventions are inequitable. In general, separate conventions can arise for those in different social categories, and sometimes this can mean inequality of a different sort.

The case we have been using to illustrate how social categories can act as a solution concept—gendered division of labor—is thus addressed. Gendered division of labor is what we should expect from cultural evolution given a set-up where labor takes specialized skills. Details about the skills and preferences of men and women, notice, have not played any role in the story so far. This is because it is the simple fact of symmetry breaking that makes social categories functional, not any features of the categories themselves. This also helps explain the conventional nature of gendered division of labor. Why do we see different jobs performed by men or by women cross-culturally? Because while it matters that labor be divided for coordination purposes, it does not particularly matter how. In the next chapter, we will complicate this story, and use evolutionary models to explain why, despite this conventionality, some jobs are primarily performed by men, some by women, and some by either gender cross-culturally.

Before continuing to the next chapter, where some of the ways this framework shines light on gender will be addressed at greater length, I'd like to drive home a point made in the Introduction. The models presented so far, I think, are how-potentially models, potentially representing the sorts of causal processes that, in the real world, occur in groups with social categories. They are also how-minimally models—showing that remarkably little is needed to generate inequity via cultural evolution. All we need is a group with two types, playing a complementary coordination game, and undergoing an adaptive process of cultural change. There is no bias in this model, no stereotype threat, not much psychology in general, but nonetheless we see group-level patterns that mimic inequitable conventions. We see something like stable norms where the men who control the plough continue to reap benefits over cultural evolutionary time.