2

Social Categories, Coordination, and Inequity

In cultures where agriculture is mostly plough-based, men tend to have more power, and greater economic stability. In cultures where agricultural practices depend on hoes or digging sticks, not so (Alesina et al., 2013). The reason? Because ploughing requires great upper body strength, and is generally incompatible with child care, men tend to control crop production in societies that use ploughs. And as a result, these societies tend to be patriarchal.

The end of the last chapter introduced a special challenge for coordination—how can groups of people deal with complementary coordination problems when any profile of strategies adopted by the group will lead to instances of mis-coordination? (Or, alternatively, how can groups redefine a complementary coordination problem as a correlative one?) In this chapter we'll start to see why social categories provide ways to solve such problems. Men can step forward and women can step back. Or old people can go through the door first, and young people second. In other words, a group with types will be able to achieve a level of efficiency unavailable to a group without them.

As we'll see though, there is a catch. The efficiency gained by adding social categories to a group often comes at the expense of egalitarianism. In the Introduction, I introduced two sorts of inequity described by Smith and Choi (2007). Groups that use social categories to solve coordination problems often display inequity of the first sort—joint action with unequal roles can lead to unequal benefits for those involved. In the case of gendered division of labor, when men take certain jobs and women others, then only one gender will receive the benefits associated with

whatever jobs they do. When there are clear advantages to certain jobs—such as control over a family's food supply—whichever gender does that job is the one that reaps the rewards.

We'll start the chapter by considering, in general, what it means for a group to have social categories, focusing on gender in particular. Then I'll describe how social categories can be represented in game theoretic models via types and type-conditioning. After that, we'll look at what sorts of coordination can be achieved by populations with and without types, noting, in particular, how efficiency can be improved by typing. At the end of the chapter, I'll explore the tension just described—that improved social coordination comes at a cost to egalitarianism.

2.1 Social Categories in Human Groups

In the models I'll introduce, two things have to be in place for groups to use social categories in coordination problems. First, members of the group have to jointly be able to identify these categories, or *types*, and to place other group members into them. Second, everyone must be able to condition their strategies on that membership. In other words, individuals must be able to treat people differently based on what social category they are part of, or *type-condition*. I'll define these terms from a modeling standpoint in the next section. For now, let's look at the evidence that these features are, in fact, present in real human groups.

The minimal group paradigm for experimentation in the social sciences nicely demonstrates how easily social categorization comes to human groups. A typical experiment of this sort first arbitrarily divides experimental subjects into groups, using meaningless criteria (a coin flip, for example). Subjects are then given an opportunity to perform a task, like dividing a resource between the other subjects. Experimental results show that subjects tend to favor those in their own group despite the arbitrariness of this membership. The division into groups alone is enough to cause type-conditioning of a special sort in these cases (Tajfel, 1970, 1978; Haslam, 2004).

In the real world, people often use external markers—either alterable or unalterable ones—to distinguish categories of people. Alterable markers are things such as clothing, hair style, piercings, tattoos, etc. Unalterable

markers involve things such as biological sex differences, skin color, and body or facial appearance, including age or visible physical disability.[1] In particular, these sorts of markers can induce type-conditioning, especially out-group discrimination, much in the way minimal groups can.[2]

Of course, not all observable markers are used to identify in- and out-groups, or to type-condition. For example, there are no deep social patterns of behavior that depend on ear size, even though ear size is observable. Ridgeway (2011) uses the term *primary categories* to identify the groupings that societies rely on most heavily for coordination purposes. These are the sort of categories that, as Bicchieri (2005) puts it, "people tend to perceive…as natural kinds having high inductive potential and stability" (89). Cross-culturally gender and age seem to always be employed as primary categories. In some cultures, race, caste, and/or class also serve as primary categories (Ridgeway, 2011). In a meta-analysis, Kinzler et al. (2010) identify gender, age, and race as the most important social categories cross-culturally. There are reasons why these categories (and not ones based on ear size) are the ones we use. Gender is important to differentiate behaviorally for reproduction. Furthermore, heterosexual household formation means that it is a division that will be salient and present in many day-to-day interactions between household members. In particular, it means that dyads with a woman and a man are often the locus of joint production, and so benefit from categories that allow them to coordinate labor (Ridgeway, 2011). Age is also relevant when it comes to reproductive behaviors, and other behaviors where developmental stage is important.[3]

---

[1] Obviously to some degree sex and even skin color can be altered, but there is a huge divide between markers like clothing which can easily be donned and removed, and markers that take extreme medical measures to change. (Markers like tattoos occupy a middle ground, easy to get, difficult to remove.)

[2] Some theorists believe this is a result of out-group bias that was evolutionarily functional for early human groups (Cosmides et al., 2003). In other words, differentiating between those people who are kith and kin, and those who might be a threat, was important, as was reacting differently to these categories of people in some settings. Current evidence suggests that humans see out-group members as more homogenous than in-group members (Quattrone and Jones, 1980), and tend to dislike and discriminate against out-group types (Brewer, 1999). Modelers have shown how in-group preference can emerge in populations with types (Axelrod and Hammond, 2003; Hartshorn et al., 2013).

[3] Cudd (2006) argues that in explaining oppression, it is essential to look at social categories of this sort. The argument in this book will go further—as we will see, the bare fact that social categories exist is often sufficient to generate significant inequity.

As mentioned above, for these sorts of categories to play a role in social coordination, they must be recognizable. When it comes to gender, biological sex differences provide grounds for this sort of recognition in humans, but, in addition, people amplify the signal through dress, ornamentation, and indicative behaviors. As West and Zimmerman (1987) point out, "Neither initial sex assignment . . . nor the actual existence of essential criteria for that assignment (possession of a clitoris and vagina or penis and testicles) has much—if anything—to do with the identification of sex category in everyday life" (132). As mentioned in Chapter 1, one central aspect of gendered behavior, besides division of labor, is that humans adopt signals and displays to communicate their gender identity (Goffman, 1976; Lorber, 1994).[4]

What do signals of sex category or gender look like? Chances are that you can easily think of examples, though, as Lorber (1994) points out, "Gender signs and signals are so ubiquitous that we usually fail to note them—unless they are missing or ambiguous" (14). In modern Western society there are different standards of dress and ornamentation for men and women—women rarely wear three-piece suits, and men rarely wear lipstick, for example. In other cultures, likewise, it is very typical to see differences in dress and ornamentation for men and women. In Kerala (southern India), men wear a lungi (or mundu), which is a long cloth wrapped around the waist. Women typically add a blouse to this. Traditional Korean dress, or hanbok, consists of a long skirt for women and pants for men.[5] There are also behavioral signals of gender identity where the behaviors are not themselves strategically important, but nonetheless telegraph type membership. "Man-spreading" is a phenomenon recently much documented on the Internet, where people post pictures of men on the subway who spread their legs to take up too much space. Notably, no one is worrying about "woman-spreading," because this is a body position that women essentially never take.

---

[4] There might seem to be some circularity here—to efficiently divide labor by gender groups need gender signals, but these signals emerge to facilitate the division of labor by gender. As I will point out later, it is usually the case that a bundle of social features must be in place for gendered division of labor to work. While I do not develop a full account of how these features could emerge in concert, a plausible story involves some bootstrapping as both gender signals and full gendered division of labor slowly develop.

[5] The suggestion here is not that the only role of gendered dress is to successfully convey sex category, but that this is one functional role it plays.

What does it look like when members of a society recognize categories and use them to condition behavior? This can be observed when we see members of a society acting differently towards those in different primary categories under the same circumstances. Such behaviors are not limited to forms of in-group favoritism. Instead, there can be many ways in which actors type-condition, including favoritism for other types. Ridgeway (2011) argues that innate out-group bias primes the recognition of different types in human groups. Once we categorize people into types, this can set the stage for other sorts of type-conditional behavior to emerge.

Here is a short list of examples of research performed in the US finding type-conditioning. Of course, as noted, type-conditioning need not involve discrimination. These studies tend to focus on discriminatory behavior, though, because much identifiable type-conditioning is discriminatory, and discrimination is of obvious research interest. Researchers have found that when sending otherwise identical resumes out for jobs, those with male names and "white" names are more likely to receive job offers, and more likely to be offered higher pay for jobs (Steinpreis et al., 1999; Bertrand and Mullainathan, 2003; Moss-Racusin et al., 2012). Similar findings have been garnered for job applications where the candidate reveals that he or she is LGBTQ (Tilcsik, 2011). Black house-hunters are not given access to the same housing opportunities as whites (Yinger, 1986; Hernandez, 2009). Women, but not men, often receive negative backlash for assertive behavior in bargaining scenarios (Bowles et al., 2007; Tinsley et al., 2009). Black and female patients are less likely to be recommended for heart catheterizations than white males (Schulman et al., 1999). In bargaining for cars, female and black customers receive higher starting offers, and meet more resistance in bargaining prices down (Ayres and Siegelman, 1995). Employers treat Hispanic and white job seekers differently (Cross, 1990). This list could be extended ad nauseam, and to include other categories such as age and disability, as well as alterable markers. The point is hopefully made.

It must be noted that there are many cases where type-conditioning is necessary for a group to function well and social categorization sometimes allows for this. It is arguably a *good* thing to make a distinction between elderly and young bus passengers and to treat them differently by giving your seat to one type but not the other. It is a good thing for doctors to recognize that male and female patients, or patients of different racial backgrounds, may have different needs and to treat them appropriately.

2.2 Social Categories in Models

How do we go about capturing social categorization in a model? Through-out much of the literature, these are referred to as models with "tags," where a tag is a marker observable by other actors. Axtell et al. (2000) describe tags as having "no *inherent* social or economic significance—they are merely distinguishing features, such as dark or light skin, or brown or blue eyes. Over time, however, they can acquire social signifi-cance due to path dependency effects" (2).[6] In fact, in their models Axtell et al. refer to tagged groups by colors (i.e., "blues" and "reds") to make clear that these markers do not carry the significance that real social categories do. In this book, I use the term *type* rather than "tag" because I want to capture the way that types tend to be stable, persistent, and general identities in human groups.

Type-conditioning occurs when actors use these tags to differentiate their behaviors toward different types. As discussed in the last section, type-conditioning behaviors in the real world are often associated with psychological phenomena like out-group bias. In the models we will discuss, though, type-conditioning will be a thinner phenomenon. It will simply involve the possibility that actors develop different strategies when interacting with in- and out-group members. (Though in some places, I will introduce out-group biases to see how they influence outcomes.) This choice is part of the general strategy discussed in the Introduction of exploring the minimal conditions necessary to generate inequity. As it turns out, type-conditioning in this thinner sense will be enough, in many cases, to generate persistently inequitable social patterns.

Note that even with this stripped-down understanding of types and type-conditioning, the models have the capacity to represent situations where actors suffer distributional injustice, or discrimination. For exam-ple, we will see many models where groups of actors have evolved con-ventions such that members of one type receive fewer resources than members of another. Even without positing internal biases to explain how such a convention is maintained, we can say that the convention itself is inequitable.

---

[6] For examples of models with tags, see Holland (1995); Epstein and Axtell (1996); Axelrod (1997); Axtell et al. (2000); Bruner (2015). Many authors have focused on the role tags can play in facilitating cooperation (Hales, 2000; Alkemade et al., 2005).

2.2.1   Types and signals

There is an interesting connection between typing models and another widely used concept in game theory and evolutionary game theory—preplay signaling. In models with preplay signals, before actors play a game, they have the option to send a signal to the other player, and condition play based on any signal received. One way to conceive of types is as signals between agents that transfer information and, as a result, allow for conditioning behavior. First, player 1 signals that she is a woman, and player 2 that he is a man, then they play the dancing game.[7] In signaling models, though, an important part of what makes signals signals is that they themselves evolve. In a cultural evolutionary model, actors have the option to adopt a new signal if the current one is not working for them. Obviously when it comes to social categories, this is not so easy. This is why I use "types" here instead of directly importing the signaling paradigm. I am mostly unconcerned with how types themselves evolve for strategic purposes, and much more concerned with cultural evolution of other behaviors in populations with types that are essentially fixed. (Though, Bowles and Naidu (2006); Hwang et al. (2014) look at relevant models of class inequality where actors can change classes.) Furthermore, we will be interested, in Chapter 4 and in the second half of the book, in models where types encounter asymmetric interactive roles, and in these cases "types" cannot be treated as pre-play signals because such signals do not capture the asymmetric features of the multiple subgroups in a population.[8]

Note that the models discussed in this book do not represent cases where "types" consist in strategies in a game. For example, the types in a population cannot be those who drive on the left side of the road and those who drive the right side of the road.[9] Conditioning on an observable

---

[7] Relatedly, Skyrms (2004, 78) explores signaling models that are much like type models, where he sees the evolution of coordination in the dancing game.

[8] Of course, types can be amplified or diminished through dress, style, and behavior. This means that there are ways in which individuals can strategically change their type, at least to some degree, which would better correspond to a preplay signaling model. As mentioned, though, the goal here is to focus more on what happens to other strategies when types are held fixed. Thanks to Kevin Zollman for this point.

[9] In the literature on the evolution of cooperation, it is common to look at models where actors are able to recognize whether other agents will or will not engage in certain behaviors (whether they take certain strategies or not). For a few examples, see Frank (1988); Frank et al. (1993); O'Connor (2016).

strategy by an opponent moves away from tags as having "no inherent social or economic significance."

2.3 Social Categories as Solutions to Coordination Problems

As I have said, complementary and correlative coordination problems are different. Correlative problems do not pose a particular challenge to groups where members interact pairwise with many others. Complementary problems, on the other hand, do pose such a challenge. Suppose once more that we have a homogenous group of people playing the dancing game—who will step forward and who will step back? We should expect a lot of people stepping on one another's toes. Now we'll use game theoretic models to say, in greater detail, how types can solve this problem.

In order to do so, I must first introduce the concept of *expected payoff*. An expected payoff for an actor is a number representing what they should expect to receive as payoff in a situation where outcomes are probabilistic. For example, suppose we have a population of actors where 60% of them play A in the driving game and 40% play B. For an actor playing A, we can say that her expected payoff when she goes out driving is the chance that she meets another A player times the payoff she receives should she do so plus the chance that she meets a B player times that payoff. If we assume that she meets other players randomly, this comes out to $.6 * 1 + .4 * 0$, or $.6$.

I also must introduce the concept of a *population equilibrium* (as opposed to a Nash equilibrium in a game between two individuals). This is a set of strategies for an entire *population* such that no individual wants to change what they are doing. In a correlative coordination game, these consist in every person doing the same thing. (If everyone is driving on the right side of the road, no one wants to change to the left.) These equilibria, as we will see in the next chapter, are very important from a cultural evolutionary point of view, because they tend to be the endpoints of evolutionary processes. For now, we need this concept to compare how well people do in terms of expected payoff in groups with and without types that are at population level equilibria.

Let's start with our homogenous group playing the dancing game. As discussed, the group does very poorly if everyone adopts the same

strategy. The best this population can do is for 50% of people to always step forward and 50% to always step back, and, indeed, this is an equilibrium. (If anyone switches what they are doing, their expected payoff will be slightly worse.) At this mix, when two random people meet there is a .25 chance they collide, a .25 chance they both step back, and a .5 chance they choose complementary actions. This means that, on average, the population will achieve coordination (a payoff of 1) half the time and fail to coordinate (a payoff of 0) the other half. This yields an average expected payoff of .5. We can immediately see that this isn't particularly good. In the very best-case scenario, everyone coordinates only half the time.

Now suppose that we add types to the model—half of the dancers are now men, and half are women, and everybody is able to recognize the two categories and condition their behavior on this. The best-case scenario is that men dance with women, and women with men, and everybody adopts one of two strategies. Either all the men step forward and all the women back, or all the women step forward and all the men back. (And, again, these are the equilibria of the population.) At these sets of strategies, coordination will be perfect in every case, so that the average expected payoff will be 1. The intuition here is that by recognizing types actors gain extra information that allows them to coordinate by picking one of the two equilibria of the game. When a man and woman meet, they can use the gender difference as an asymmetry to create expectations for who takes which role. When members of the same gender meet, this is not possible.

So, the addition of types to a model, in this case, can double successful coordination, and, as a result, double the payoff members of the group expect to receive. But, something extra was actually sneaked into the scenario just described, which is that men only met women, and women only met men to play the coordination game. We can also imagine a scenario where a group has two categories, but every pair of individuals sometimes engage in coordination problems. This sort of case occurs, for example, when deciding who will walk through a public door first. There are no restrictions on which members of a group will engage in this problem. Men and women and people of all races and classes approach doors at the same time and must decide who goes first. I will call this a *two-type mixing* model, in contrast to a model where everyone only meets their out-group, which I will call *perfectly divided*. For those familiar with evolutionary game theory, the perfect division corresponds to traditional

two-population models.[10] Note that in both sorts of model the group is partitioned into two types, and everyone knows the type of everyone else. The difference lies in who meets whom.

Even in two-type mixing groups, where everybody can meet, types can still help with coordination. Suppose our group encounters the coordination problem where pairings of people have to decide who will hold the door and who will walk through. The best possible scenario for coordination is that when men and women play this game, they use their types to coordinate—women always hold the door for men, or vice versa. Whenever someone meets their own type half of people go to open the door, and the other half walk forward. In such a scenario, half the time coordination will be perfect because people have types to coordinate with (meaning a payoff of 1). In the other half of interactions, there will be a 50% chance of coordination (meaning an expected payoff of .5). This makes the expected payoff of the whole group .75, which is better than for a homogenous group, even though there is still some mis-coordination.

To get a more general sense of this phenomenon, consider Figure 2.1. (Readers who are less interested in technicalia may wish to skip to the next section.) This is a general, symmetric two-player game. A game is a complementary coordination game, remember, whenever $b > a$ and $c > d$. Let us further restrict our attention to games where $a + a$ and $d + d$ are $\leq b + c$. This means that actors at coordination equilibria cannot jointly do better by switching to a non-equilibrium outcome. Note that all the games we have looked at meet this requirement.[11]

Consider the possible payoffs for a homogenous group playing this game. Let p be the proportion of the population playing 1 and $(1 - p)$ the

---

[10] In actuality, one can imagine a spectrum of situations where types interact with higher or lower levels of correlation. On one extreme are situations where actors only ever interact with their own types. These models are identical to models with no types at all, since they boil down to two individual, homogenous groups. On the other extreme are situations where actors meet only the other type (perfectly divided populations). Directly in the middle are the two-type mixing populations where interaction is not based on type at all. Henrich and Boyd (2008) consider two-population models that vary along this continuum. They find that when actors can learn from the other population, the level of homophily can impact how likely inequitable conventions are to arise. Because I consider models where actors learn from their own type only, we see similar between-group outcomes for all levels of homophily.

[11] We can make a version of hawk–dove where this would not be the case, though. Suppose that the Dove vs. Dove payoffs are 2.5 and 2.5, while the payoffs when hawks and doves meet are 3 and 1. The joint payoff for doves is 5, and the joint payoff at equilibrium is only 4.

		Player 2	
		1	2
Player 1	1	a, a	c, b
	2	b, c	d, d

Figure 2.1 A general payoff table for a symmetric two-person, two-strategy game

proportion playing 2. Then the average population payoff will be $p^2 * a + p(1 - p) * c + p(1 - p) * b + (1 - p)(1 - p) * d$. This equation weights the probabilities that each of the two types meet by the payoffs they receive in meeting. (For example, because p of the population play 1, they meet with probability $p * p$ and receive a payoff of a when they do so.) The best this population can do is to divide so that $p = .5$, or each strategy is played by half the population. This maximizes the weight on the better coordination outcomes compared to the non-equilibrium outcomes. The best average payoff is $.25a + .25c + .25b + .25d$.[12]

Now suppose we have a perfectly divided population where one side plays strategy 1 and the other strategy 2. Half of this population will always receive the payoff of c and the other half will always receive b. The average payoff will be $.5c + .5b$. Notice that this second equation will always be greater than the first. Why? Because $b > a$, $.5b > .25a + .25b$, and because $c > d$, $.5c > .25c + .25d$. A two-type mixing population (where everyone encounters everyone, but they all recognize both types) will always have an intermediate payoff between the perfectly divided and single populations, so their payoff will be greater than that of the single population as well.

2.3.1 Types all the way down

We have just seen how a group with two social categories can get better payoffs than a homogenous group. In many real-world populations, though, there are multiple primary social categories that actors use to condition behavior. What happens in populations with multiple kinds of types?

Consider a group with both men and women and young and old people. Furthermore, assume that these types intersect one another so

---

[12] Note that is will not always be a population equilibrium, but it will maximize payoff, so that if the equilibrium is $p = x$ where $x \neq .5$, the payoff will not be greater than at this point.

that there are young men, young women, old women, and old men. Forget about perfectly divided populations for now—in perfectly divided populations the group coordination problem is completely solved with two types. In two-type mixing groups, where everyone meets everyone else, however, there are still failures to coordinate when people meet their own type, which happens 1/2 the time. In *four*-type mixing groups, it will be less likely that this ever happens, because each person is a member of a relatively small subgroup and so meets their own type only 1/4 of the time. Whenever they meet another type, they can use the asymmetry between them to coordinate. In the dancing game, this means that their average expected payoff is $.25 * .5 + .75 * 1 = .875$. This is higher than the .75 that actors in the two-type group get.

This same argument can be given for more and more type divisions. (These need not be created as intersections of base types. What matters is simply to have shared divisions that all agents are able to recognize and use as symmetry breakers.) People interact inefficiently in these games only when they meet their own type. If the size of each type is smaller, this is less likely to happen, and so there is always a benefit of adding more type divisions.[13] Of course, at some point the benefits gained from further type divisions will be outweighed by the impracticality of identifying and remembering how one interacts with all these different types. It has been observed many times that humans are only boundedly rational, and that cognitive limitations will prevent some sorts of optimal strategies, from a game or decision theoretic perspective, from emerging.[14] Suggestively, Brewer (1988) argues that owing to cognitive limitations human groups typically employ only three or four primary social categories. Hoffmann (2006), in a model where actors use types to coordinate play of hawk–dove, finds that groups improve their payoff only by adding up to three type divisions. After that, as he says, "... the trait diversity of the population rises more rapidly than the sophistication of agent discrimination" (244).[15]

---

[13] This continued benefit of types is very similar to the continued benefit of extra pre-play signals that Skyrms (2004, 79–80) describes for a dancing game.

[14] For a compelling defense of bounded rationality in modeling human cultural evolution see Alexander (2007, 1–8). For empirical discussion of bounded rationality see Gigerenzer and Selten (2002).

[15] This result occurs in part because Hoffmann's agents apply their experiences to all the social identities of their opponents, rather than thinking of combinations of types as separate

2.3.2 *Gradient markers*

There is another way to use tags to solve complementary coordination problems in a group. Suppose we have a population with tags where the tags do not form categories or types, but instead have gradient values. In the model, this might look like a population of agents with tags that take values between 0 and 1—.012, .566, .78, etc. In the real world, this might represent a group where actors display different levels of femininity and masculinity, or engage in displays of social class to different levels, or have finely different skin color, or are of different ages. In this case, the solution to the coordination problem can be a unified rule such as "if my tag is of higher value choose step forward, if of lower value step back." This sort of tag can allow for perfect coordination in every interaction, assuming that actors' tags are finely varied enough that they never meet others with the same value.

2.4 Coordination and Discrimination

To this point, it might sound as if social categories are an unequivocally good thing. When we add them to our groups, we improve coordination, and everybody benefits. One of the messages of Part I of this book is that, indeed, social categories have the capability to do something for us. They can play a functional social role. But, as we will now discuss, this functional role opens the door to the possibility of inequity.

Before beginning this discussion, let's introduce some basic terminology from work on distributive justice—or the study of the ideal distribution of goods in a society. There are different norms one might want a social distribution of resources to meet. *Equality* refers to a norm where individuals ought to receive equal amounts of resources. *Equity,* on the other hand, is the principle that individuals ought to receive equal amounts of resource given something like equal contributions. Skyrms (2004) identifies the equity norm thus: "If the position of the recipients is symmetric, then the distribution should be symmetric. That is to say, it does not vary when we switch the recipients" (18). In the following, we will discuss unequal outcomes, which also can be interpreted

---

types of their own, but the general idea is that the benefit of breaking into further types grows smaller and smaller.

as inequitable in many cases—although actors in division of labor cases end up doing asymmetric things, there is no reason to think their contributions are deserving of less. When members of one gender control agricultural production, they benefit not because they are working harder, but because the type of work they are doing garners better direct payoffs.[16]

In the dancing game, the homogenous group gets an expected payoff of .5. When types are added, everyone gets a bump up to 1. In other words, types benefit everybody equally. But this is not so for all games. Recall the leader-follower game from Figure 1.4. In this game, remember, everyone would prefer to play B (be the CEO, not the clerk) where they get a payoff of 2, rather than A where they get a payoff of 1, even though everyone prefers coordination over mis-coordination.

At equilibrium, a homogenous group playing this game will be 1/3 followers and 2/3 leaders. (While a 50–50 profile would make the average payoff better, some followers would do better to switch and be leaders, meaning that it isn't an equilibrium.) When the group takes these strategies, everyone gets an expected payoff of 2/3.[17] Once we add types, however, at equilibrium we see different payoffs for the two types. If men play B and women A, then men get a payoff of 2 and women a payoff of 1. (In the two-type mixing group, where all meet one another, women get 5/6 and men get 4/3.) The reason this payoff difference is possible is that no one can switch types—if they could, they would until payoffs were the same and equilibrium was reached. Social categories provide the extra information needed to facilitate coordination, but they also block people from taking certain social roles, and, in doing so, create the possibility of unequal outcomes. In fact, one of the things that make social categories so effective at solving coordination problems is that people can't switch categories—they are stuck in their complementary roles.

Outcomes where members of one type play B and the other A are unequal, potentially inequitable, and furthermore can represent discrim-

---

[16] It is beyond the purview of this book to discuss implications of the modeling work here for theories of distributive justice. Okin (1989) argues compellingly that because of the sort of type-based inequities modeled here, any such theory must take gender into account.

[17] Followers meet leaders 2/3 of the time, and when they do they get a payoff of 1. Leaders meet followers 1/3 of the time and when they do they get a payoff of 2. $2/3 * 1 = 1/3 * 2 = 2/3$.

inatory conventions.[18] People treat those in different social categories differently, and as a result one sort gets poorer outcomes, on average, than the other. This could represent a situation where because women work at home they have less freedom and control over their lives, or where members of a lower caste do a job that yields fewer material benefits.

So now the story is more complicated—types allow for coordination, but they also allow for inequality. There is another kink though. For many complementary coordination problems, even though type–conditioning creates inequalities, it is still better for *both* types. To see what I mean by this, consider the last case just described. Men do better than women (2 versus 1). But they *both* do better than the homogenous group where everyone gets 2/3. From a payoff point of view, the discriminatory outcome with coordination is better for everyone than the egalitarian outcome without it. The type that ends up discriminated against should still choose a state of unequal coordination over one of equal disarray.

This last observation isn't necessarily true for games where actors care about what happens when they do not manage to coordinate. In hawk–dove, a homogenous group can get an average payoff of 1.5.[19] If a group is perfectly divided, hawks expect to receive a payoff of 3, because they always meet doves, and doves expect a payoff of 1. In other words, doves receive a *lower* payoff than they would expect in the homogenous group.

One might well ask—how common are these different sorts of scenarios in human populations? Which of these games represent real-world interactions? Understanding this may help us understand whether typing actually hurts people as opposed to creating an unequal situation that nonetheless benefits all involved. The answer to this question must necessarily be a very complicated one, and different for different groups. As discussed in Chapter 1, there are many, many complementary coordination problems that groups of people solve. These will be best modeled by various of the games that I have presented. One thing that I will elaborate on in Chapter 5 is that advantage will tend to accrue for

---

[18] For previous authors who have used this sort of outcome to represent discrimination see Axtell et al. (2000); López-Paredes et al. (2004); Phan et al. (2005); Stewart (2010); Poza et al. (2011); Bruner (2017); Bruner and O'Connor (2015); O'Connor and Bruner (2017).

[19] Doves meet doves half the time and get a payoff of 2. They meet hawks the other half of the time and get a payoff of 1. These outcomes average to 1.5. Hawks meet doves half the time and get a payoff of 3, and receive 0 the rest of the time upon meeting hawks. Again this leads to an expected payoff of 1.5.

members of the same types across situations. This exacerbates concerns about inequity. Ultimately, types can emerge to play a role in mutually beneficial coordination (the first sort of inequity) but come to facilitate the second sort of inequity, which entails advantages to only one group, and fails to fulfill any beneficial social function. On this picture, even when there is sometimes a benefit gained for the first sort of inequity, the overall result is a disadvantage to one type.[20]

2.5 Other Solution Concepts for Complementary Coordination

As I've hinted at, breaking groups into social categories isn't the only way to solve complementary coordination problems. In particular, there are a few solution concepts that bear some relationship to type-based solutions. The first is from Aumann (1974, 1987), who describes what he calls *correlated equilibria*. These are equilibria where actors use some external correlating device to determine which equilibrium to choose when confronted with games with multiple equilibria. For an example, consider hawk–dove, but let us consider the "chicken" interpretation— you and another actor are both deciding whether to swerve or drive straight in a game of chicken.[21] Suppose that you and the other player really don't want to crash and you secretly decide to use a nearby traffic light to avoid this worst-case scenario. If it is green, you will be the one who swerves. If it is red, your opponent will swerve. In this way, you use further information from the world—the state of the traffic light— to choose an equilibrium.[22]

---

[20] Notice that, of course, this entire discussion centers around payoffs. One might point out that there are ways this framework will fail to capture relevant aspects of inequity inherent in gender roles, such as restrictions on personal freedom. Okin (1989), for example, points out that a Rawlsian should require the abolishment of gender, since restriction to gender roles based on irrelevant types is inconsistent with his notions of political justice.

[21] Because you are apparently in a 1950's greaser gang?

[22] For work in philosophy arguing for the importance of correlated equilibrium type concepts see Vanderschraaf (1995). For an example from biology, consider organisms that broadcast their gametes. This usually occurs in the ocean, and can pose a problem because gametes are tiny, and each faces the challenging task of finding another gamete of the right type to combine with in order to mate. To do this, they must coordinate spatially, and also temporally. This is a correlative coordination problem, then. There are many possible times available for gamete release and individuals do best to release all at once. If females broadcast their eggs three days after males broadcast their sperm so much for all of them.

Skyrms (2014) discusses the evolution of correlating devices to solve the coordination problem in hawk–dove. He points out that strategies incorporating an external correlating device take over populations without such strategies. Where do such correlational devices come from, though? (We can't use traffic lights for everything.) In complementary coordination games, another way to think about the role types and type-conditioning can play in coordination is as a correlating device. The observable difference between players takes the role of the traffic light in providing the extra information necessary for equilibrium selection.

The second concept, arguably more relevant to that of types and type-conditioning in coordination games, is that of uncorrelated asymmetry, from Maynard-Smith (1982) who addresses complementary coordination games in the biological realm. In these games, as he points out, actors can use extra bits of information unrelated to the game in question, or uncorrelated asymmetries, to break symmetry. These allow actors to figure out who plays which role and so turn complementary coordination games to correlative coordination games.

To see how this works, consider another example employing hawk–dove, but now where the players involve a property owner and an intruder on the property. In this problem, there is extra information, which is that one player is the owner and the other the intruder. A single convention such as "play hawk if owner and dove if intruder" can solve the population coordination problem (Maynard-Smith and Parker, 1976; Skyrms, 2014).

With types and type-conditioning, something very similar is happening. Actors are using observable types to turn the game into a correlative coordination game. Now the uncorrelated asymmetries are social categories, and the choices that the entire population arrives at are things like "play A if woman and B if man." There is a disanalogy, of course, because men and women have no chance to play the other role. This also means that they need not keep track of a conditional behavioral rule, (i.e., they don't need to know what to do when in the other situation).

Note both of these concepts can be phrased in terms of information. A correlating device gives extra information to the players which they

Moonlight can play the role of an external correlating device that allows actors to correlate their behavior (Kaniewska et al., 2015). In this case, the solution is both conventional—because there are multiple possible solutions—and functional, because moonlight itself correlates with the state of the tides, so there are better and worse levels of moonlight to use for coordination.

use to choose roles in a coordination game. An uncorrelated asymmetry builds extra information into the structure of the game. This, of course, ties into the idea of conceiving of types as signals to other players. Under this framework, types are literally a way of transferring information to the other opponent to allow for coordination.

•   •   •

Social categories are a solution for groups engaged in complementary coordination problems. They create asymmetries that answer the question, who should do what? In doing so, though, they also create a situation where a preferred role is always taken by members of one category—where men always plough and control livestock, for example. Thus we see a case where inequity is the result of features that play a social functional role. Now we'll fill out this picture further by adding the dynamics of social change. As we'll see, in cultural groups with social categories, cultural evolutionary patterns are radically shifted in ways that lead to norms where different groups play different complementary roles, and where inequity results.