Measuring Conventionality

Cailin O'Connor

June 4, 2020

Abstract

Standard accounts of convention include notions of arbitrariness. But many have conceived of conventionality as an all or nothing affair. In this paper, I develop a framework for thinking of conventions as coming in degrees of arbitrariness. In doing so, I introduce an information theoretic measure intended to capture the degree to which a solution to a certain social problem could have been otherwise. As the paper argues, this framework can help improve explanation aimed at the cultural evolution of social traits. Good evolutionary explanations recognize that most functional traits are also conventional, at least to some degree, and vice versa.

1 Introduction

Driving on the right side of the road is a classic case of convention. This is a stable pattern of behavior that benefits those involved by preventing car crashes, but it could have been otherwise. In many countries, drivers stay to the left. Another classic case is that of linguistic conventions of word meaning. In English, the word "fork" refers to the metal, four-tined tool we use to eat. This pattern of usage allows English speakers to communicate, but could have been

otherwise. Any number of words might have come to refer to this tool, just as successfully.

In his 1969 *Convention*, David Lewis used game theory to develop an influential account of convention. We will say more later about the specifics. One key aspect, though, is that conventions are univocal. Either something is a convention, or not. But in many cases conventions come in degrees. A central aspect of all accounts of convention is that they are patterns of behavior that are *arbitrary* in the sense that they could have been otherwise. (And throughout the paper, when I speak of conventions being arbitrary, this is what I will mean.) But in some cases, there is a more real sense in which a convention could have been otherwise. In the driving case just described, for instance, there are only two options for which side of the road to drive on. On the other hand, there are are an enormous number of words that might have been used instead of "fork", and the likelihood that another culture would develop the same linguistic convention is very low. Arguably the sense in which the "fork" convention could have been otherwise is stronger.

Some previous authors have argued that conventions can come in degrees of arbitrariness. The goal of this paper is to develop a framework for thinking about what this involves, including a concrete proposal for how we might measure these degrees. In particular, I draw on cultural evolutionary modelling tools, and on information theory, to develop this framework and measure. The central idea is that situations where the endpoint of a cultural evolutionary process is less predictable, or more surprising, are more informative, and thus, we might wish to say, more arbitrary.

As will become clear, there are some difficulties for the account I sketch. Most importantly, there are multiple ways to represent any one situation. Because the measure I develop is representation dependent, the choice of repre-

sentation will determine the level of measured arbitrariness. But, as I argue, attending to the explanatory aims of a project can help determine what sort of representation is best suited. And in the paper, I will describe multiple cases where the framework here has useful applications, including to thinking about gendered division of labor, indirect requests, color categories, and patterns of word order.

One main goal of this framework is to improve evolutionary explanation. In explaining the presence or persistence of a cultural trait, we often appeal to functionality, i.e., a trait emerges and persists by dint of improving the well-being or fitness of those employing it. Conventions, on the other hand, are often explained as arbitrary, or the result of chance occurrences. The framework here contributes to recent work that attempts to break down this division. Most cultural traits are functional, to some degree, and yet still could have been otherwise, to some degree, and can only be successfully explained as such. In addition, as we will see, when theorists argue over the degree to which a cultural trait is innate, the measure presented here can be useful.

In section 2 I discuss what conventions are, and where they come from, beginning with Lewis, and focusing on philosophers who argue that conventions come in degrees. I also draw on the work of Brian Skyrms to introduce an evolutionary framework that can represent the cultural emergence of convention. This will set the stage for section 3 where I introduce the measure of arbitrariness. In this section, I also discuss some worries about the measure, including about choosing a representation, and about dealing with probabilities and counterfactuals. The next two sections introduce several applications of the measure. Section 4 discusses gendered division of labor. Section 5 discusses human language, including indirect requests, word order, and color categories. Section 6 concludes with some brief thoughts about the applicability of this

framework to biological conventions.

2 Games and Conventions

In this section I outline how Lewis, and subsequent authors, have used game theory to analyze convention. Then I address how evolutionary models can provide a fuller dynamical picture of their emergence. Throughout, I pay special attention to the idea that conventions might come in degrees of arbitrariness.

2.1 Coordination Games and Conventionality

The account of convention presented by Lewis (1969) starts with games. A game, in the game theoretic sense, is a representation of a strategic interaction, i.e., one involving multiple actors where it matters to each what the others do. Games are defined by four things. *Players* are the individuals involved. *Strategies* are the behavioral options for these players. *Payoffs* define, for each combination of strategies, the utility each player gets.[1] In addition, games typically define *information*, or what each player knows about the structure of the game, and their interactive partner(s).[2]

Lewis introduces a *coordination game*, where actors would like to coordinate their actions, but have multiple ways to do so. The game has two players—player 1 and player 2—with two strategies each—A and B. The payoffs are illustrated in the *payoff table* shown in figure 1. Player 1's strategies are listed in the rows and player 2's in the columns. Each entry gives payoffs for a combination of strategies, with player 1 first.

---

[1] Although there are worries about the concept of utility, these are beyond our purview here. We can take utility to be a representation of what an agent wants, desires, or acts to get. When we move to evolutionary models payoffs will instead represent drivers of evolution.

[2] We will not say too much about information in this paper, since ultimately we will focus on evolutionary models where information becomes less important. But as we will see, for Lewis information, and especially common knowledge, is a central aspect of his account of convention.

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

Figure 1: A simple coordination game.

Because the goal is to coordinate, players get a payoff of 1 whenever they choose the same strategy, and 0 otherwise. This game could represent, for instance, the driving situation described in the Introduction, with A and B corresponding to the left and right side of the road.

For Lewis, conventions are usefully thought of as a special sort of equilibrium in games with multiple equilibria. The coordination game just introduced has two *Nash equilibria*, or sets of strategies where neither player can unilaterally switch and improve their payoff—A v. A and B v. B. Lewis introduces a refinement which he calls a *proper coordination equilibrium*—where no player can unilaterally change strategies and improve payoff for *any* other player. He uses these to identify possible conventions in games, including the two Nash equilibria just described.

For Lewis, game theory gives a precise notion of why conventions might emerge as stable behavioral patterns—at equilibrium each player has a payoff incentive to keep playing the same strategy. In addition, it gives a clear notion of arbitrariness—there are two proper coordination equilibria in this game that the players might have selected. Notice that for Lewis "arbitrariness" of a convention is synonymous with a live possibility that the pattern of behavior in question could have been otherwise. (And, as noted, this is the sense of arbitrariness relevant to the project here.)

To complete his account, Lewis describes specific conditions under which a group playing a proper coordination equilibrium will be at a convention. Members of a group must conform to an equilibrium, must expect other group mem-

bers to conform, must prefer to conform when others do, and must prefer each member to conform when the group does. In addition, Lewis stipulates that group members must have very similar preferences over the equilibria, and must all have common knowledge that the conditions of the convention obtain.

Some of Lewis's conditions have been convincingly critiqued. It is beyond the purview of this paper to overview this literature. And ultimately, one of my aims will be to show that giving necessary and sufficient conditions for whether something is or is not a convention is not a useful way to go in any case. For our purposes, the important take-away from Lewis is that many conventions are usefully represented as populations playing equilibria of games.

Gilbert (1992) in critiquing Lewis, points out that in some cases all of Lewis's conditions will hold, but one equilibrium will be notably better, or more attractive, than the other . Consider the game in figure 2, and suppose that $x = 100$. Both A v. A and B v. B are still proper coordination equilibria, and thus candidates for conventions on Lewis's account. But suppose the players settle on the B equilibrium. Although the choice is arbitrary in that it *could* have been otherwise, that *could* does not seem very strong. Furthermore, if we were to *explain* why the players chose the B equilibrium, we would want a functional explanation that appeals to payoffs, rather than just to chance.

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

Figure 2: A coordination game where B v. B is the preferable equilibrium assuming $x > 1$.

Lewis argues that when the alternatives to behaviors are of a "deficient sort", a group will not be able to generate common knowledge that others will conform to the behavior. Thus, for Lewis the worse equilibrium will not actually

be a viable alternative convention in this case. But we might imagine a set of games where x ranged from 1 to 100. Instead of following Lewis in claiming that at some point the situation changes from one with conventions to one without, it seems more perspicuous to say that there is a range of cases where the conventions are more or less arbitrary.

Millikan (2005), concerned primarily with human language, defines a convention as something that is unlikely to emerge a second time. However, if a group developed a new convention for driving, it is actually quite likely they would end up with a right-side driving convention. (Absent other factors, we might think there is a 50% chance of this.) Again, we might want to say in this case that there is a range of arbitrariness between driving conventions, with two possible equilibria, and word choice, where the possibilities are highly numerous. In other words, both differences in the payoffs of the game and the number of strategies seem to determine how arbitrary the resulting conventions will be.

Motivated by such cases, some authors have argued we should think of conventions as coming in degrees of arbitrariness (though, in general, philosophical accounts have not focused on this aspect of convention (Rescorla, 2019)). Regarding linguistic implicature, Morgan (1977) argues that there are many cases that are both conventional and "natural". Following Morgan, Simons and Zollman (2019) introduce the notion of a *natural convention* to discuss indirect requests. In section 5 we discuss their arguments, but for now what is relevant is their thinking about degrees of naturalness, which provides a starting point for the rest of the paper. They define three dimensions along which a convention could be more or less natural (vs. arbitrary). (1) Some conventions are better, from a payoff standpoint. (2) Some conventions are more likely to emerge. And (3), some conventions are more stable than others. Once they have emerged, they are not likely to be abandoned.

For the game in figure 2, when $x > 1$ the B convention is more natural in sense (1)—it is superior from a payoff perspective. To see what Simons and Zollman (2019) have in mind for their other two scales of naturalness, and also to set us up for the measure introduced in the next section, we will now switch to an evolutionary game theoretic perspective.

2.2 Modeling Evolving Conventions

Lewis (1969) arguably set the stage for such a perspective, but he did not develop it. Notice his definition does not specify how a group arrives at a convention. He supplemented it with a claim that conventions involve mutual expectations about future behavior. I expect you to conform, and I expect that you expect me to conform, etc. These expectations can develop for different reasons, but often result from past behavior. If we have conformed in the past, and both benefited, we have reason to expect conformity in the future.

Subsequent authors, especially Brian Skyrms, have used evolutionary game theory to develop a fully dynamical picture of the emergence of conventions (Rescorla, 2019; Skyrms, 2010; O'Connor, 2019a). In such models, a group engages in repeated strategic interaction. Over time they change strategies as a result of natural selection, learning, or cultural evolution. The models start with a game and add *dynamics*—a set of rules stipulating how the population will change.

In such models, groups often end up at equilibrium. This need not involve common knowledge or complicated sets of expectations, but we might nonetheless think of it as a process by which a convention emerges. For instance, a group playing the game in figure 1 might learn to all play A, and thus solve their coordination problem. The equilibrium involves a self-reinforcing pattern of behavior that might have been otherwise.

8

Evolutionary models allow us to specify what Simons and Zollman (2019) mean by conventions being (2) more or less likely to emerge, and (3) more or less stable. Suppose a group learns to play the game in figure 2. And suppose they change strategies by imitating group members who do well. We can use the *replicator dynamics* to represent this state of affairs. Over time this dynamics selects for strategies that do better than average. For this reason, it tracks both natural selection and cultural imitation well (Weibull, 1997). Under these dynamics, when $x = 1$ the population evolves to play A whenever more than 50% of its members play A, otherwise B.[3]

Figure 3 shows a *phase diagram*, a representation of all the possible population states of a group with two strategies—A or B. To the far left of the line, we have the state where every actor plays A, and to the far right, B. The center point is where 50% of actors play each. The two black spots on the ends of the line show the equilibria (all A or all B). When the population is exactly split, neither strategy does better than the other, so neither is winning. For every other point, though, the population moves towards the two equilibria. We can also identify what are called *basins of attraction*—the set of starting points that end up at each rest equilibrium. The two basins of attraction are the two halves of the diagram, and are of equal size.
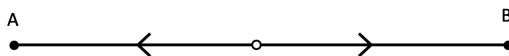

Figure 3: A phase diagram for a simple coordination game.

For a coordination game where $x > 1$, the B equilibrium is more "natural"

---

[3]For the game in figure 1 the replicator dynamics yield one equation for the change in A (the change in B is 1 minus this): $\dot{x} = 3x^2 - x - 2x^3$. Notice that $\dot{x} = 0$ when $x = 0, .5, 1$. These are the rest points of the model. When $0 \leq x \leq .5$ the ride hand side of the equation will yield a negative number. I.e., when the proportion of A is less than half, it decreases further. When $.5 \leq x \leq 1$ the proportion of A is increasing.

in sense (1) of yielding higher payoffs. And as a result it will also have a larger basin of attraction. If $x = 50$, the phase diagram for this model will look like the one shown in figure 4. The basin of attraction for A makes up only 2% of the space, and 98% for B. If we do not know much about the starting place of the population, we can take these basins to represent the probability that each outcome evolves. So on the second scale, again, B is a more natural convention.
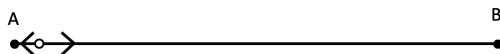

Figure 4: A phase diagram for a simple coordination game where one option yields significantly higher payoffs.

Here, the first two senses of naturalness align because actors playing B tend to do better on average due to its high payoffs. There will be cases where these senses pull apart, though, describing these goes beyond the purview of this paper.

The notion of stability of an equilibrium (naturalness (3)) can typically be assessed by asking: what portion of the population needs to deviate from the equilibrium to disrupt it? Imagine a population at the A equilibrium in figure 3. If a few individuals switch to B, the group will evolve right back to A. But for the model in figure 4, only 3% of individuals need to switch to disrupt the A equilibrium, i.e., it is less stable. Again, for this model the more stable convention (3) also tends to evolve (2) and have higher payoffs (1). And, again, there are cases where these pull apart, but we do not explore them here.

At this point, we have an evolutionary framework that illustrates the argument from Simons and Zollman (2019) about naturalness. As these authors point out, thinking of conventions as coming in degrees matters to cultural evolutionary explanations. (See also O'Connor (2019b).) In particular, these explanations often appeal to adaptiveness or to chance, but rarely to both. If we

recognize that conventions come in degrees, an accurate explanation will appeal to both their adaptive character and to chance.

3   Measuring Conventionality

I now present a measure capturing the degree to which a conventional outcome could have been otherwise. In doing so, I move away from the three division scale proposed by Simons and Zollman (2019), and focus mostly on their second sense of naturalness—probability of emergence. This arguably best tracks whether a convention really could have been otherwise. (Though shortly we will see that sense 3 can also be addressed using the measure.) Could some other outcome have emerged on a cultural evolutionary time-scale, and how likely is it to have done so? In answering this, I draw on *information theory*, first introduced by Shannon (1948), which it typically used to address questions related to the transfer of information.

What is the connection between information theory and convention? As we have seen, different equilibria can have different probabilities of emerging. Equilibria with large basins of attraction are more likely to evolve than small ones.[4] The sizes of the basins thus tell us something about how much information we gain, for some domain, upon seeing what evolved. This level of information will be our measure of arbitrariness. It is higher when there is a lot of uncertainty about what will evolve, and lower for less uncertainty. To be clear, the reason I use information theory here (rather than some other framework) is that it simply is the theory that deals with questions like: how much do we learn upon observing the outcome of a variable process?

The relevant information theoretic measure is *Shannon entropy* which tells us, for some channel, how much information on average is transmitted.[5] An

---

[4]This is not exactly right. More on this shortly.

[5]A channel is just a variable that can transfer information by being in different states. The

amount of information corresponds, intuitively, to how much we learn upon observing something, or how surprised we are to see it. The entropy measure is:

$$H(x) = \sum_i P(x_i)I(x_i) \tag{1}$$

$H(x)$ is entropy. It is calculated by taking the sum, over all the possible signals, i, that can come through a channel, of the probabilities of those signals ($P(x_i)$) multiplied by the information contained in each ($I(x_i)$). The information per signal $I(x_i)$ is equal to $-log_2 P(x_i)$. Why? The negative log of $P(x)$ is 0 when $P(x) = 1$, so a completely certain signal carries no information. This makes sense as we learn nothing upon observing it. As $P(x)$ decreases, $-log_2 P(x)$ increases arbitrarily, meaning that the more unlikely a signal is, the more information it carries. I.e., we learn more upon observing it. Using $-log_2$ will give us the number of *bits* in a signal. Another log gives a different unit of information. Altogether, the measure weights the probability that each signal is sent by the amount of information in that signal to give the average information in the channel.

For example, consider a channel that reports the results of an unbiased coin flip. The probability of heads is $P_H = .5$, and tails $P_T = .5$. The information gained upon observing either of these is $-log_2(.5) = 1$. The entropy is then $.5 * 1 + .5 * 1 = 1$. So on average there is 1 bit of information transferred.

Other things being equal, a channel has higher entropy when the signals being sent are closer to equiprobable. If so, we learn more on average than if, say, one signal is sent 99% of the time, and the other 1%. The entropy of

choice of what counts as a channel is somewhat arbitrary, i.e., channels can be conceptualized in different ways.

this biased channel is just .08. A channel will also have higher entropy, holding other things equal, when there are more signals. The more messages we might receive, the more we learn upon observing any single one.

Let us connect this to our evolutionary models. We take the sizes of the basins of attraction to be our probabilities $P(i)$. Thus, arbitrariness is higher whenever we learn more, on average, upon observing the outcome of an evolutionary process. If there are more equilibria, or their basins of attraction are closer to equally sized, then, holding other aspects fixed, the problem will be more arbitrary, or, we might say, "more conventional".

Notice that while Simons and Zollman (2019) defined naturalness for a particular convention (i.e., B is more natural than A), this measure gives us a level of arbitrariness for a domain. Solutions to x social situation are more arbitrary than solutions to y situation. We can also specify, though, the information value of any particular outcome, $-log_2 P(x)$. We can use this to specify the naturalness of a particular convention a la Simons and Zollman (2019). The closer $-log_2 P(x)$ is to 0, the more natural the convention. In other words, information theory gives us the tools to measure and discuss both how natural a particular evolutionary outcome is, and also how arbitrary solutions to some problem are, on average. In the rest of the paper, though, we will focus on the latter.

Once we employ the measure, we can see that there is only one entropy value, 0, for which cultural traits are not conventional at all. In these cases, one outcome is guaranteed to emerge. For all other cases, traits are at least somewhat conventional. Furthermore, there is no other "end" to the measure. Traits can be arbitrarily arbitrary. This pushes strongly against a framework where we class outcomes into "conventional" and "not conventional". Instead we should expect that almost everything is at least a little conventional, and focus on the diversity of cases within the category of "convention".

While I have described this measure using basins of attraction, it can actually be applied whenever we have some clear way to define probabilities for different evolutionary outcomes. Let us discuss some variants where this is possible.

First, we sometimes have information about the starting point of an evolving population. Maybe a nation is developing a driving convention, but has many immigrants from a right-side driving country, i.e., a high probability of starting with a lot of A players. In such cases, basins of attraction will not adequately track the probability that each equilibrium emerges. What we can do is specify a probability distribution over possible population starting places. This gives us new probabilities for whether a population will start in each basin of attraction, and thus allows us to measure the arbitrariness of the process.[6]

Alternatively, unlike the replicator dynamics, some dynamics are stochastic—each starting point can end up at multiple equilibria depending on chance events. Once again, though, we can redefine the inputs to the entropy equation to get a conventionality measure. Suppose we model ten individuals playing the game in figure 2, with $x = 10$. And suppose we simulate it 10k times, and find that the A equilibrium emerges for 8% of these and B for 92%. We can take these as the probabilities for the two equilibria.

There is another kind of stochastic model, often used by economists (Foster and Young, 1990). These models include continued experimentation. At any point, enough individuals may experiment with a new behavior that the group changes equilibrium. An analysis of such a model sometimes measures how much time, on average, a population spends at each equilibrium. For instance, a group might spend 1/5 of its time at A and 4/5 at B. Once again, we can use these probabilities to measure arbitrariness. Now, though, we measure how much we learn upon observing what convention a population is engaged in at any

---

[6]Other than payoffs and numbers of strategies, we can now see how historical facts and other constraints shape the probabilites of emergence for different outcomes. We will return to this theme in discussing the way this measure is dependent on choices of representation.

given time. For a model like this, notice, stability is reflected in the probability that a population is at that equilibrium. Using this kind of model, the measure can capture actually naturalness (3), stability, as well as naturalness (2) from Simons and Zollman (2019).

As I will discuss presently, the measure can also be applied directly to probabilities derived from empirical data. But first let us address a potential problem.

There are always multiple ways to represent or model any scenario, and this choice will influence what the arbitrariness measure yields. For example, suppose two friends decide which of their two favorite movies to watch every week—Moonstruck or Raising Arizona. We could model this with the game in figure 1. But suppose that while there are only two movies they like, they might also decide to sit and talk instead. Perhaps this is quite unlikely because they love Nicholas Cage, meaning figure 1 is a decent model. But if we develop a game that includes this further possibility, it will yield a slightly higher level of arbitrariness. We might imagine adding more unlikely possibilities to the model and thereby increasing the measured conventionality.

Let us be a little more specific about what is at issue here. Actual chances are not observable. We cannot know the real underlying probabilities for different evolutionary outcomes. Indeed, many argue there are no such probabilities, only our uncertainty about states of affairs. There are deep philosophical issues here, and it is beyond this paper to address them. The point is simply that the probabilities feeding into our measure must be representations of some sort. Given this, there will always be different, plausible ways to choose probabilities for any scenario. For example, in many real world scenarios it will not be plausible to exactly define each possible strategy, and the exact payoffs associated with each.

As we will see in the next sections, in some cases there is rich empirical

data reflecting which conventions have emerged across human cultures. We might take these to be the entries in our conventionality measure, and avoid the thorny problem of choosing a model. This does not actually solve the problem, though. For empirical data, just like models, choices have to be made. Different conventional outcomes in a data set must be operationalized by the scientists involved, and this almost always involves degrees of freedom (Longino, 1990; Anderson, 2004). I.e., the data is also a (somewhat) flexible representation of the underlying truth. And, again, the measure can come out different ways depending on the choice. In addition, models cannot always be abandoned because cross-cultural data often reflects historical influence between societies. Some of the regularity in patterns of behavior results from these influences. Thus there will be cases when, in thinking about degrees of conventionality, researchers may want to use models to represent the "true" arbitrariness of a process, absent historical details.

A further issue arises because, in addition to flexibility in representation, there are often different things one might be trying to represent when it comes to cultural evolutionary processes. These processes are extended in time and space, constrained by other cultural factors, and shaped by surrounding cultural influences. For example, if women in some culture hunt small game, this may mean it is more likely that they also process skins from small game. How, then, does one think about the level of arbitrariness of the emergence of gendered division of labor in skin processing? In choosing a representation one has to pick how many of these constraints and influences are included. These choices will sometimes strongly shape the degree to which something "could have been otherwise".

There is not a final, ultimate solution to these worries. The best response is that representations should be sensitive to the explanatory projects they are

aimed at. If a researcher asks, "how conventional is the emergence of gendered division of labor absent other constraints?" her best representation might be an idealized model that considers how sex differences shape cultural evolution. If she asks, "how conventional is food preparation in cultures with plough-based agriculture?" she might want a data set of real patterns of conventional behavior in plough-based cultures. As we will see, in some cases researchers might have reason to compare arbitrariness across several different representations of the same case. Of course, even a well tailored representation will not be perfect. There will still be degrees of freedom in how it is created. The inherent limitations in representations, and thus the measure here, should be recognized, but should not be taken to prevent good (enough) inquiry.

With this picture in hand, let us briefly return to one more worry about probabilities before continuing on to cases. If there are no real physical chances, there is no real sense in which any convention could actually have evolved to be otherwise. And, if so, there are no conventions in the sense developed here. Recognizing that the measure is representation-dependent solves this problem, though. We can specify what we mean by probability in the context of a particular model or data set (basins of attraction, probability of emergence over many rounds of simulation, time spent at equilibrium, percentage of real societies that adopt a behavior) and set aside worries about probability and chance in the world.

We now apply the measure to four cases of real conventions. The goal is to demonstrate how this framework is useful in thinking about cultural evolutionary explanation. Sometimes just conceptualizing of conventions as coming in degrees improves explanation. Sometimes having specific numbers allows for comparisons across different evolving conventions. Sometimes such numbers allow for comparison between data sets and models that might help with debates

17

over just how innate different cultural traits are. In each case the goal is not to settle the issues at hand, but rather to show in principle how this framework can be applied to such problems.

4   Gendered Division of Labor

Murdock and Provost (1973), in a classic article, look across 185 cultures and document who performs 50 different "technological activities" including things like rope making, house building, and vegetable harvesting. They find significant division of labor by gender. They also find, though, variation in the patterns of division of labor for different tasks. Some tasks, like big game hunting, are almost always performed by men cross-culturally. Some are predominantly performed by women, such as spinning, laundering, and dairying. There are many other tasks generally performed by one gender in any particular society, but with variation across cultures as to which gender it is. Caring for small animals, carrying burdens, and making rope all fall into this category.

In other words, we see a range of conventionality. In some cases, there is a reason for one gender to adopt certain tasks—males have stronger muscles, and females give birth, and thus big game hunting tends to be done by men (who, in most cultures, are typically male). In other cases, there are no clear reasons for either gender to adopt the task. In yet other cases, there may be some small advantage to a task being performed by one gender. Let us develop a model to represent this situation, and discuss how the measure would apply. Then we will briefly discuss a real debate over the emergence of gender roles that benefits from the account developed here.

Gendered division of labor is well-modelled by an *anti-coordination game.* This moniker is somewhat misleading since the actors still want to coordinate, but do so by taking different, or complimentary, roles instead of doing the same

thing. In one household, for example, it is beneficial to have a few individuals doing the dairying and others doing the well-digging, but you typically do not want all individuals doing the same tasks. Most tasks require extensive skill learning, and it is much more efficient when actors do not need to each learn all the necessary skills of household management (Blood and Wolfe, 1960).

Figure 5 shows a game. Two actors in a heterosexual household choose between some target task (dairying, rope making, etc.) and performing other work that benefits the household.[7] If they assign the task to just one of them, they manage to coordinate. But because of sex based asymmetries, or pre-existing gender norms, one of these assignments may be more payoff-beneficial than the other. This asymmetry is tracked by the values of α and β. These values might be equal in some cases, or a particular division of labor could benefit one partner more than the other.

Player 2

	T. Task	Other Work
T. Task	0,0	1,1
Other Work	1+α,1+β	0,0

Player 1 (label to the left of the rows)

Figure 5: An anti-coordination game with the potential for payoff asymmetries. Assume α, $\beta > 0$.

As we increase the values of α and β, the better equilibrium will slowly become more and more likely to evolve under most dynamics. Figure 6 shows basins of attraction for this game evolved under the replicator dynamics.[8] As $\alpha = \beta$ increases, the basin of attraction for the preferable equilibrium increases as well. On the measure, then, this evolutionary scenario becomes less and less arbitrary as $\alpha = \beta$ increases. When $\alpha = 0$, the arbitrariness level is 1. When $\alpha = 4$, it is .47. When $\alpha = 9$, it is .24. (Note that the particular values are not

---

[7]The assumption of heterosexuality obviously limits the analysis here.

[8]To be precise, this is a model using the two-population, discrete-time replicator dynamics. Basins of attraction are estimated using 10k runs of simulation.

important here, just the idea that we can formally characterize these cases as varying with respect to conventionality.)
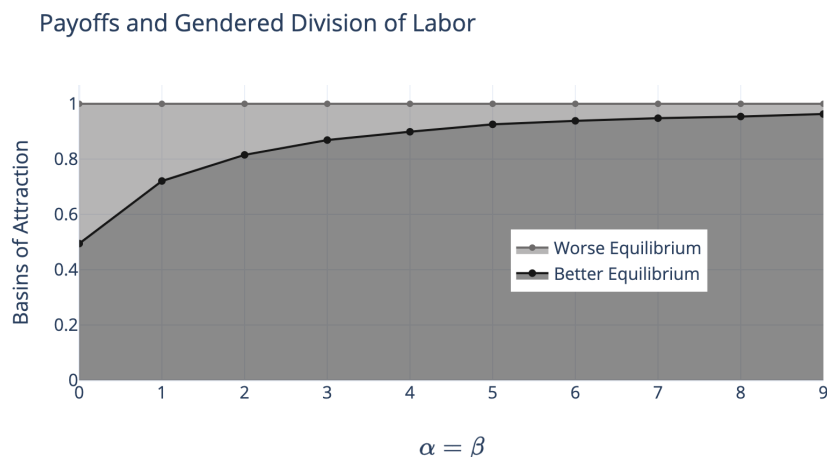
Payoffs and Gendered Division of Labor

Figure 6: In a replicator dynamics model, the basins of attraction for a preferred equilibrium get larger as it provides more payoff benefits.

Another alternative is to apply the measure directly to the cross-cultural data gathered by Murdock and Provost (1973).[9] Across their sample the hunting of large aquatic animals has a conventionality of 0. (Only men ever carry out this task.) Spinning has a conventionality of .52. (It is performed mostly by women cross-culturally.) The making of leather goods has a conventionality of .99, as it is performed by men and women almost equally cross-culturally. Having the measure is useful here in that it allows us to compare these different arenas, and the degree to which gender roles are conventional in each.

In the previous section, I argued that the representation used should depend on the explanatory goals of a project. Notice that the two approaches just

---

[9]In doing so, I ignore tasks that are performed both by men and women in a significant number of cultures. I also combine the ratings "performed exclusively by males/females" and "performed predominantly by males/females" in the initial study to get numbers that are more comparable to the game shown in figure 6. Of course, there are other ways to go about this analysis. This reflects my earlier claim that there will always be degrees of freedom in the production and use of data.

described each have advantages, and disadvantages, that make them more or less appropriate given explanatory goals. The real data does not reflect the level of conventionality absent influence between cultures. The model, on the other hand, avoids such possible influences, but is so simplified that it yields probabilities that may not track real processes well.

Beyond the details of the measure, the general picture can help us think about explanations for the emergence of gendered division of labor. Some evolutionary psychologists claim women have evolved psychologies that disincline them towards competition, and that thus explain why women tend to hold certain social roles. (See Buss and Schmitt (2011).) Alternatively, social structural theorists argue gendered division of labor comes first, and creates psychological differences between genders as they fill different roles (Wood and Eagly, 2012). One complaint against social structural theory is that it fails to explain why we have the gender roles we do. Because it does not appeal to function, it suggests the roles are arbitrary, which does not seem to fully fit the data. Social structural theorists respond that ecological conditions determine which gender roles emerge in the first place (Wood and Eagly, 2012).

But the picture developed here gives a much better response. Gendered division of labor *is* arbitrary, but this arbitrariness comes in degrees. In many cases, there is a strong sense in which patterns of labor are conventional, and could have been otherwise. In other cases, explanations of gendered division of labor should appeal to function, but only partially. In such cases, evolutionary psychology's emphasis on function as the primary determinant of cultural traits is misleading.

5   Language

Human language is a classic case of conventionality. It was so important to Lewis (1969) that he introduced the *Lewis signaling game* to capture the strategic situation of those engaged in communication, and to argue that linguistic conventions can emerge as a result of expectation and precedent. But do linguistic conventions come in degrees? And, if so, does the measure introduced here apply?

Let us first consider conventions of semantic meaning. Some of these conventions are more natural than others. For example, there is cross-linguistic regularity in the use of "ma" sounds to refer to mothers because of constraints on early infant speech.[10] But it makes little sense to model most of these cases with a game that has some set number of strategies. There are an extremely large number of terms that might be successfully used to mean fork, or even to mean mother.[11] Some of these will be more or less likely to emerge for the task (for instance, very long words may be less likely than shorter ones). It is not practical, or useful, though, to give a model estimating the number of possibilities, and their probabilities of emergence.

There are many other types of linguistic convention, though. And some of these are more reasonable targets of analysis here. Below we consider three examples.

5.1   Indirect Requests

As mentioned, Morgan (1977) and Simons and Zollman (2019) introduce notions related to degrees of naturalness in thinking about conversational implicature, and, in particular, indirect requests. An example is, "can you hand me the

---

[10]To give another example, the word "bo" apparently means cow in both Gaelic and Vietnamese. Booooooo.

[11]For this reason, cross-cultural data also will not be very helpful. There will be too few societies that evolved the same sorts of conventions.

salt?" which is stated as a question about the ability of the recipient, but is actually a request for salt. The meaning, in this case, depends upon implicature. The recipient is aware that there is no particular reason for the speaker to be interested in their ability to pass the salt, but that there is a reason for the speaker to want the salt. From this, and from knowledge about how human communication usually works (for example, we usually speak about things that are relevant) the recipient can infer that the speaker wants the salt, and is not looking for a reply like, "yes, I can".[12]

Previous authors have argued that pragmatic meaning of this sort is not conventional. It depends on inference, rather than convention, in allowing effective communication (Gordon and Lakoff, 1975). Searle (1975), however, points out that certain phrases are used to make indirect requests, and not others. For instance, we would not usually say, "do you have the capacity to pass me the salt?". For this reason, we might suspect there is some level of conventionality at play. Morgan (1977) argues that indeed it is convention that determines which phrases are commonly used and immediately understood to carry their implied meaning. Simons and Zollman (2019) agree and, in addition, argue that the use of indirect requests itself is conventional, though using indirect requests is a highly natural convention on their account.

Let's consider the emergence of strategies that either use indirect requests (IR) or do not (¬IR). Simons and Zollman (2019) give the payoff table in figure 7 to capture payoffs actors might garner for these strategies. This is a coordination game with two proper coordination equilibria—IR and ¬IR. If actors do not coordinate on the use of indirect requests, there will be miscommunication. For instance, Orla might ask Liam "can you pass the salt?" and Liam might respond "yes". However, in this game, the IR equilibrium garners a higher payoff because

---

[12]Grice (1991) gives an influential account of implicature that includes certain maxims of conversation, such as that one should be relevant. Though see Lepore and Stone (2014).

in real languages indirect requests provide an easy way for speakers to be less direct, and thus less rude, in making requests. Also, those using IR tend to do better in communicating with those who do not than vice versa. Those who do not conventionally use indirect requests will still be able to use pragmatic considerations to understand such requests, but will fail to avail themselves of the polite language their peers expect.

		Player 2	
		IR	**¬ IR**
Player 1	**IR**	10,10	5,3
	¬ IR	3,5	8,8

Figure 7: A game representing the use of indirect requests in a language.

In a game like this, with two options, the conventionality level of the problem would be 1 if each option were equally likely to emerge. Here, though, the IR equilibrium is more likely to evolve both because it does well against itself and because it does well against the other strategy. The level of arbitrariness for the problem is thus only .88 bits. These values can support the notion that this strategy is, indeed, conventional (i.e., well above 0), but also partly natural in the sense that there is some explanatory reason for why IR might typically evolve.

5.2   Basic Word Order

Consider another language case. Basic *word order* refers to the order in which subject (S), verb (V), and object (O) will typically be arrayed in a sentence. There are six possibilities: SVO, SOV, VSO, VOS, OSV, and OVS. Different languages demonstrate different canonical word ordering. (English, for example, is SVO—dog bites man.) In other words, word order is conventional across languages. Furthermore, the logical structure of this example gives us six well-

defined possible outcomes of the cultural evolutionary process. Thus, unlike the case of semantic meaning, this is a plausible place to apply the conventionality measure.

Across 402 languages, Tomlin (2014) finds that 45% of these are SOV, 42% are SVO, 9% are VSO, 3% VOS, 1% are OVS, and no languages are OSV. There are various theories about why there is such an uneven distribution. Maurits et al. (2010) give a partly functional account which appeals to the concept of *uniform information density*. In order to reduce errors and maximize efficiency, there are many features of language which spread information transfer out as evenly as possible. As these authors point out, the word orders that best preserve uniform information density are also common ones. I.e., there are functional reasons to choose one word order or another, and yet also conventionality with respect to what different languages settle on.[13]

There are different ways to answer the question: how conventional is a language's choice of word order? One option is to apply the measure directly to the linguistic data. If word order were completely arbirary over the six various options, its level would be 2.59. Given the real distribution, the measure yields 1.57. In other words, comparing the data to a model where word order is as conventional as possible indicates that there may be some contraints on the evolutionary process, maybe the functional ones outlined by Maurits et al. (2010).

However, these is also another way to think about the issue. Real languages bear historical relationships to one another, which means that they sometimes share common features as a result of these relationships, rather than in response to payoff demands. This may explain some of the word order data. We could instead construct a game to model how beneficial it is for agents to use different

---

[13]Their account does not perfectly fit the data. Our goal, though, is not to assess the goodness of their account but to use it as an example.

word orders. We could then use this model to measure arbitrariness on the assumption of functional differences, abstracted away from historical influence.[14] The arbitrariness in this model, compared to that of the real data, could give a sense of how well functional factors might explain real patterns of word order.

5.3 Color Categories

Let's consider one last case related to human language—that of conventionality in color terms. Color terms are conventional in that in each language has different sounds to represent different colors. They are also conventional in that each language groups hues into color categories differently. Some have just a few color terms, while others have many more. In addition, the boundaries between these terms are conventional. For example, some languages group green and blue into a single color category, while in others light green and dark green are two different colors.

There is a debate over the degree to which color terms are "universal". Human color space forms (approximately) a spindle shape where the different colored hues—red, orange, yellow, etc.—circle the equator of the figure, and the poles correspond to light and dark. Color categorizations divide this shape into regions associated with different terms. Relativists argue that these divisions are largely conventional, and mostly respond to human needs for clear communication given a set of perceptual and cultural constraints (Roberson et al., 2000). Universalists, on the other hand, think that within the space are perceptually salient hues—red, green, yellow, blue, white, and black—that act as anchors for the locations of different terms (Hardin, 2005; Kay, 2005). These anchors, they argue, explain cross-cultural regularities in color categorization. Both sides of this debate recognize that color terms are at least partly shaped by perception,

---

[14]Although I will not provide such a game here, the work of Maurits et al. (2010) provides a natural way to develop one. They use real data sets to calculate how close each word order is to ideal, which provides error rates that could determine payoffs in a signaling game.

and partly arbitrary, or conventional. The question is: to what degree do facts about human perception constrain this evolutionary process?

The measure here could tell us something about both 1) the number of basic terms used per language and 2) the locations of category centers and boundaries. Let us consider 1), since this is a much easier topic to tackle. The number of basic color categories across human cultures ranges from two to eleven (Berlin and Kay, 1991).[15] Berlin and Kay (1991) consider 98 languages and give numbers of basic color terms for each.[16] On our measure the conventionality of this arena is 3.01. If there was no regularity to these numbers, i.e., if we saw equal numbers of languages using these different numbers of terms, the conventionality would be 3.32. On the other hand, if there were full regularity, i.e., if every language was so constrained by perception that they used the same number of color terms, the conventionality would be 0. One might argue that the comparison here sheds light on the debate about universality. Alternatively, one might compare these real conventionality levels to those that emerge in evolutionary models of color terms that do or do not include perceptual salience. This would give some information about how conventional the process is expected to be when there is no salience, and when their is such salience, and thus provide more apt comparisons for the actual data.[17]

In this section, we have seen three examples of linguistic conventions that sit along a continuum of arbitrariness. In each case, understanding the emerging patterns of behavior as both conventional and natural is perspicuous. Furthermore, this understanding can help with explanation. In the case of indirect speech acts, we see that there is no reason to debate whether they are conven-

---

[15]A basic color term is recognized by all language users, used consistently, and applicable to any domain (unlike, say, blonde) (Hardin, 2005).

[16]See their data presented on page 22.

[17]Universalists do have an explanation for cross-cultural variance. They argue that there are stages of development for color languages. Again, the goal here is not to successfully contribute to this contentious debate, but to give a proof of purpose about how the framework in this paper might be useful to it.

tions or not. They are conventions, but this does not mean there are no reasons related to implicature that explain their presence. In the case of word order, we can see that while information needs may give some reason for certain word orders to evolve, this evolution can also be partly conventional. In addition, the measure can give concrete ways of comparing data to models to get a sense of the role of information needs play in constraining linguistic evolution. In the case of color categories, we see two sides of a contentious debate where the topic is the level of conventionality or arbitrariness of categorization. How well can categories be explained by appeal to human color perception? To what degree do we need to appeal to the communication needs to human groups? To adjudicate this, formal tools like the measure here may be useful.

6   Conclusion

In this paper, I have supported the idea that conventions should be understood as coming in degrees of arbitrariness. We have seen a number of examples where only this sort of conception can make sense of the relevant phenomena, and allow for appropriate evolutionary explanation. In addition, I have used several formal tools, including game theoretic models and an information theoretic measure, to clarify and support this picture.

By way of concluding, I want to briefly discuss biologial conventions. Under Lewis's account, there are no conventions outside of human societies. No other animals have the abilities to meet his common knowledge requirements. But when we move to an evolutionary picture, many of the models used to represent the emergence of convention in human societies can be aptly applied to the evolution of behavior in other species. The signalling game, for example, has been widely applied to non-human animals, from scrub jays to vervet monkeys (Skyrms, 2010; Searcy and Nowicki, 2005). If we take a bare bones notion of

convention—a stable pattern of self-reinforcing behavior that could have been otherwise—we can see that there are many biological cases that fit the bill. Perhaps most importantly, there are many biological cases where explanations of patterns of social behavior should appeal both to chance and to function. In such cases, the measure developed here might be useful in thinking about the degree to which evolutionary processes are constrained. As with word order, and color categories, it might provide a useful tool, for instance, in comparing data and models to give a better sense of how much ecological conditions, and current phenotype, constrain the emergence of behavioral traits

References

Anderson, E. (2004). Uses of value judgments in science: A general argument, with lessons from a case study of feminist research on divorce. *Hypatia*, 19(1):1–24.

Berlin, B. and Kay, P. (1991). *Basic color terms: Their universality and evolution.* University of California Press, Oakland, CA.

Blood, R. and Wolfe, D. (1960). *Husbands and Wives: The Dynamics of Married Living.* Free Press, New York.

Buss, D. M. and Schmitt, D. P. (2011). Evolutionary psychology and feminism. *Sex Roles*, 64(9-10):768.

Foster, D. and Young, P. (1990). Stochastic evolutionary game dynamics. *Theoretical population biology*, 38(2):219–232.

Gilbert, M. (1992). *On social facts.* Princeton University Press, Princeton, NJ.

Gordon, D. and Lakoff, G. (1975). Conversational postulates. In *Syntax and*

Semantics: Speech Acts, volume 3, pages 83–106. Academic Press, New York, NY.

Grice, H. P. (1991). *Studies in the Way of Words.* Harvard University Press, Cambridge, MA.

Hardin, C. L. (2005). Explaining basic color categories. *Cross-Cultural Research*, 39(1):72–87.

Kay, P. (2005). Color categories are not arbitrary. *Cross-Cultural Research*, 39(1):39–55.

Lepore, E. and Stone, M. (2014). *Imagination and convention: Distinguishing grammar and inference in language.* Oxford University Press, Oxford, UK.

Lewis, D. (1969). *Convention: A philosophical study.* Blackwell, Hoboken, NJ.

Longino, H. E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry.* Princeton University Press, Princeton, NJ.

Maurits, L., Navarro, D., and Perfors, A. (2010). Why are some word orders more common than others? a uniform information density account. In *Advances in neural information processing systems*, pages 1585–1593.

Millikan, R. G. (2005). *Language: A biological model.* Oxford University Press, Oxford, UK.

Morgan, J. L. (1977). Two types of convention in indirect speech acts. *Center for the Study of Reading Technical Report; no. 052.*

Murdock, G. P. and Provost, C. (1973). Factors in the division of labor by sex: A cross-cultural analysis. *Ethnology*, 12(2):203–225.

O'Connor, C. (2019a). *Games in the Philosophy of Biology.* Cambridge Elements in the Philosophy of Biology.

O'Connor, C. (2019b). *The Origins of Unfairness.* Oxford University Press, Oxford, UK.

Rescorla, M. (2019). Convention. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy.* Metaphysics Research Lab, Stanford University, summer 2019 edition.

Roberson, D., Davies, I., and Davidoff, J. (2000). Color categories are not universal: replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, 129(3):369.

Searcy, W. A. and Nowicki, S. (2005). *The evolution of animal communication: reliability and deception in signaling systems.* Princeton University Press, Princeton, NJ.

Searle, J. (1975). Indirect speech acts. In *Syntax and Semantics: Speech Acts*, volume 3. Academic Press, New York, NY.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423.

Simons, M. and Zollman, K. (2019). Natural conventions and indirect speech acts. *Philosophers' Imprint*, 19(9):1–26.

Skyrms, B. (2010). *Signals: Evolution, learning, and information.* Oxford University Press, Oxford, UK.

Tomlin, R. S. (2014). *Basic Word Order (RLE Linguistics B: Grammar): Functional Principles.* Routledge, Abingdon on Thames, UK.

Weibull, J. W. (1997). *Evolutionary game theory.* MIT press, Cambridge, MA.

Wood, W. and Eagly, A. H. (2012). Biosocial construction of sex differences

and similarities in behavior. In *Advances in experimental social psychology*, volume 46, pages 55–123. Elsevier.