PLAYING BY THE RULES: USING GAMES TO STUDY SOCIAL NORMS

Martina Valković

In this article, classic game theory and evolutionary game theory are used to explain how social norms might come into existence. The norm of distributive fairness is taken as a case in point, and illustrated by a simple example of dividing a cake.

However little we can say about a topic as controversial as human nature, one thing is surely beyond reasonable doubt – we are social animals. We feel the need to be around other people, to interact with them and to be acknowledged by them. Social interaction is a source of joy and learning. It enables us to form valuable and lasting relationships and it is the basis for the formation of our personal and social identities. There is also the other side of the coin – social interaction is often a source of less pleasant experiences and phenomena, such as enmity and conflict. Even so, when we are denied the presence of others for a certain time, we become lonely, sad, and our world is greatly impoverished. Indeed, shunning, social rejection and ostracism were often considered as some of the worst punishments that can be inflicted on an individual. Social groups and communities in which we live shape our experience and often constitute the range of our potential life choices.

Furthermore, our social worlds are not nearly as chaotic as they may first seem. On the contrary, they are shaped

doi:10.1017/S1477175622000045
Think 62, Vol. 21 (Autumn 2022)

by numerous rules, the most obvious of which are laws that prescribe or prohibit certain behaviours, and the violation of which results in a punishment carried out by an authority. However, a great majority of the rules governing our social interactions are not codified in law; they remain unwritten and are enforced not by some central authority, but by the individual members of the society themselves, more through negative reactions to violations than through some more palpable punishment. These rules are called social conventions and norms. They constitute what we deem to be acceptable social behaviour, and we are all aware of them. We all know that we should keep a certain spatial distance from our interlocutors, put our hands on our mouths when sneezing or yawning, divide a cake into two equal pieces between the two of us, or offer a helping hand to someone whose groceries have spilt out of her shopping bag. Equally so, we know that we should refrain from calling people nasty names in their presence, wearing a three-piece suit at an informal children's birthday party, or lowering our pants in public. These conventions and norms may vary between populations and contexts, and we are again, most of the time, aware of them and adjust our behaviour accordingly. So, when we visit Japan, we learn to stand further away from our interlocutors than we do in Argentina, and although we would never think of lowering our pants in the middle of an office meeting or a grocery store, we do so cheerfully at beaches and saunas.

Even though many conventions and norms vary greatly between various populations, some are less variable and even seem largely universal. In all or almost all cultures, asocial behaviour such as theft or unprovoked attack are frowned upon. Other norms, such as those guiding the fair distribution of resources or mutual aid, vary considerably less than, say, rules of personal space or rules of etiquette. We usually think of these norms as non-arbitrary and as reflecting important social phenomena, such as justice. Considering the great importance of these social rules for our daily lives, an interesting question is: where did they

https://doi.org/10.1017/S1477175622000045 Published online by Cambridge University Press

come from? An easy answer would be that we inherited them from our parents and grandparents or that they were once imposed on us by someone, but that just pushes the question further – where did our ancestors get these rules from and how did they survive until the present day? In other words, how were our social norms established? How could they have evolved spontaneously, as a result of cultural evolution, without some grand design on anybody's part?

One way to approach this question is by utilizing game theory. Game theory studies the underlying logic of social interactions, or games as they are called within the field, and it has shown itself to be a useful and fruitful tool in philosophy and the social sciences. It aims to describe, explain and/or predict the behaviour of agents in interactive scenarios by modelling their strategic behaviour, using agents' incentives and outcomes of their actions and assuming a level of rationality.

Although the modern field of mathematical game theory was established only in the 1950s by the mathematician John von Neumann and the economist Oskar Morgenstern, game-theoretical thinking has existed for a long time. For instance, it can be seen in the work of Thomas Hobbes, who tried to ground his social contract theory on his account of what rational agents would do to escape a dire situation of the state of nature in which they find themselves. Similarly, Jean-Jacques Rousseau gave an example of a group of hunters, who can only catch a stag, their preferred catch, if each and every one of them stands at their designated spot. If just one of them abandons the position, the stag gets away and they all end up hungry. However, a single hunter can catch a rabbit, which is a less desirable, but certain, prey. So, what should a hunter do? It is in her best interest to stand in her place to participate in catching the desirable stag, but only if the others do so, too. If she has reasons to doubt that the others will remain at their positions, the best option for her is to go solo and catch that rabbit, so that at least she will have something to

https://doi.org/10.1017/S1477175622000045 Published online by Cambridge University Press

eat. In this way, her best option depends on actions of others, or, rather, on what she expects those actions will be. For obvious reasons, the stag hunt later became a general name for the type of game in which we can gain a lot only if we are confident enough that others will do their part of the deal; otherwise we are destined to settle for a smaller gain. It is not hard to see why interactions like these, if repeated often enough, could result in the establishment of a convention or a social norm, which could prescribe staying in place when hunting stag.

However, if we are interested in how our social norms have or could have developed, classic game theory might not be the most suitable approach. One reason for this is because it requires high levels of rationality and deliberation on the part of agents, a requirement which may be too high for many contexts. Instead, we might be more successful by turning to evolutionary game theory. This approach does not have to assume *any* rationality on the part of the agents involved in the interaction, because it models their decision-making strategies as a result of learning. This considerably weaker requirement allows for the application of the models to a much wider range of situations.

As its name might hint, evolutionary game theory originated in the application of mathematical models of game theory to biological contexts, particularly the study of sex ratios in mammals and evolutionary biology (Alexander 2009). More recently, philosophers and social scientists got into the field, due to the insight that evolutionary game theory does not have to apply only to biological evolution, but also cultural evolution, and can be a useful device for studying the emergence and development of social norms. Possibly the most important work in this regard has been that of the philosopher Brian Skyrms in his influential books *Evolution of the Social Contract* (1996) and *The Stag Hunt and the Evolution of Social Structure* (2003). In the former, Skyrms uses evolutionary game theory to show how social phenomena such as distributive justice, mutual aid and ownership behaviour might have evolved. In what follows, I

https://doi.org/10.1017/S1477175622000045 Published online by Cambridge University Press

will use his analysis of the evolution of distributive justice norms, modelled on a simple bargaining game, to highlight the difference between classic game theory and evolutionary game theory, and show how evolutionary processes could have resulted in the emergence of fairness norms. Since the norm of distributive justice is, arguably, one of the most important norms governing our behaviour, success in 'retrieving' this norm by using evolutionary game theory would be an important result.

Imagine finding yourself in a room with another person, who will be your fellow player, a delicious cake to split between the two of you and a referee monitoring your actions. You and your partner, or opponent, are both equally (un)deserving of the cake – neither of you worked harder for it, your hunger levels are about the same, and your positions are symmetrical in every other sense. In other words, neither of you has any special claim to the cake. Each of you wants as big a share of the cake as possible and, let us assume, does not care about the share your opponent gets. You are forbidden from communicating with each other, and you are requested to write a percentage of the cake you want to claim for yourselves on pieces of paper which you then hand to the referee. If your requests amount to a total of 100% or less, you get the percentage you requested. But there is a catch – if your claims total over 100%, the referee keeps the cake to herself and neither of you gets any.

At this point, you may think that the obvious solution is to ask for a half of the cake, that is to say, to split it into two equal parts. If this is your intuition, you are not alone, since the experiments similar to our divide-a-cake game regularly result in subjects requesting such a division. But the game still remains interesting, since we can ask ourselves: why does this rule of 50:50 split assert itself as the right or fair one? Again, the question is: how did our norm of fair division evolve?

We use game theory to approach the subject. Hence, we can argue that the fifty-fifty split is the result of people

https://doi.org/10.1017/S1477175622000045 Published online by Cambridge University Press

being governed by their informed rational self-interest. As we said, you want to get as much cake as possible, and so does the other person. Your outcomes are intertwined in such a way that what is the best claim for you to make depends on what the other person requests. You do not want your claims to add up to a total of less than 100%, since that means that there is still some unclaimed cake that you could have had. More importantly, you do not want your claims to add up to over 100%, because that means that you will not get any cake at all. Your positions being perfectly symmetrical, the solution of your problem is a combination of your requests such that neither player could do better in the game by changing their request, given the other player's request. Such a combination of players' strategies is called the Nash equilibrium. A stronger concept is the strict Nash equilibrium, in which you would certainly do worse by changing your strategy. Thus, we can say that the situation in which both players ask for a half of the cake is a strict Nash equilibrium. A player who would ask for less than a half would get less, and the one who would ask for more would get nothing, since the total would be more than 100%.

The concept of the Nash equilibrium, however, cannot be used to account for our norm of equal distribution. As Skyrms points out, our little division game has many strict Nash equilibria. More precisely, every combination of requests which equals 100% constitutes such an equilibrium. If one participant requests, for example, a third of the cake, and the other one two thirds, their combination is still a strict Nash equilibrium. The reasoning is the same as in the equal split. A person requesting less (the assumption is always that the other person's request remains the same), would have got less cake, so they would do worse by changing their request. Similarly, a person changing their strategy by asking for more would also do worse, because the total of requests would exceed 100%, meaning that they would not get any cake. This indicates that informed rational self-interest, and the concept of equilibrium coming

https://doi.org/10.1017/S1477175622000045 Published online by Cambridge University Press

with it, does not deliver when it comes to the evolution of norms.

Fortunately, evolution comes to the rescue. Skyrms turns to studying the effects of evolution on our strategies, and builds a model with a large population from which participants in the game are chosen at random. In this model, our cake turns into a measure of Darwinian fitness, meaning the expected number of offspring, and the idea is that individuals do not pick their strategies, but come preprogrammed and pass on their strategies to their offspring. The evolutionary fitness, the measure of success, is determined by the interactions, and decides which strategies will survive and in what quantities.

This model gives the following results. Individuals in a population where the usual request is over 50% do poorly, as they do not get anything from their interaction. This enables individuals who ask for less than 50%, and thus do not break the 100% limit, to do a bit better than the average member of the population. The same goes for an individual who asks for a bit more than 50% in a population where the usual demand is under 50%. Thus, when we add evolution into the mix, our equilibrium possibilities are narrowed to two strategies: Demand 50% and Demand 100%.

But it gets even better. We get to exclude Demand 100%, on the grounds of it not being a stable equilibrium. Unstable equilibria are prone to 'mutant invasions'. Specifically, since in a population where everyone requests the whole cake for themselves, no individual ever gets any cake, a small number of 'mutants', individuals who settle for a half of the cake or less, can fare better than the 'hosts' asking for 100%, and this will eventually cause their numbers to increase. Demand 50%, on the other hand, is a stable equilibrium. It cannot be invaded by mutants asking for either less or more, since such individuals will always fare worse than their fair hosts. This leaves the equal distribution as the only evolutionary stable equilibrium strategy of our game, a result which is in agreement with our intuitions and norms.

https://doi.org/10.1017/S1477175622000045 Published online by Cambridge University Press

So, we can show that the norm of distributive justice must emerge as a result of evolutionary processes working on populations in which individuals interact while following different strategies. That means that cultural evolution forces us to be fair and that there is, after all, hope for a robust naturalistic account of our social norms and moral intuitions, right? Well, no. It turns out that we should not get our hopes up as yet. As Cailin O'Connor shows in her forthcoming book *Dynamics of Inequity*, cultural evolution can also result in less rosy conventions and social norms, with consequences that include, but are not limited to, inequity, discrimination and, yes, distributive injustice.

Recall that in our divide-a-cake game above we assumed symmetrical roles for both you and your fellow player. You not only had the same needs and merits, you were also indistinguishable from each other, and the same was the case for the evolutionary version of the game. Now, let us change that condition by introducing asymmetrical roles, or certain characteristics which players can notice and use in order to divide a group into types. These characteristics may vary from the colour of your shirt or eyes, to the colour of your skin, your sex or gender, or some outward sign of religious affiliation. It turns out that this division into types makes possible the emergence of stable equilibria which we would deem anything but fair. Imagine that the population somehow agrees that, say, men get 80% of the cake, and women 20%. This is a robust equilibrium which cannot be reached in a group undivided by social categories. Even more worrying is the fact that all participants in the interaction can be perfectly rational – men and women trapped in the 80:20 split are doing the best they can, considering their social environment. For example, women cannot simply break the harmful norm by demanding more than 20%. This would result in them getting no cake at all, and 20% is still better than nothing, especially if your bargaining position is not that strong to begin with (which is highly probable, considering that you have been getting 20% all the time, while others were

https://doi.org/10.1017/S1477175622000045 Published online by Cambridge University Press

getting four times as much). An apparently perverse result is that the 80:20 split is, however inegalitarian, mutually beneficial in this context – unilateral straying from that norm would result in both participants faring worse.

So, how should we interpret the results gained from our game-theoretic models, and what future venues should we consider for exploration within the theory? Some authors, such as Ken Binmore (2006) and Robert Sugden (2005), have argued that we can draw normative conclusions from studying evolutionary game theory, and that it can serve as a basis of our ethical systems. In other words, by describing how our norms evolved we already provide justification for them. On the other hand, Skyrms and, understandably, O'Conner, instead consider it to have a purely explanatory value, which can be of use to ethicists and political philosophers only in delineating the domain of plausible demands on human behaviour. In this view, the evolutionary story does not give any justification of the social norms, but only shows which norms are likely to persist under certain conditions, and which not. However, the debate remains very much open. Another open path for research is the design of more complex models, taking into account not only different social roles, but also the varying speed of learning between individuals or the strength of the starting bargaining position and other important parameters. The results obtained by such models will resemble more closely our real world, and will thus be of greater help in understanding our real-world social norms.

Martina Valković is a Research Assistant at Leibniz University Hannover and a Visiting Researcher at Radboud University Nijmegen. martina.valkovicphilos.uni-hannover.de; mvalkovic@gmx.com

References

Alexander, J. M. (2009) 'Evolutionary Game Theory', *The Stanford Encyclopaedia of Philosophy*.

https://doi.org/10.1017/S1477175622000045 Published online by Cambridge University Press

Binmore, K. (2006) *The Origins of Fair Play*, Papers on Economics and Evolution, 614 (Jena: Max Planck Institute of Economics, Evolutionary Economics Group).

O'Connor, C. (forthcoming) *Dynamics of Inequity: How Categories like Gender and Race Impact Cultural Evolution, and What it Means for Fairness* (Oxford: Oxford University Press).

Ross, D. (2016) 'Game Theory', *The Stanford Encyclopaedia of Philosophy*.

Skyrms, B. (1996) *Evolution of the Social Contract* (Cambridge: Cambridge University Press).

Skyrms, B. (2003) *The Stag Hunt and the Evolution of Social Structure* (Cambridge: Cambridge University Press).

Sugden, R. (2005) *The Economics of Rights, Cooperation and Welfare* (Basingstoke: Palgrave Macmillan).

Verbeek, B. and Morris, C. (2018) 'Game Theory and Ethics', *The Stanford Encyclopaedia of Philosophy*.

https://doi.org/10.1017/S1477175622000045 Published online by Cambridge University Press