Toward a Robust Game Theoretic Explanation of the Cultural Evolution of Justice

By Hannah Stowe

In the preface to his Evolution of the Social Contract, Brian Skyrms describes two traditions which have sought to explain the social contract in different ways. The first and, according to Skyrms, the "best-known" field of thought in this area "approaches the social contract in terms of rational decision. It asks what sort of contract rational decision makers would agree to in a preexisting 'state of nature'" (ix). The second tradition, Skyrms writes, in which he himself participates, "asks different questions": "How can the existing implicit social contract have evolved? How may it continue to evolve?" (ix). In an attempt to answer these latter questions, Skyrms develops an explanation of the idea of justice using game theory. In response, Justin D'Arms, Robert Batterman and Krzyzstof Górny critique crucial aspects of Skyrms's account. In order to achieve his results, Skyrms incorporates correlation into his model. D'Arms et al challenge Skyrms's use of correlation and show how a revised, more robust version of his model does not explain justice; on the contrary, it suggests that justice would not have evolved in human culture.

I will respond to both Skyrms's account and the account by D'Arms et al, identifying another way in which both these accounts fail to be robust. My more robust account indicates how game theory can be useful in explaining the cultural development of justice after all; like Skyrms, this project defines justice as the equal sharing of resources between players in the Nash bargaining game (5).[1] While the conclusions of D'Arms et al are convincing and succeed in taking Skyrms's account to task, ultimately their critique is not threatening to a game theoretic account of justice. D'Arms et al illustrate that Skyrms's account is deficient in robustness, but I will argue that both models (that of Skyrms and that of D'Arms et al) lack robustness in a different way. These authors only model interactions between two individuals. However, it is necessary to take into account interactions—"games," to use game theory terminology—involving more than two players. Once one does this, the cultural development of the just strategy seems more likely.

In order to understand what D'Arms et al are critiquing, we must first examine Skyrms's own argument. Skyrms's account of justice, according to D'Arms et al, begins with an examination of the Nash bargaining game. This game involves two players, one referee, and one resource which the players must share. Nash uses cake as an example of such a resource. For the game, each player writes down the portion of the cake ze[2] wishes to receive on a piece of paper and gives the paper to the referee. Neither player knows what the other is writing, but both know the following rule: If their "bids" add up to the whole cake or less than the whole cake, each receives the portion ze wrote down; if their bids add up to more than the whole cake, neither player receives any cake (D'Arms et al 1998, 77). So, if each player bid one-half, each would receive one-half. If one bid one-fourth and the other bid three-fourths, the players would receive one-fourth and three-fourths of the cake, respectively. If both were to bid three-fourths, neither player would get anything. Experimentation reveals that, in a game like this one, most people bid one-half. "Demand 1/2" appears to be the most popular, "intuitive" strategy (78).

Skyrms seeks to explain why the strategy "demand 1/2," which he takes to be the just strategy, is so prevalent. While empirical evidence reveals that "demand 1/2" is the most popular strategy, Skyrms suggests that the reason for the popularity of this strategy is not immediately clear (78). Why, for example, do we not see some players bidding for one-fourth of the cake, or three-fourths, etc.? In an attempt to answer this question, Skyrms constructs the following thought experiment: Begin with a group of individuals over which the strategies "demand 1/2," "demand 1/3," and "demand 2/3" are evenly spread. These individuals play a round of the bargaining game described above in random groups of two. Successful players—those who end up with cake or whatever resource is at stake—survive and reproduce. More importantly, they pass their strategies on to their offspring. In the next "round" of the game, which represents the next generation of a species, the three strategies will not be evenly spread as before (78). Since certain strategies are more successful than others and a player's success results in more offspring for that player, and since, again, individuals carry on their parents' strategies, a strategy's success in the first round will result in more players using that strategy in the second round (78). As D'Arms writes, "the number of individuals playing each strategy is adjusted after every round to reflect the differential success of the various strategies" (D'Arms 1996, 616).

According to Skyrms, if any one strategy wins out eventually, it will be "demand 1/2" (D'Arms et al 1998, 78). Of the three strategies, this is the only one that is "evolutionarily stable," meaning that "if almost everyone in a population is playing it, any mutant strategy will do worse, and hence cannot invade the population through a process of natural selection" (78). There could not be an entire population made up of "demand 2/3" strategists, because any time two "demand 2/3" strategists play each other, they both lose. Remember, in order for the players to receive cake, their bids cannot add up to more than one whole, and

$$2/3 + 2/3 \approx 1.333;\ 1.333 > 1.$$

Therefore, "demand 2/3" is not a stable strategy in a population where everyone has this strategy. Players who demand one-third, on the other hand, always receive cake when they play each other:

$$1/3 + 1/3 \approx 0.667;\ 0.667 < 1.$$

Therefore, we can imagine a population in which everyone was of the "demand 1/3" strategy. However, this strategy is only stable until a "mutant" is born who realizes that ze can get more cake in these bargaining games where everyone else is demanding one-third (ze can demand and receive up to 2/3 of the cake). So, even if a population develops consisting of all "demand 1/3" strategists, this strategy will not last long.

Still, Skyrms continues, it is not clear that only one strategy should survive over the long term in a population (79). Why, for instance, should a population not develop consisting of both 2/3ers and 1/3ers? When 2/3ers play with 1/3ers, everyone wins:

$$2/3 + 1/3 = 1.$$

Such a state is termed a "polymorphic equilibrium," and it appears to be just as stable as the "pure" equilibrium achieved in a population consisting of all 1/2ers (79, 78).

However, argues Skyrms, the polymorphic equilibrium is not as likely to evolve as the pure equilibrium (82). This is due to "correlation," the tendency of like individuals to attract each other (81). According to the idea of correlation, 1/2ers will be more likely to play with other 1/2ers, 1/3ers with other 1/3ers, and 2/3ers with other 2/3ers. As we have seen, when 2/3ers play with 2/3ers, no one wins, and when 1/3ers play with 1/3ers, everyone wins only until someone with a different strategy shows up. Therefore, concludes Skyrms, the pure equilibrium in which everyone demands one-half is, when all is said and done, still the most likely to occur.

D'Arms et al challenge Skyrms's conclusion by disputing the robustness of his account. Robustness refers to whether or not "[t]he desired result is achieved across a variety of different starting conditions and/or parameters" (90). First, D'Arms et al remark that the results Skyrms achieves when he brings correlation into the picture seem to contribute to the robustness of his account (82). Then, however, they go on to raise problems with Skyrms's use of correlation, thereby challenging that apparent robustness (92). They argue that, while it makes sense to allow for correlation between "demand 1/2" individuals, it does not make sense to expect that "demand 2/3" strategists will only play with each other, or that "demand 1/3" strategists will seek each other out (94). One of Skyrms's assumptions, according

[1]Thus, in one sense, "justice" is a technical term in this paper. However, I admit that my use of the word at times also appeals to a perceived cultural consensus on the meaning of justice. While this appeal entails a problematic assumption, it is somewhat warranted by my own (and Skyrms's) deeper assumption that humans have a long, shared cultural (pre/) history, such that introspection is arguably a useful methodological tool. Still, I will not deny that I am on shaky ground when I speak of "our sense of justice"; at this point, I hope to provoke critical discussion of the justifiability of these assumptions, rather than to defend myself entirely against such critique.

[2]In this paper, I use the gender-neutral, third-person singular pronouns "ze," "hir," and "hir" in place of the English subject pronouns "she/he," object pronouns "her/him," and possessive adjective pronouns "her/his," respectively.

to D'Arms et al, is that the individuals in these hypothetical stages of the development of the species are freely, rationally choosing their strategies (93). Given this premise, they argue, 2/3ers will not make a point to play with each other; on the contrary, they will avoid each other (95). Therefore, rather than correlation, D'Arms et al include anticorrelation between 2/3ers in their model (95). Moreover, they argue, there should be no correlation between 1/3ers, because it does not matter whom 1/3ers play: they always receive a payoff. "Demand 1/3" strategists will not go to the trouble to seek out only each other (96). Finally, note D'Arms et al, it should be taken into account that 1/2ers and 2/3ers do have to go to some trouble in determining their partners in a bargaining interaction, while 1/3ers do not. "Being choosy," they argue, "…could leave an individual without a partner in a given round" (96). Therefore, they associate a "cost factor" with correlation and anticorrelation in their model to account for possible cases where 1/2ers and 2/3ers do not have the opportunity to play (96). When D'Arms et al take these different kinds of (anti)correlation into account and include the cost factor, their model illustrates that a polymorphic equilibrium is much more likely to develop than a pure equilibrium in which everyone demands one-half (97). D'Arms et al have illustrated, therefore, that Skyrms's model does not, in fact, provide a robust explanation of the cultural development of human justice.

Nevertheless, again, I will show that a game theoretic explanation of the evolution of justice is possible. Both Skyrms's account and the account by D'Arms et al contain a basic flaw in their starting points. Both accounts use the model of a two-player game in order to sketch the development of bargaining strategies in a population. I submit, however, that the kind of "games" we are imagining in the (cultural) evolutionary history of the species would have involved more than two players. A more robust explanation of justice will model interactions among more than two individuals, for instance among three players and among four players. As I will show, in bargaining games involving three players from a population in which it is present, the just strategy emerges as the dominant strategy.

Importantly, in bargaining games which include three or four players, the strategy "demand 1/2" no longer seems just. It is not generally considered just for one individual to claim half the cake when there are four people present. Instead, just division of the cake should involve the players receiving equal portions to each other. Therefore, I state the just strategy as "demand 1/n," where n= (the number of players in the game). An individual with the strategy "demand 1/n" demands the portion which each player would receive if each were to receive an equal portion to the rest. So, in a game with two players (n=2), the "demand 1/n" strategist demands one-half. In a three-player game (n=3), ze demands one-third. In a four-player game (n=4), ze demands one-fourth, and so on.

In bargaining games involving three players (n=3) from a population in which the strategies "demand 1/n," "demand 1/2," and "demand 2/3" are evenly spread, "demand 1/n" strategists receive the only payoff. In a three-player game, the "demand 1/n" strategist will demand one-third. Out of all the possible games that could take place, the only one in which anyone wins anything is the one in which three players of the strategy, "demand 1/n," play each other.[3] Of course, one could argue that this is only the case because the other strategies I chose to represent are too large in their demands. What happens if we include more so-called "modest" strategies in our population?

In bargaining games involving three players (n=3) from a population in which the strategies "demand 1/n" (just), "demand 1/6" (modest), and "demand 2/3" (greedy) are evenly spread, each strategy results in payoff some of the time. The strategy "demand 1/6" results in payoff the most often, "demand 1/n" results in payoff nearly as often, and "demand 2/3" rarely results in payoff.[4] These results seem to suggest that, in a population containing a very modest strategy, the just strategy, and a greedy strategy, a polymorphic equilibrium will emerge. Instead of a single strategy like "demand 1/n" taking over the population entirely, it appears that each strategy will survive because each results in payoff some of the time.

The polymorphic equilibrium is not, however, likely to develop in a population with interactions involving several players. In theory, it would, but in reality it probably would not, because very modest strategies are not adaptive. It is important to keep in mind that we are modeling bargaining situations in early human cultural development. More often than not, these situations would likely have involved competition over a necessary resource, rather than a luxury like cake. Presumably, then, the likelihood of an individual's survival and reproduction would increase or decrease depending upon the amount of the resource ze received in such situations. A modest strategist would receive little of the resource, while a greedy and even a just strategist would receive more of it. If the resource was important to the individual's health, the just and greedy strategists might receive more benefit to their health from their payoffs in the bargain than the modest strategist. The just and greedy strategists would then be more likely to continue to survive and produce offspring than the modest strategist. The just and greedy strategies would then be more prevalent in the following generation of the population, if we assume with Skyrms that parents pass their strategies on to their offspring. The very modest strategies, although they are successful in individual bargaining games with more than two individuals, would nevertheless probably disappear in time because they would not be as beneficial to an individual over a lifetime as the just and greedy strategies.

Moreover, as modest strategists grew scarce, greedy strategists would as well. The greedy strategist's success in a bargaining game entirely depends on the presence of a modest strategist in the interaction. In interactions involving only just and greedy strategists or only greedy strategists, no one ever gets anything. In bargaining interactions that do not involve modest strategists, the portion which the just strategy demands is the maximum amount any player can receive; any higher demand will cause the total demands of the group to exceed the whole of the resource at stake. Therefore, greedy strategies will not survive long without modest strategies. The just strategy, on the other hand, depends on neither modest nor greedy strategies for its perpetuation: just strategists receive payoff playing each other. Practically speaking, then, out of a beginning population in which modest, just, and greedy strategies are evenly spread, a polymorphic equilibrium is not as likely to evolve as a pure equilibrium consisting of all just strategists.

This conclusion might not sound any different than the conclusion at which Skyrms arrived in the first place. Importantly, however, this paper's account of justice is not subject to the same critique which D'Arms et al give Skyrms's explanation. As stated, the models of Skyrms and D'Arms et al both fail to take into account bargaining games that involve more than two players. It is, however, important to account for such games, because humans likely worked together in groups larger than two in early cultural development (as they do now). Again, the results of three-player bargaining games suggest that the just strategy, "demand 1/n" (n=number of players in game), is in fact most likely to prevail in a population's cultural evolution.

While these conclusions provide for a more robust game theoretic explanation of human justice than Skyrms's account, at least three questions remain regarding the representativeness and the robustness of both game theoretic models in general and this paper's model in particular. Two of these have to do with the relevance of merit to justice. The third question regards the application of the strategy "demand 1/n" (n=number of players in game) to situations involving a great number of individuals and only a small resource. I will address each of these in turn.

The first objection questions whether the strategy "demand 1/2" in the two-player Nash bargaining game, or "demand 1/n" in the n-player game, is in fact a just strategy. The suggestion here is that justice is not in fact blind, but takes into account things like merit. For example, take the case where four individuals, players A, B, C, and D, go fishing and player A does nothing but sunbathe the entire time while the others catch fish. In this instance, it does not seem to be just that player A take home the same amount of fish as the other three. According to our sense of justice (if a common "sense of justice" between myself and the reader can be presumed), it would not be just

for hir to relax all day and then demand one-fourth of the day's catch. This objection challenges the "representativeness" of my account, which essentially refers to whether my theoretical models correspond to everyday reality (D'Arms et al 89). The question is whether I (and Skyrms) have captured the human idea of justice when I state the just strategy as the strategy which demands the amount that everyone would get if everyone were to get the same amount.

While this is a weighty objection, I do not think it is insurmountable. It certainly seems true that player A ought not to receive hir demand of one-fourth. In fact, it does not seem obvious that ze should receive anything. I suggest that, given player A's lack of participation in the fishing endeavor, ze might not be included in the other players' computation of just division of the fish at all. It seems reasonable instead that players B, C, and D divide the fish equally only among themselves. The strategy "demand 1/n" would then translate to "demand 1/3," not "demand 1/4."

I am suggesting that "demand 1/n" strategists would have to have in place a system for determining the value of "n." This system would allow individuals to distinguish between those individuals who participated fully in the pursuit of whatever resource was at stake, and those who did not. Players would mentally weed out non-participants and calculate "n" based on the number of deserving participants that remained. Players would account for merit, then, but among the number (n) of those who had merit, justice would still dictate that each receive "1/n."

Still, there is a second possible objection having to do with merit which my account does not yet seem able to address. I have explained how players might discriminate based on merit by calculating "n" based only on those who helped acquire the resource. Even among those who earn a share in the resource at stake, however, humans sometimes deem it just to divide that resource unequally based on other determinants of what the players deserve or need. For example, say two strangers come across a hundred-dollar bill on a sidewalk at the same time. One of these individuals is a successful businesswoman, and the other is a struggling waitress saving to go to college. They do everything possible to try to return the cash to its owner, but ze cannot be found. In the process, however, the two women talk and become familiar with each other's situations. At this point, we would expect the wealthy financier to allow the other woman to keep all or most of the cash. In fact, we would probably even think the former individual was greedy and unjust if she demanded half the find. Our sense of justice (again, presuming that one can speak of such a thing) dictates that the money go to the woman in need (assuming the original owner is not an option), despite the fact that they both spotted it at the same time. In this case, it does not seem quite right that the women split the money evenly. "Demand 1/n" does not, therefore, appear to be the just strategy in every instance, namely, in those instances where the resource at stake is gratuitous to one individual and necessary to another. This means that my account is not as representative as it could be.

The third objection I mentioned points to cases where a great number of individuals must share one resource. In such cases, the argument says, we do not feel that it is just to receive a portion of only "1/n." For example, imagine a poorly planned wedding reception where there is only one bottle of champagne to be distributed among three-hundred guests. Here, it would not seem just (or practical) to divide the champagne equally among the guests, giving each only a few drops. There seems to be some point, then, at which the ratio of players to resource is simply too large for the strategy "demand 1/n" to apply. In such cases, it seems justice would have to be determined in some other way. Therefore, "demand 1/n" does not seem like a reliably just strategy, according to this objection. This objection questions the robustness of my account, because I have not modeled bargaining games involving more than three players.

In situations where there are not enough resources to go around and where no individual has greater claim to the champagne than any other, however, I submit that humans do not have an idea of what just division should look like. In these cases, justice simply does not seem to apply, and therefore it is incorrect to speak of any strategy, including "demand 1/n," as unjust. Humans' behavior in such instances seems to suggest that they have not developed fine-tuned strategies for dealing with such situations.

In a situation like the wedding reception I just described, individuals would often decide simply to reserve any claim to the champagne. Rather than attempting to receive a portion of champagne that was equitable or otherwise, it seems likely that some guests would opt out of bargaining altogether. To some, if not many, it would simply not appear worthwhile to exert time and effort with such little assurance of payoff. This behavior suggests that humans have not developed culturally widespread strategies for receiving payoff when resources are scarce and competitors are many.

Of course, there are individuals (myself, for example) who would choose to go for a portion of the champagne even if they knew there was not enough for everyone. It is unlikely, however, that these champagne-loving individuals would describe their action as just. In these cases, like the cases where the guests opted out, it seems that the individuals are not employing learned bargaining strategies at all. On the contrary, they might be going against the rules for social interaction that their parents (or other authority figures) have taught them. In the end, then, human behavior in cases where there is not enough to go around does not challenge the idea that the strategy "demand 1/n" (n=number of individuals in interaction) is just. Rather, it illustrates that cultural strategies for dealing with such situations have not (yet) evolved in populations with which I am familiar.

This paper has laid the groundwork for a game theoretic explanation of justice that is more robust than those by Skyrms and D'Arms et al. Such an explanation will include bargaining games that involve more than two individuals, where the just strategy can be stated as "demand 1/n" (n=number of players in game). As seen, question regarding the representativeness of game theoretic accounts of human justice arises upon consideration of cases where one individual's need outweighs another's. Such instances appear to refute the idea that the strategy "demand 1/n" (n=number of players in game) is in fact just. Future inquiry will need to consider this complication and how it impacts the representativeness of game theoretic accounts of human justice that use the Nash bargaining game model.

Bibliography

D'Arms, Justin. "Sex, Fairness, and the Theory of Games." The Journal of Philosophy 93.12 (1996): 615-627. Philosopher's Index. EBSCO. Web. 16 Feb. 2011.

D'Arms, Justin, Robert Batterman and Krzyzstof Górny. "Game Theoretic Explanations and the Evolution of Justice." Philosophy of Science 65.1 (1998): 76-102. Philosopher's Index. EBSCO. Web. 16 Feb. 2011.

Skyrms, Brian. Evolution of the Social Contract. New York: Cambridge University Press, 1996.

[3]For the mathematical calculations involved in my making this conclusion, please see the online version of this article.

[4]The calculations behind these statements, as well, can be found in the online version of this article.