Philosophy and Phenomenological Research
Vol. LIX, No. 1, March 1999

Games Social Animals Play: Commentary on Brian Skyrms's *Evolution of the Social Contract*

PHILIP KITCHER
University of California at San Diego

Brian Skyrms discusses a variety of game-theoretic topics, including bargaining games, ultimatum games and commitment, prisoner's dilemma and correlated equilibria, symmetry breaking, and signal systems. On each topic, he has much to say that is important, novel, and insightful. For this reason, *Evolution of the Social Contract* deserves careful study. In what follows, I intend not so much to offer correctives to what he has written as to explore ways in which his book might be extended. More specifically, I shall start from the connection between the various game-theoretic situations he analyzes and the problem posed at the very beginning of his book, bearing in mind an important qualification that Skyrms offers on the next-to-last page. There he writes:

> In bargaining situations between more than two people, coalitions may play a crucial role. If I had, or knew of, a good account of the dynamics of coalition formation, I would have written a longer book. (108)

It will be useful, I think, to see why a longer book might be even more valuable, and why it would be difficult to write.

Skyrms starts by distinguishing two ways of thinking about social arrangements: one might ask how a particular form of group behavior could be seen as the product of rational decisions on the part of the group members, or one might try to understand how that type of group behavior might evolve. Conceiving the latter enterprise in Darwinian terms, he approaches it in terms of differential replication on the part of appropriate entities. These appropriate entities might be genes, associated with phenotypes that directed the pertinent pieces of behavior (in suitable environments), but this need not be so. An alternative choice, compatible with the same dynamics, is to assume that they are cultural entities that can be invented, transmitted and

copied. Skyrms's discussion is to be neutral about the targets of the processes he considers.[1]

The style of argument is evident from the opening discussion of bargaining, where Skyrms shows that most populations with a variety of demand strategies will evolve to states in which everyone asks for 50%. Given sufficient granularity in the divisibility of the good, or a relatively small amount of correlation among the players, populations can avoid polymorphic traps and reach the "fair" equilibrium. So, Skyrms concludes, we may have "a beginning of an explanation of the origin of our concept of justice" (21).

A beginning, perhaps—but we ought to look closely at just what is being explained and just what is assumed. If we are serious about accounting for the origin of a human conception of justice, then we have to suppose that the dynamics of the populations corresponds to a process that occurred in human prehistory. Let's make the connection in the most obvious way. We'd explain the origins of the concept of justice by supposing that our remote ancestors belonged to populations in which situations isomorphic to the bargaining game repeatedly arose. Successive generations followed one of the trajectories in Skyrms's simulations, reaching a situation in which all of us have the propensity to demand 50%.

There are two obvious forms of concern about this way of making the link, one that focuses on the explanandum and one that worries about the realism of the assumptions. We can bring out the first by noting the difference between explaining a propensity to behave in a particular way in a particular situation and explaining our having a special attitude to others' behavior in that situation. At best, it seems that the simulations of the bargaining game reveal why we might have evolved to conform to arrangements to which we attach the label 'just', but it would appear that accounting for the origin of our concept of justice would require showing why we distinguish cases in which others do not do their part in such arrangements from other instances in which they don't do what we would like them to have done. Suppose a greedy mutant arises in a population of fair bargainers. Skyrms's simulations show that this mutant is going to be driven out of the population, but there's so far no explanation for any special attitude on the part of other population members—resentment at the attempt to take more than one's fair share, say—no reason for thinking that this mutant should be differentiated from any other member of the population with whom interests conflict.

I don't want to underestimate what Skyrms has achieved. Although the old idea that Darwin's world is inevitably harsh has faded in light of recent work,

---

[1] In fact, there is one interpretation of the evolutionary process that might not conform to the mathematical models he offers. If there is differential replication both of genes and of cultural items, then it isn't obvious that Skyrms' dynamics will be applicable.

it's important to demonstrate that the forms of behavior that accord with our sense of justice and morality can originate and be maintained under natural selection. Yet we should also be aware that the demonstration doesn't necessarily account for the superstructure of concepts and principles in terms of which we appraise those forms of behavior. Throughout the book, it seems to me, what we get is the former and not the latter, although occasional sentences (such as the one quoted from p. 21) seem to promise more.

The second worry is more serious. Consider the circumstances of the bargaining game. We are to suppose that there is a divisible good to which two parties, *A* and *B*, can make claims, and that there is a third individual, *C*, who referees, obtaining the entire resource if their demands total more than 100% and the residue if their total demands fall short of 100%. How representative is this of situations that might have figured in our evolutionary past?

Now there were surely many occasions on the savannah or in the forest on which organisms ancestral to us came across divisible goods—not cakes, but fruit trees, areas of relatively safe territory, care from conspecifics, and so forth. Surely in most instances, though, matters were not symmetrical, and one of the parties had the strength, experience, or speed to take what he/she wanted with relative impunity. Furthermore, *C*'s supervisory role is either unrealistic or, at least, in need of prior evolutionary explanation. Why should *A* and *B* defer, foregoing everything if it turns out that their total demand exceeds 100%? Why does *C* even bother to intervene, for, as writers on the evolution of punishment have pointed out, there are evolutionary costs to policing the behavior of others? Why, if *C* benefits from whatever part of the good comes his/her way and if *C* has enough power to enforce the terms of the arrangement, doesn't *C* simply take everything? And what are the other members of the group doing while the bargaining is occurring? Are they watching what kinds of offers the parties make? Or are they intervening, buttressing the claims of one of the parties?

Bargaining did emerge at some stage in human, hominid, or pre-hominid, social life. However, I submit that its emergence was preceded by the evolution of other social institutions, and that the structure of the bargaining game can't be specified until we have a clear view of these other institutions. One precondition for the kinds of situations Skyrms treats is the abandonment of force to resolve various kinds of conflicts, particularly those in which there are asymmetries among the participants. Now if the abandonment of force depends on the attitudes and actions of other members of the group, and if the attitudes and actions of those members are modified in light of an animal's current behavior, then the payoffs even in an apparently simple situation will be affected. As I'll try to show below, probably the most important determinants of the reproductive success of weaker group members (juveniles, for example) stem from the ways in which their current actions affect the treatment they will receive from the more powerful organisms around them. If

this is correct, then the simple specifications of utility in the kinds of games that Skyrms considers miss the major terms.[2]

There are two ways of looking at the evolution of social interactions. One is to specify a simple game that seems to bear some resemblance to situations recurring in the lives of our presocial ancestors; the other is to try to identify the structures of the interactions that seem most fundamental to the social lives of our closest evolutionary relatives. My chief reservation about Skyrms's book is based on pessimism about the first approach. In the rest of this essay, I'll try to expose the difference between the two and explain why the second is so hard.

* * *

Styles of primate social life are highly variable. Even among our relatively close evolutionary relatives, there are species that do not form groups with more than one adult male (orang-utans) and species in which the social units are male-female pairs and their offspring (gibbons). The pre-historic record indicates that our ancestors lived in larger social groups, comparable to those of present-day chimpanzees and bonobos. Our most basic question about the evolution of the social contract ought to be "How did the evolution of social groups containing several unrelated adults of both sexes originate under natural selection, and how has it been maintained?"

Following suggestions of Richard Wrangham,[3] I propose that differences in the forms of social life among the various groups of higher primates are best explained by features of the distribution of resources: in some instances, adults or pairs of adults can commandeer a territory containing abundant resources and their encounters with others are relatively infrequent and agonistic. The precondition for sociality is that animals are forced into a common space to obtain the things they need, and, when this occurs, there are two possibilities—either the strongest win and the weaker die off (or fail to reproduce), or the animals arrive at a scheme of mutual toleration. So the fundamental structure of the games social animals play involves a competition for

---

2 I have been equally guilty of the oversimplification identified here. In "The Evolution of Human Altruism" (*Journal of Philosophy*, 90, 1993, 497–516), I offered a game-theoretic picture that was mathematically tractable, but one that didn't take into account the complexities of the interactions among organisms whom we could reasonably take as ancestors of ourselves, and how those complexities might affect the payoffs. As I note in "Psychological Altruism, Evolutionary Origins, and Moral Rules" (forthcoming in *Philosophical Studies*), I continue to regard the approach pursued by Skyrms and my earlier self as profitable, if set in the right context.

3 See R. Wrangham "On the Evolution of Ape Social Systems", *Social Science Information*, 18, 1979, pp. 334–68; "An Ecological Model of Female-Bonded Primate Groups", *Behaviour*, 75, 1980, pp. 262–300; "Social Relationships in Comparative Perspective", in R. Hinde (ed.) *Primate Social Relationships: An Integrated Approach* (Oxford: Blackwell, 1983), and "Evolution of Social Structure" in Barbara Smuts *et. al.* (eds.) *Primate Societies* (Chicago: University of Chicago Press), pp. 282–96.

scarce resources with asymmetries in power and ability, and the task is to understand when and how the animals transcend the brutal *bellum omnium contra omnes*.

Studies of the social behavior of social primates provide an important clue. The lives of chimpanzees, bonobos, and some baboon species are dominated by coalitions and alliances, often maintained or repaired by prolonged bouts of social grooming. Animals' chances of securing food, receiving care and protection, exercising mate choice, and other important determinants of fitness depend on which other members of the troop will come to their aid. At the earliest stages of an animal's life, support is typically provided by the mother and other relatives (although, even here, the well-being of the infant depends on the mother's alliances), but a recently-weaned juvenile needs friends to thrive at all. I conjecture that the strategy young chimpanzees and bonobos use is a relic of the way in which our remote common ancestors made the transition to sociality.

Imagine a population of N organisms in a region with r resources, each of value v. Suppose that each organism can visit k resources in the period before those resources are renewed. Let an animal's fitness be directly proportional to the number of resources collected. If the world is benign (Rousseauian), then there's enough to spare for all, and $Nk < r$. However, it's unlikely that this condition will prevail for long, and the principal interest lies in what occurs when it's violated. Order the animals in terms of decreasing strength, and let their strengths be $s_1 \ldots s_N$. To keep life simple, let's suppose that if two animals, with strengths s_i and s_j, meet in the vicinity of a resource, then the stronger gains the resource and the weaker receives nothing—thus if $i<j$, the ith organism takes the spoils. Abstracting from luck (a weak animal may chance on a resource and not encounter anyone stronger), the distribution of resources, and thus of fitness, can be expected to approximate an assignment of maximum possible fitness (proportional to kv) for the animals early in the ordering, and minimal fitness to the organisms lower down.

What can weak organisms do in this predicament? Just what social primates do—form alliances. Suppose that the strength of a coalition is the sum of the strengths of the coalition members, and that a coalition beats an individual if the strength of the coalition is greater than the strength of the individual. Let A and B be two animals with strengths s_A and s_B respectively ($s_A > s_B$), such that neither A nor B can gain anything in individual competition. If there is an individual C who is garnering some resources, such that $s_A + s_B > s_C$, then A and B working together can obtain the resources that C would have controlled, for a maximum of k to divide between them. It will be necessary for A to concede some part of this booty to B if there is to be some fitness advantage to B in joining the coalition. However, it is already possible to recognize that there may be a complex dynamical process if partners are to make fitness-maximizing choices. (Assume that you are A, cur-

rently a have-not but one of the stronger of the have-nots; you may have a choice of potential partners; if you choose stronger partners, you will be able to beat stronger individuals, and maybe stronger coalitions that may arise in the future; but stronger partners have more choices among their partners, and can therefore demand a larger share of the spoils.)

It isn't hard to see that, under almost every distribution of strengths in a non-Rousseauian world, there'll be some pair of animals who can profitably team up. Once they've done so, it'll be almost inevitable that the individual they displace can form a coalition with another have-not to regain some resources. So there will probably be a process of escalation, generating further coalitions, including some that are larger than two-organism alliances. Of course, once there are three-member coalitions, coalitional substructure comes into play in the determination of the division of the resources acquired by the coalition. Notice, however, that a coalition works by visiting resources together, so that, if there are m coalitions, they only visit km resources, and, although $Nk > r$, it may be that $km < r$. The original habitat may thus become subdivided into large coalitions that act together against neighboring coalitions, that effectively control a number of resources, and that resolve the distribution of those resources through subcoalitional alliances. This, of course, is the abstract structure observed among some social primates, notably chimpanzees and bonobos.

I have sketched an N-person game and some qualitative features of its solution. It would be satisfying to be able to go further and to be more precise about the numbers and sizes of the coalitions, but I doubt that this is possible. Unlike the elegant games studied by Skyrms, this one looks mathematically messy. Yet, even without knowing much about the kinds of coalitions that form, we can still recognize ways in which coalitional structure affects the interactions among primitive social animals.

Strength is one important feature of a potential coalition partner, but it is by no means the only asset. Conspecifics are valuable as allies if they are strong, reliable, and inclined to make relatively small demands. Our prehistoric ancestors surely had the cognitive capacities to make rough judgments about the merits of those around them on all three of these dimensions, and our intelligence may even have been shaped by the importance of such judgments in determining fitness (this is a non-standard version of the popular hypothesis that we became smart because of social demands on our powers of detection). Again, let's simplify, supposing that organisms in the same habitat keep track of one another's activities by means of a composite index that records the *standing* of each for each. I'll offer two general hypotheses:

(1) X's standing for Y is affected by Y's perception of X's strength, X's past performance in interactions that bear on Y (or, perhaps more

generally, on Y's kin or on animals that have high standing with Y), and X's willingness to make few demands.

(2) The probability of X's being accepted as a coalition partner by Y depends on X's standing with Y.

I think that (1) and (2) offer partial explanations of broad features of primate social behavior, particularly of the efforts that animals make in maintaining and repairing their alliances, and that one can also provide Darwinian reasons for thinking that successful social animals would conform to these hypotheses.

Let's now return to the kinds of games that Skyrms studies—bargaining, ultimatum, and prisoner's dilemma in particular. If we take seriously the idea that situations like these occurred repeatedly in our evolutionary past, then we have to recognize that they would have taken place in a public setting. Not only the participants, but also onlookers, would have been able to witness the actions performed, and to use those actions in modifying their attributions of standing. Hence we shouldn't think about the payoffs in the short-term way that Skyrms (following game-theoretic tradition) does. Each cell of the matrix must be considered in light of the consequences for future interaction opportunities and the expected behavior of other group members.

Consider, for example, the ultimatum game. In accordance with (1) and (2), we might believe that perceived strength increases the probability that one will play the role of the agent who issues the ultimatum and that perceived willingness to accept unfavorable offers increases the chance that one will participate in situations where one is on the receiving end of an ultimatum. For imagine that the ultimatum situations occur after two animals have teamed up to gain a resource that neither could have acquired on its own, that the animal perceived to be stronger gets to issue the ultimatum, and that one can choose partners for the joint effort. Animals that perceive themselves to be strong will then look out for weaker partners who are relatively undemanding. If you are weaker than most, it may be a good strategy to accept quite unfavorable demands, since, although you gain little on each occasion, you are chosen on more occasions, and a larger number of small returns may be the best you can manage.

I haven't analyzed the dynamics of this amended ultimatum game in any detail, and, in any event, it's important to see that performance in one ultimatum game not only affects future possibilities for playing games of this sort but other types of games as well. But the general story is prefigured in my earlier analysis of situations like the prisoner's dilemma: as I argued, when animals have the chance to choose their partners and to opt out of any interaction, conduct now has consequences for future payoffs. That earlier analysis was too simple because I didn't consider all the factors that enter into the

determination of standing, nor did I appreciate the prevalence of coalitional structure in animal interaction. But it was a first step in a direction I now believe should be followed more thoroughly.

* * *

Where does this leave Skyrms's enterprise? There seem to me to be two possible ways of advancing, the *top-down* and the *bottom-up* approaches. The top-down strategy would build directly on Skyrms's achievements. We would take the mathematically tractable games that seem to capture various social interactions, amending them in light of hypotheses about the effects of coalitional structure. In effect, this procedure gives priority to modelling in ways that allow for precise analysis, trying to inject greater realism in a piecemeal fashion.

The bottom-up approach, by contrast, would try to tackle the difficult problems of coalitional structure head-on, attempting to represent the multiperson interactions that go on in primate societies, and to explore the kinds of social arrangements that might be expected as strategies for participating in those interactions are shaped by natural selection. Here, we would give priority to a realistic understanding of the games social animals play, but, quite possibly, forego the possibility of precise solutions.

It's hard to say in advance which (if either) of these strategies will work best in fathoming the evolution of the social contract. My current thoughts are, perhaps, the product of my own struggles with kindred problems: having reached what I thought was an elegant and satisfying treatment of the evolution of cooperation, I found confrontation with the rich literature on the details of primate social behavior sobering. Yet, even if only the bottom-up approach will succeed, Skyrms's subtle analyses should serve as paradigms of the standards to which we should aspire, and, if the top-down approach is workable, then future discussions are likely to be descendants of his. Either way, we cannot ignore attention to coalitions—and the result is likely to be a much longer book.