

# Explaining the Social Contract

## Zachary Ernst

---

### ABSTRACT

Brian Skyrms has argued that the evolution of the social contract may be explained using the tools of evolutionary game theory. I show in the first half of this paper that the evolutionary game-theoretic models are often highly sensitive to the specific processes that they are intended to simulate. This sensitivity represents an important robustness failure that complicates Skyrms's project. But I go on to make the positive proposal that we may none the less obtain robust results by simulating the population structures that existed among our evolutionary ancestors. It is by extending the evolutionary models in this way that we should pursue the project of explaining the evolution of the social contract.

- 1 *Introduction*
  - 2 *Skyrms's strategy*
  - 3 *Other models and model sensitivity*
    - 3.1 *Cultural evolution and other models*
    - 3.2 *Effects of simplifying assumptions*
  - 4 *Introducing group selection to evolutionary game theory*
    - 4.1 *The Island Model*
    - 4.2 *Constructing a group-selection explanation*
  - 5 *Conclusion*
- 

## 1 Introduction

In his book *Evolution of the Social Contract* ([1996]) and in an earlier pair of articles ([1994a], [1994b]), Brian Skyrms argues that evolutionary game theory is more promising than rational choice theory as a method for explaining how the 'existing implicit social contract' could have evolved ([1996], p. ix).

Evolutionary game theory uses mathematical models to determine which behaviors will tend to spread through a population. Here, the behaviors are identified with simple game-playing strategies. As game-players in a population repeatedly play some game with each other, each player will enjoy a payoff which depends on her strategy and also on the strategies of those players with whom she plays the game. In the evolutionary models,

differential reproduction<sup>1</sup> will cause the more successful strategies to increase in frequency and the less successful ones to decline.

Skyrms begins his analysis with a game called ‘Divide-the-Cake’. This game, which I will explain below, models a particular kind of bargaining situation. As a model of bargaining, some of the strategies in Divide-the-Cake seem intuitively ‘fair’ and others intuitively ‘unfair’. For Skyrms, we have ‘the beginning of an explanation’ (Skyrms [1996], p. 21) of the social contract if the evolutionary models can show that ‘fair’ game-playing strategies will tend to evolve in a population.<sup>2</sup> Skyrms is careful not to come to too strong a conclusion on the basis of such simple models. Rather than propose a full-blown theory of the evolution of the social contract, Skyrms is merely proposing and defending a particular methodology—evolutionary game theory—which he hopes can be used eventually to construct such a theory.

My purpose in this paper is to suggest a method for pursuing Skyrms’s project. Specifically, I propose that the evolutionary models be extended so as to take into account the population structures of our evolutionary ancestors. My argument in support of this suggestion proceeds in two parts. First, I show that without modeling such population structures, the evolutionary models are not robust enough to explain the evolution of fairness norms. I illustrate this robustness failure by showing that a standard evolutionary model of cultural evolution can behave very differently from the dynamics which Skyrms uses. I also briefly consider how the simplifying assumptions of Skyrms’s replicator dynamics model affect the behavior of the model.

In the second part of my argument, I show how the evolutionary models can be extended so as to take into account the population structures of our evolutionary ancestors. In particular, I consider how a simple metapopulation model of migration can be superimposed onto an evolutionary dynamics. When this is done, we can see that such a population structure significantly favors the evolution of fairness. Surprisingly, this result holds no matter what underlying evolutionary dynamic is used.

The purpose of the metapopulation model is *not* to suggest a particular explanation of the origin of fairness norms. Rather, its purpose is to illustrate

<sup>1</sup> Or some other mechanism. See below.

<sup>2</sup> In order for Skyrms’s explanatory strategy to be complete, we need to understand a ‘social contract’ merely as a tendency for members of a population to interact with each other in a ‘fair’ manner. If a ‘social contract’ is more than that, then we need to tell a detailed story about (say) how the evolution of ‘fair’ behaviors promotes the spread of fairness norms, or conventions which give rise to a social contract. Skyrms does not spell out the relationships between behaviors, norms, conventions, concepts of justice or fairness, and the social contract. In fact, Skyrms seems to change the explanandum of his analysis from ‘the evolution of the social contract’ ([1996], p. ix), to ‘the origin of our concept of justice’ (*Ibid.*, p. 21) to the evolution of ‘norms of fairness’ (*Ibid.*, p. 28). This point raises extremely difficult questions which I cannot hope to address in this paper. So for the remainder of the paper, I will simply put the project in terms of explaining the evolution of fairness.

how the standard evolutionary models can be extended to take various population structures and other features of group life into account. The behavior of the model strongly suggests that by taking such population structures into account, we can achieve results which are both robust and explanatory. I conclude that for Skyrms's project to be carried through, the evolutionary models must be informed by the empirical work of evolutionary biologists, anthropologists, and primatologists.

## 2 Skyrms's strategy

For an analysis of the emergence of fairness, we want to study a game which is simple enough for the mathematics to be tractable, yet complex enough to capture some simple pre-analytic intuition about fairness. The game which Skyrms focuses on is called 'Divide-the-Cake'. In it, we imagine that there is some surplus to be divided between two players—here, we imagine that the surplus is a cake. Without any communication between them, each player demands some portion of the cake from zero to one. If the sum of the two players' demands does not exceed one, then each gets exactly the portion she demanded. But if their demands add up to more than one, then both players receive nothing. There is an infinite number of possible demands from zero (demand none of the cake) to one (demand the entire cake). So to make the analysis simpler, we imagine that there are only three demands: Demand 1/3, Demand 1/2, and Demand 2/3.

In research conducted by experimental economists, test subjects are made to play the Divide-the-Cake game both repeatedly and in one-shot contests (Davis and Holt [1993], p. 243). As we might expect, test subjects tend to play the strategy Demand 1/2, which is the strategy we would intuitively judge as the 'fair' or 'just' one. When the test subjects are asked why they demanded half of the good, their answers suggest that the subjects' behavior is guided, not by self-interest, but by norms of fairness. The question we address is whether there is some feature of the 'fair' strategy that accounts for the evolution of fairness norms.

Skyrms is concerned with showing that evolutionary game theory can explain why norms for such 'fair' behavior have evolved. We are to consider a population which is initially seeded with all of the game-playing strategies and then we model the strategies as replicators. Each strategy will have a fitness which is its expected payoff in the population. Those with greater fitnesses will reproduce more than their less-fit counterparts, and will come to comprise a larger share of the population. Using an appropriate mathematical model, we can watch the spread of the successful strategies and determine under what conditions particular strategies will spread in any given population.

Since Skyrms proposes to model the strategies as replicators which reproduce according to their fitness, the evolutionary model seems to be most appropriate for modeling *biological* instead of *cultural* evolution. But because the social contract probably evolved through a process involving both cultural *and* biological evolution, he must argue that both evolutionary processes can be modeled in the same way.<sup>3</sup> Skyrms does provide a quick argument for this. In the introduction, he says:

These biological concepts also have qualitative analogues in the realm of cultural evolution. Mutation corresponds to spontaneous trial of new behaviors. Recombination of complex thoughts and strategies is a source of novelty in culture. Using these tools of evolutionary dynamics, we can now study aspects of the social contract from a new perspective ([1996], p. x).

Later, I will argue that this analogy is flawed. But for now, we should look at how the analogy to biological reproduction leads directly to a set of equations called the ‘replicator dynamics’.

Suppose we have a population of individuals, each of which is programmed to play some strategy  $i \in K$ , where  $K$  is the set of all the strategies. Then we denote the proportion of the population playing strategy  $i$  as  $x_i$  (so that if half the population is playing strategy  $i$ , then  $x_i = \frac{1}{2}$ ), and the expected payoff of strategy  $i$  in a population  $x$  as  $u(i, x)$ .<sup>4</sup> The average payoff to the individuals in the population is denoted  $u(x, x)$ . For the replicator dynamics, we want the growth rate of each strategy  $i$  to depend on  $i$ ’s expected payoff, compared to the average payoff in that population. If  $i$  does better than average, we want it to spread; but if it does worse than average, we want it to diminish. Since the growth rate of a population is equal to the fitness of the individuals multiplied by the number of individuals,<sup>5</sup> the growth rate of any strategy must be equal to the fitness of that strategy multiplied by the current population share of that strategy. So if we denote the rate of change of the population share playing strategy  $i$  as  $\dot{x}$ , we immediately get the replicator dynamics:

$$\dot{x} = [u(i, x) - u(x, x)]x_i \quad (1)$$

We should understand Equation 1 as a differential equation with a suppressed time argument. That is, the value  $\dot{x}$  refers to the growth rate of the population share playing strategy  $i$  at some time  $t$ . By solving<sup>6</sup> Equation 1 for

<sup>3</sup> See Robert Boyd and Peter Richerson ([1976]) for a discussion of the relationship between cultural and biological evolution, and how the conflict between both evolutionary processes can be modeled.

<sup>4</sup> In this discussion, I generally follow the notation in (Weibull [1995]).

<sup>5</sup> The term ‘growth rate’ is commonly used to refer to either (1) the rate of change of the size of the entire population or (2) the average reproductive fitness of the population. Here, I use the term ‘growth rate’ to mean the former. See (Deevey [1972], pp. 6–7).

<sup>6</sup> Often, the differential equations will be nonlinear, and the best we can do is approximate their solution.

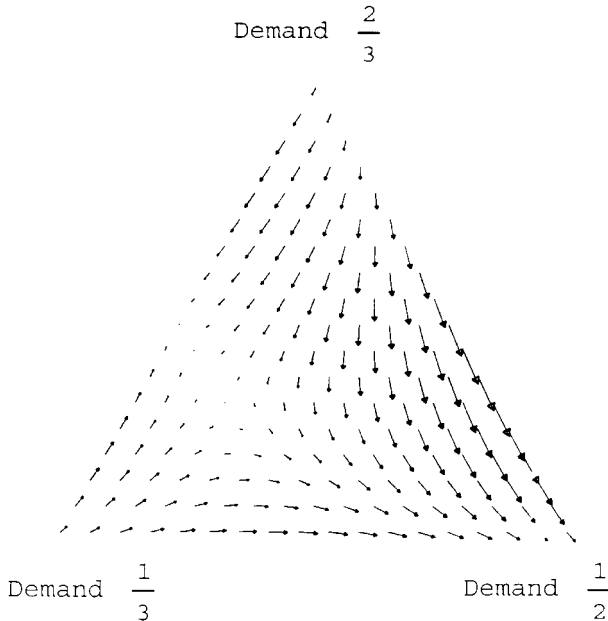
each strategy in a particular game, we can derive a function which takes an initial population state and time, and yields the population share of that strategy at that time. By doing so for each strategy and a variety of initial population states, we can often get a good picture of what states the populations will evolve toward, how quickly they will evolve to those states, and under what initial conditions.

This is the method Skyrms uses in his analysis of Divide-the-Cake. Since there are three strategies in the game, we represent possible states of the population as points in a triangle, where each vertex represents one of the three strategies. The points closer to each vertex represent a population where a larger share of individuals plays that strategy. So the point right on a vertex of the triangle represents a population in which every agent plays that strategy. The point in the middle of the triangle represents an equal representation of strategies in the population, and so on. At a sample of points in the triangle, we can program a computer to calculate the replicator dynamics for that population. In Figure 1, I have represented the direction in which the population evolves by an arrow, where the length of that arrow represents the rate at which the population shares change.<sup>7</sup> The strategy Demand 1/3 is the left-hand vertex, the strategy Demand 2/3 is the top vertex, and Demand 1/2 is represented by the right vertex.

Examining Figure 1, we immediately notice a few things. First, the majority of initial populations will eventually approach a state in which every agent plays Demand 1/2. Unless the population begins at a point very far from the right-hand vertex, the population shares of the other strategies will eventually go to zero. The region in which this happens is called the ‘basin of attraction’ for Demand 1/2. According to Skyrms, this basin of attraction comprises about 62% of the triangle ([1996], p. 16). The second thing we notice is another basin of attraction which leads to a state in which half of the population plays Demand 1/3 and the other half plays Demand 2/3. This is represented by a point on the edge of the triangle midway between the left and top vertices. Because this basin of attraction leads to a polymorphic state in which no player is a fair-dealer, Skyrms refers to it as a ‘polymorphic trap’ ([1996], p. 21).

The size of these polymorphic traps can be reduced by invoking an additional assumption—namely, that players are more likely to interact with others who play the same strategy. In fact, if the correlation is strong enough, the polymorphic traps can be eliminated entirely. This assumption has been questioned in depth by D’Arms *et al.* ([1998]), so I will not discuss it here. But if we grant the assumptions in Skyrms’s model, then it looks like evolutionary game theory does indicate an important asymmetry between the ‘fair’

<sup>7</sup> So the figure is actually a vector field. For a discussion of the use of vector fields in evolutionary game theory, see (Weibull [1995]).



**Fig. 1.** Evolution of a population playing Divide-the-Cake under the standard replicator dynamics.

Demand 1/2 strategy and the other strategies, and suggests circumstances under which the ‘fair’ strategy will evolve.

### 3 Other models and model sensitivity

The formal results which Skyrms reports are incontrovertible. The Divide-the-Cake game under the replicator dynamics is well understood, and all of Skyrms’s formal claims can be easily confirmed. Thus, Skyrms ends his initial discussion on a hopeful note when he says that:

In a finite population, in a finite time, where there is some random element in evolution, some reasonable amount of divisibility of the good and some correlation, we can say that it is likely that something close to share and share alike should evolve in dividing-the-cake situations. This is, perhaps, a beginning of an explanation of the origin of our concept of justice ([1996], p. 21).

For Skyrms, the evolutionary models suggest a method for explaining ‘the origin of our concept of justice’ because they seem to show that the evolution of fair behavior is probable. But do simple models, such as the replicator dynamics model of Divide-the-Cake, really show this? The way to answer this question is to examine the simplifying assumptions in Skyrms’s model, and

determine whether those simplifications relevantly affect its behavior. If it is found that the simplifying assumptions *do* have a significant effect on the behavior of the model, then we will have to modify our explanatory strategy.

### 3.1 Cultural evolution and other models

The first assumption I would like to question is that biological and cultural evolution may be modeled in the same way. The first thing to notice about this assumption is that it *is* an assumption. For, at least on the surface, the processes underlying learning and imitation are quite different from the processes of biological natural selection, so it is not as obvious as it may seem that biological and cultural evolution are really so closely analogous.

The second thing to notice about this assumption is how crucial it is to Skyrms's argument, because it is obviously a very difficult and unresolved question to what degree each process operates in the evolution of social norms. If cultural and biological evolution are relevantly different from each other, then it is quite possible that the two processes might lead to significantly different outcomes. The importance of this assumption has been noted by Philip Kitcher in his commentary, but he seems to accept it when he says that cultural evolution is 'compatible with the same dynamics' as biological evolution ([1999], p. 221).<sup>8</sup>

But is Skyrms's claim that 'the Darwinian story can be transposed into the context of cultural evolution, in which imitation and learning play an important role' ([1996], p. 11) really true? We are used to thinking of cultural and biological evolution as very closely analogous.<sup>9</sup> But on closer inspection we can see that the analogy between biological and cultural evolution breaks down much faster than is usually assumed. To see this, consider an alternate dynamics which models how game strategies spread through a population of agents who learn by imitation. Intuitively, the picture is this.<sup>10</sup> Players in the population live and interact with each other forever. Periodically, each player reviews how well her strategy is doing. If she judges that her strategy's payoff is falling below some threshold, she may choose to revise it. The process of revising a strategy consists in sampling a segment of the population and switching to a strategy which she judges has a higher payoff than her own. This dynamics is called the 'aspiration-imitation model' because players 'aspire' to a certain threshold of payoffs, and try to increase their payoffs by imitating others.

<sup>8</sup> Kitcher does express a little concern over this assumption in a footnote ([1999], p. 221), but he believes that the two processes are analogous enough to support Skyrms's analysis.

<sup>9</sup> This assumption is probably influenced by Dawkins's famous analogy between genes and 'memes', where a meme is 'a unit of information in a brain' which sometimes is copied from one communicating brain to another (Dawkins [1999], p. 109). On this rough analogy, genes and memes are both replicators, and would therefore seem to get the same evolutionary analysis.

<sup>10</sup> This discussion comes from (Samuelson [1997]) and (Weibull [1995]).

Given such a picture of the evolutionary process, we have to attend to two new variables. The first is the ‘review rate’  $r_i(x)$  which is the probability that a player playing strategy  $i$  in a population  $x$  will review her strategy in a given unit of time. The second is the ‘choice probability’  $p_i^j(x)$  that a reviewing player with strategy  $i$  will switch to strategy  $j$  in a population  $x$ . With these two functions in hand, we can easily formulate the rate of change  $\dot{x}$  for any population share of  $i$ -strategists. It will be equal to the proportion of players who switch to strategy from some other strategy, minus the proportion of  $i$ -strategists who switch to other strategies:

$$\dot{x} = \sum_{j \neq i} x_j r_j(x) p_i^j(x) - \sum_{j \neq i} x_i r_i(x) p_i^j(x) \quad (2)$$

Obviously, there are many different ways in which we could fill in the review rate and choice probability in Equation 2. For example, suppose that less successful agents review their strategies more often, so that the review rate is a linearly decreasing function:

$$r_i(x) = \alpha - \beta u(i, x) \quad (3)$$

Furthermore, suppose that players switch strategies by randomly selecting a member of the population and adopting that player’s strategy. So the probability of a reviewing  $i$ -strategist switching to strategy  $j$  is just the probability that the switching player will randomly pick a  $j$ -strategist, which is just the proportion of the population playing strategy  $j$ . So we have:

$$p_i^j = x_j \quad (4)$$

Equation 4 might be interpreted as modeling an environment in which the players have no information about the actual payoffs of other strategies and assume that more successful strategies are more common.

In a reply to critics, Skyrms defends his reliance on the replicator dynamics by correctly pointing out that a variety of different models yield behavior that is ‘qualitatively identical’ to it ([1999], pp. 244–5). Skyrms is right about this. In fact, if we insert Equations 3 and 4 into Equation 2, we get a system whose behavior is identical to that of the replicator dynamics, except that it may evolve at a different rate.<sup>11</sup>

Unfortunately, such equivalences between the replicator dynamics and other models does not justify our exclusive use of the replicator dynamics. When a different dynamics such as the aspiration-imitation model turns out to be equivalent to the replicator dynamics, that is an exceptional case. For the ‘aspiration-imitation model’ is not a single model, but rather a family of models whose members are generated by choosing among the many possible review rates and choice probabilities. In fact, the equivalence between the

<sup>11</sup> For proof, see Weibull ([1995], p. 154).



replicator dynamics and the aspiration-imitation model depends crucially on our choosing Equations 3 and 4 as the model's review rates and choice probabilities. But there are many environments for which other functions are more appropriate; in those environments, the aspiration-imitation model behaves quite differently.

For example, consider an environment in which the review rate is nonlinear. We might imagine players who are forced to update their strategies if their performance dips below a certain threshold value, but are highly reluctant to give up their strategies otherwise. If those players also had relatively good information about how well other strategies were performing, then they would be more likely to switch to high-performing strategies rather than merely popular strategies. To model this type of situation, we might define the review rate and choice probability as follows:

$$r_i = \left( 1 - \frac{u(i,x) - \min_j u(j,x)}{\max_j u(j,x) - \min_j u(j,x)} \right)^\mu \quad (5)$$

$$p_i^j(x) = \frac{\sigma^{u(j,x)}}{\sum_h \sigma^{u(h,x)}} \quad (6)$$

Equations 5 and 6 deserve some unpacking. In Equation 5, the quantities  $\max_j u(j,x)$  and  $\min_j u(j,x)$  refer to the highest possible payoffs in the population  $x$  and the lowest possible payoffs in  $x$ , respectively. Thus, for any strategy  $i$  and population  $x$ , the value of

$$\left( 1 - \frac{u(i,x) - \min_j u(j,x)}{\max_j u(j,x) - \min_j u(j,x)} \right)$$

will range from zero to one—zero if  $i$  is the best-performing strategy, and one if  $i$  is the worst-performing. In the special case where the exponent  $\mu = 1$ , the function is linear. But if we have  $\mu > 1$ , then the range of  $r_i$  still remains between zero and one, but will rise nonlinearly. The greater the value of  $\mu$ , the more nonlinear the function, so that in the limit as  $\mu$  approaches infinity,  $r_i$  becomes a step function.

Equation 6 may be interpreted in the following way. When a reviewing player decides which strategy to adopt, she looks at the payoff of every strategy in the population. For each strategy  $j$ , she assigns it a weight equal to  $\sigma^{u(j,x)}$ . If  $\sigma > 1$  then reviewing players will be more likely to switch to strategies with higher payoffs. By dividing this weight by the sum of all the weights for all the strategies, we guarantee that  $p_i^j(x)$  is a probability function.<sup>12</sup> As with Equation

<sup>12</sup> Where a probability function must have a range between zero and one, and the sum of all its values for every element in its domain must be equal to one.

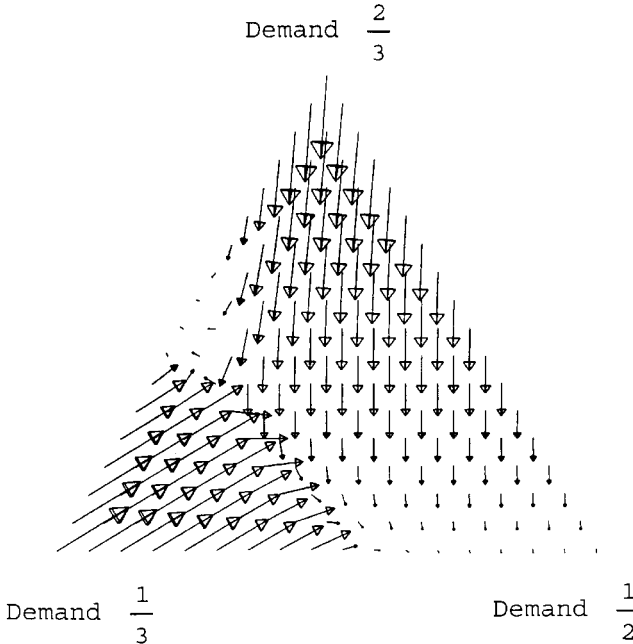
5, the larger the constant (this time the constant is  $\sigma$ ), the more sharply nonlinear the function.

If we use nonlinear functions such as Equations 5 and 6 for the review rate and choice probability, we will frequently be left with a dynamical system which differs significantly from the replicator dynamics. To see this, consider what happens in the model as  $\sigma$  increases. The constant  $\sigma$  can be interpreted as representing the quality of the information each player receives about the payoffs of other strategies. So as  $\sigma$  increases, players become increasingly likely to choose the best-performing strategies when they switch. Similarly, as  $\mu$  increases, the review rate becomes more sharply nonlinear. This means that players with inferior strategies will be much more likely to switch, while other players will be much more likely to keep their strategies, even if some benefit could be achieved by switching. The evolution of the system is given in Figure 2.

This system obviously behaves quite differently from the replicator dynamics pictured in Figure 1. Although we see regions of the triangle which roughly correspond to the same basins of attraction, the actual behavior of the system is quite different. The most striking feature is that if the initial population contains a mix of all three strategies, this version of the aspiration-imitation model *never* approaches a monomorphic state of 100% Demand 1/2. The reason is that as the greedy Demand 2/3 strategy dies out (as the population approaches the bottom of the triangle), these individuals switch either to Demand 1/3 or Demand 1/2. But the players playing Demand 1/3 are content with their constant payoff of  $\frac{1}{3}$ . The result is that the population shares of Demand 1/3 *and* Demand 1/2 *both* increase, although they do so very slowly, as the population share of Demand 2/3 approaches zero. Since two different strategies are always increasing in that region, the population cannot go to a monomorphic equilibrium of Demand 1/2.

It is important to note that there is nothing contrived about the aspiration-imitation model with nonlinear review rates and choice probabilities.<sup>13</sup> In fact, it's easy to think of many environments in which those probability functions are nonlinear. For example, we might imagine a group of well-informed investors who want to increase the return on their investments, but who must pay a large tax penalty for withdrawing their funds. Such investors will switch investments by buying and selling only when the difference between their own return and some other is greater than the tax penalty. This is a situation in which there is a clear threshold separating those who will switch strategies from those who will not. Given such a threshold, their switching behavior must be described by nonlinear functions such as the one in our model.

<sup>13</sup> The behavior in Figure 2 also appears over a wide range of values for  $\mu$  and  $\sigma$ .



**Fig. 2.** Evolution of Divide-the-Cake under the aspiration-imitation model with  $\sigma = 1.5$  and  $\mu = 4$ .

### 3.2 Effects of simplifying assumptions

In addition to the existence of alternative, non-equivalent dynamics, there is a second source of difficulty in employing evolutionary game theory in an analysis of this kind. This difficulty lies in the use of idealizing assumptions in the replicator dynamics. Although I do not believe that there is anything wrong with idealized models as such, we should take the time to recognize those idealizing assumptions and perform a careful inventory of their implications.

The assumption I will discuss here is what I will call the ‘continuous time’ assumption. In the derivation of the replicator dynamics, we think of the difference in population shares of some strategy  $i$  at some time  $t$  and later at time  $t + \tau$ . This is approximated by considering the expected payoff of strategy  $i$  at  $t$  and comparing it with the average payoff of the population. When we let  $\tau$  approach zero, we get Equation 1. Since we’ve moved from considering discrete intervals of time to thinking of a smooth evolution of the population over every instant of time, we say that this is a ‘continuous time’ model.

Although the assumption of continuous time may seem harmless enough, there are environments in which it makes a significant difference. To see why, consider a population whose makeup is changing in some interval of length  $\tau$ . If  $\tau$  is large, there will be many births and deaths (or many individuals changing strategies) during that interval. But as we shrink that interval of time, the probability that there will be a large number of individuals dying or giving birth (or changing strategies) during that time diminishes. In the limiting case where  $\tau$  has approached zero, and when we're assuming an infinitely large population, it is as if there is exactly one individual changing state each moment. In other words, when we have a model which operates in continuous time, the picture is one where the individuals change state *one at a time*.

On the other hand, in discrete time, the number of individuals who change state at a time is proportional to the size of the interval  $\tau$ . When a large number of agents make changes within the same interval, the dynamical system makes a non-continuous 'jump' from one state to another. As a real-world example of this phenomenon, think of the stock exchange. Overnight, many investors receive news and place orders which will all be executed as the market opens the next morning. Accordingly, we often see a sizable jump in the state of the stock market as all of these orders are (more or less) simultaneously executed. But if we imagine a fictitious stock exchange which operates in continuous time, the investors' orders will be executed one at a time. In this case, each investor would be able to respond to the prior movements of every other investor, and the market would not experience those sudden lurches.

In the kinds of dynamical systems studied in evolutionary game theory, the behavior of discrete-time systems differs from that of continuous-time systems in an important way. The population trajectories in continuous-time systems usually change according to smooth curves such as those in Figure 1. In a discrete-time system, the population shares move in the same *direction* as in a continuous-time system, but jump in that direction without occupying any intermediate states. The size of those jumps will depend on the size of the interval  $\tau$ . If  $\tau$  is large enough, the dynamical system may jump across important basins of attraction, just as the stock market may sometimes 'overcorrect' as large numbers of investors receive information or place orders simultaneously.<sup>14</sup>

In Divide-the-Cake, as in other games, the effect of using a discrete-time dynamics depends on the size of the interval  $\tau$  as well as on the choice of underlying dynamics (e.g. whether you use the replicator dynamics in discrete

<sup>14</sup> Sober gives an example of a system which is either cyclic, unstable, or a stable spiral attractor, depending on the underlying dynamics ([1994]).

time or the aspiration-imitation model in discrete time). For example, in a discrete-time version of the replicator dynamics, the behavior of the system will be arbitrarily close to that of the continuous-time system, provided  $\tau$  is sufficiently small. If  $\tau$  is larger, the system may cycle around a point which is a stable attractor in the continuous time dynamics. In a discrete-time version of the aspiration-imitation model with highly nonlinear review rates and choice probabilities, entire strategies may become extinct when they would never become extinct in a continuous-time model. This is because when we assume that  $\tau$  is very large and the review rates are nonlinear, the entire share of the population playing a particular poorly performing strategy may all decide to change strategies simultaneously (that is, within a single interval of length  $\tau$ ). This might happen, for instance, when Demand 2/3 has the lowest payoff. If everyone switches strategies simultaneously, then all of the Demand 2/3 players will switch, and Demand 2/3 will become extinct. But if time is continuous, then it is as if players review their strategies one at a time. In that case, as some of the Demand 2/3 players gradually switch to (say) Demand 1/3, the expected payoff of Demand 2/3 will go up. And if the payoff of Demand 2/3 goes up as some of the Demand 2/3 players switch strategies, then, after a certain point, we should see the remaining Demand 2/3 players decide to remain with their strategy.

Thus, when we construct an evolutionary game-theoretic model of some phenomenon, we have to make many choices. First, we must choose between different dynamics; in this section, I've focused on just the replicator dynamics and the aspiration-imitation model, but there are many others. Furthermore, when we analyze the behavior of those dynamics, we have to choose appropriate parameters and functions which will be plugged into the equations. As an example, we've seen how the aspiration-imitation model forces us to choose parameters and functions so that we have specific review rates and choice probabilities for our analysis. Furthermore, in order for the math to be tractable, we are forced to make various simplifying assumptions. As an example, we've seen how we must choose between discrete and continuous time. But again, there are many other simplifying assumptions which I have not discussed here, but are equally important. For example, Samuelson ([1997]) has an interesting discussion of how the assumption that populations are infinitely large affects the dynamics. Bergstrom and Stark discuss how the models behave differently if we drop the assumption of asexual reproduction ([1993]). A similar analysis of the effects of introducing sexual reproduction into the evolutionary models has recently been offered by Tennant ([1999]). All of these choices dramatically affect the probability that particular strategies will evolve to fixation. Some models, such as Skyrms's application of the replicator dynamics, may put the probability of Demand 1/2 evolving to fixation quite high. But other models, such as some versions of the

aspiration-imitation model, may place that probability vanishingly close to zero.

#### 4 Introducing group selection to evolutionary game theory

Of course, the considerations presented so far do not show that it is *impossible* for evolutionary models to demonstrate that fairness will evolve in a population. However, I believe that the previous arguments *do* show that in order for such an explanatory strategy to succeed, we have to pay careful attention to the evolutionary mechanisms the models are supposed to simulate.

If we do attend closely to the variety of mechanisms which might have encouraged the evolution of fairness, then we will be naturally led to a substantial shift in our explanatory strategies. Rather than attempting to base an explanation of the evolution of fairness on one general model such as Skyrms's application of the replicator dynamics, we should look at a variety of models which capture particular features of the lives of our evolutionary ancestors. For example, we might examine models which characterize their migration patterns, kinship relations, and other population structures. If such models were to reveal that particular features of our evolutionary ancestors' lives significantly encourage the evolution of fairness, then those models would provide an explanation. In this section, I will illustrate this approach by constructing a very simple metapopulation model which suggests how migration patterns favor the evolution of fairness.

We begin by considering a model proposed by Maynard Smith in an argument against group selection ([1976]). Maynard Smith's goal was to study under what conditions, if any, group selection might cause a gene for altruism to evolve. He defined 'altruism' as behavior which reduces the agent's own fitness, while increasing the fitness of others. It follows immediately from this definition of altruism that within any isolated population, altruistic individuals will eventually be driven to extinction while 'selfish' individuals will go to fixation.

Clearly, if individual selection works so efficiently *against* the evolution of altruism, group-selection mechanisms must operate to a very high degree if altruism is going to have a chance of spreading. Thus, Maynard Smith was forced into considering a highly unlikely set of circumstances which he thought were necessary for the evolution of altruism. Specifically, he proposed a model in which there is a species of mice which engage in either timid or aggressive behavior, where timidity is identified with altruism and aggression with selfishness. This behavior is determined by their genes: an *AA* or *Aa* mouse is of the aggressive/selfish type, whereas an *aa* mouse is timid/altruistic. Maynard Smith's insight was to recognize that even though

altruistic *aa* mice might be driven to extinction in any mixed population containing both types; a population consisting entirely of altruists would grow more quickly than other types of populations, provided that the benefit altruistic mice confer on others is sufficiently great.

Therefore, what is needed if altruism is to spread is a population structure in which the greater growth rate of altruistic groups is enough to counteract the individual-selection pressure which favors selfishness. The population structure Maynard Smith proposed was this. The mice live out their entire reproductive cycles in haystacks, where each haystack colony is founded by a single impregnated female. This founding female gives birth to a sibling group, which then reproduces for one year, during which time we assume each colony would go to fixation, leaving only pure selfish groups (unless the colony began with only altruists). The next summer, all of the mice are blended back into the global population, mating occurs, and new colonies are founded by a single impregnated female.

The Haystack Model shows that if the colonies are sufficiently isolated (that is, that they do not compete with each other for food and other resources), the vast majority of mating occurs only within colonies, and colonies of altruistic mice grow sufficiently more quickly than other colonies, then altruism can spread in the *global* population. However, these assumptions are so severe, and the population structure is so fanciful, that Maynard Smith concluded that group-selection effects will be negligibly small in the real world.

Before we give up on group-selection models, we should note that Maynard Smith's conclusion is a consequence of his extremely strong presumption that the altruistic gene will inevitably and quickly be driven to extinction within any mixed population (Sober and Wilson [1998], p. 70). However, our explanandum—the emergence of fairness—does not sanction this strong assumption. After all, fair behavior is significantly different from altruistic behavior. For what defines altruistic behavior is that the altruist is *unconditionally* worse-off because of her altruistic acts. This means that we would expect altruists to fare poorly against selfish individuals in *any* mixed population. But as we've seen above, the same is not true of fair individuals. For there are a variety of populations in which fair-dealers have *higher* fitness than their unfair counterparts.

Given that fair-dealers do not always have the lowest fitness in a mixed population, we need not examine how the highly artificial population structure of the Haystack Model affects the evolution of fairness in Divide-the-Cake. After all, that population structure was only necessary to counteract the extremely powerful individual-selection pressure which works against the evolution of altruism. However, although the motivation for the details of the Haystack Model does not apply to our problem of

explaining the evolution of fairness, the fundamental insight of Maynard Smith's model is still instructive. If populations are sufficiently isolated, and the growth rate of a population increases with the proportion of fair-dealers in the population, then the *global* population share of fair-dealers will increase significantly.

Introducing a small amount of formalism will help apply this to Divide-the-Cake. Call a metapopulation a 'set of populations occupying a cluster of habitable sites' (Wilson [1975], p. 108). To keep things simple, suppose that each population in the metapopulation is completely reproductively isolated, and that they do not compete with each other to any significant degree. Let each population in the metapopulation be denoted  $p_i$ .

Recall that  $u(x,x)$  denotes the average payoff in the population  $x$ . Economists often call this value the 'social efficiency' of the population. In our biological context, the social efficiency  $u(p_i,p_i)$  of a population  $p_i$  helps to determine the growth rate of the population. Since we interpret payoffs as fitnesses and fitness is a measure of the expected number of offspring, it follows that populations with greater social efficiency will have a higher rate of growth than populations with lower social efficiency. So for any population  $p_i$ , let  $P(p_i,t)$  be the size of the population  $p_i$  at some time  $t$ . Assuming that growth rates are linearly proportional to efficiency (which amounts to the simplifying assumption that there are no environmental constraints on population size), we have:

$$\frac{dP}{dt} = \lambda u(p_i,p_i)P(p_i,t) \quad (7)$$

So suppose we begin with a metapopulation which is initially seeded with populations, each of which has a random distribution of strategies. If the populations are isolated and there are no constraints on population size, then by assuming that the populations evolve according to some particular dynamics, we can discover what will happen in the metapopulation. For simplicity, suppose that each population evolves according to the replicator dynamics. By making this assumption, we can apply our discussion of Section 2; but as we will see, our choice of underlying dynamics is—surprisingly—largely unimportant.

As Skyrms demonstrates, there are two significant basins of attraction in the replicator dynamics model for Divide-the-Cake (when there is no correlation): a basin comprising 62% of possible populations which leads to the fixation of Demand 1/2, and another comprising the remaining 38% of possible populations, which leads to a state in which half the population plays Demand 1/3 and the other half plays Demand 2/3. Accordingly, if we use a replicator dynamics model, we should expect about 62% of the populations



in the metapopulation to evolve to Demand 1/2, and the rest to go to the polymorphic attractor.<sup>15</sup> Call the first set of populations the ‘fair populations’ and the rest the ‘greedy-modest’ populations.

To calculate the relative growth rates of the two kinds of populations, we must figure out the efficiency of each. Since the efficiency of a population is the average payoff in that population, the calculation is simple. The efficiency of the fair populations will be equal to the average payoff of Demand 1/2 in a monomorphic population of Demand 1/2, which is just  $\frac{1}{2}$ . To calculate the efficiency of the greedy-modest populations, we note that such a population contains many pairs of individuals. One-quarter of those pairs will be (Demand 1/3, Demand 1/3), in which case the average payoff is  $\frac{1}{3}$ . One-half of the pairs of individuals in the population will be (Demand 1/3, Demand 2/3), with an average payoff of  $\frac{1}{2}$ . And lastly, one-quarter of the pairs in the population will be (Demand 2/3, Demand 2/3), in which case they receive an average payoff of zero. So the average payoff in the polymorphic population is  $\frac{1}{4}(\frac{1}{3}) + \frac{1}{2}(\frac{1}{2}) + \frac{1}{4}(0) = \frac{1}{3}$ .

Since the efficiency of the fair populations is significantly larger than that of the greedy-modest populations, we should expect, by Equation 7, that each of the fair populations will eventually be much larger than the greedy-modest populations. By itself, this model does not say what will happen to the distribution of strategies in the metapopulation under a group-selection mechanism. But the model is suggestive, for it is natural to suppose that, under ordinary circumstances, group-selection mechanisms will tend to favor those strategies which predominate in the largest populations. For example, if populations send out colonies in proportion to their size so that larger populations colonize at higher rates than smaller populations, we would expect to see a large proportion of Demand 1/2 colonies. Similarly, if smaller populations are more likely to go extinct, then we’d expect the polymorphic populations to die out more frequently, leaving a larger proportion of Demand 1/2 populations in the metapopulation.

Furthermore, all of these observations hold, at least to some significant degree, regardless of whether the replicator dynamics or some other dynamics is used to describe the evolution of the populations. For the growth rate of a population will be proportional to their share of Demand 1/2 players, no matter what dynamics is chosen.<sup>16</sup> Furthermore, we should note that it is *not* necessary for any of the subpopulations in a group selection model to evolve to a state in which Demand 1/2 has gone to fixation. If a population has a relatively large share of Demand 1/2 players, then its growth rate will be

<sup>15</sup> This assumes that each possible initial population state is equally likely, which might not be true. But this assumption will not cause any problems for the analysis.

<sup>16</sup> Although we do require that higher payoffs entail higher reproductive fitness, this is not a significant constraint on the choice of evolutionary dynamics.

correspondingly higher, even if its share of Demand  $1/2$  players is less than one. So the result that group selection mechanisms will favor the spread of fair strategies should be robust across a wide range of dynamical models.

#### 4.1 The Island Model

Given the observations of the previous section, we should examine one particular group-selection mechanism in more detail. This will enable us to see how evolutionary game theory can be extended to take group-selection mechanisms into account.

In his ‘Isolation by Distance’ ([1943]) and in *Evolution and the Genetics of Populations* ([1969]), Sewall Wright examined how evolution is affected across a continuum of population structures. These structures ranged from nearly perfect isolation, as in the Haystack Model, to an environment in which all of the populations interact with each other to a high degree. Of these different population structures, the one I will examine here is the ‘Island Model’, so named by Wright because he thought that this described best a metapopulation whose populations live on different islands. In such an island environment, the majority of any particular population will be made up of individuals who are natives to that population, while some smaller proportion—call it  $m$ —are migrants. As an idealization, the Island Model assumes that the migrants to any particular island are representative of the entire metapopulation.

It turns out that if we apply the Island Model to the Divide-the-Cake game, fair-dealers will comprise an even *larger* share of the metapopulation than in the Haystack Model. To see this, note that some of the conclusions from the Haystack Model will carry over into this model. In particular, we should note that no matter what underlying dynamics is chosen, populations which have a larger share of fair-dealers will tend to have higher growth rates than other populations. For this reason, as time passes, we should expect a larger and larger share of all migrants to play Demand  $1/2$ . That is, the fair-dealers should come to comprise the greatest share of migrants to any island.

It is the higher proportion of fair-dealing migrants which causes the Demand  $1/2$  strategy to spread through the metapopulation. For consider a population in which Demand  $1/2$  has gone to fixation. If Demand  $2/3$  players immigrate to the population, they will always receive a payoff of zero, and will go extinct (or switch strategies) quickly. If Demand  $1/3$  players immigrate to the population, they will have an expected payoff of  $\frac{1}{3}$ , which will be consistently lower than the expected payoff of  $\frac{1}{2}$  which the native Demand  $1/2$  players receive. The only way for the population to move from all Demand  $1/2$  is if it is invaded by a proportionally huge number of Demand  $2/3$  players with a corresponding number of Demand  $1/3$  players. If this happened, then

the expected payoff of the Demand 2/3 immigrants could go over the expected payoff of the native Demand 1/2 population. However, since we're assuming that all players are equally likely to migrate, there will always be a compensating number of Demand 1/2 players immigrating to the population. Those players will tend to stabilize the population at Demand 1/2, so the Demand 1/3 and Demand 2/3 players will never be able to take over.

So populations of Demand 1/2 are stable. However, mixed populations can change. Consider the other stable attractor in the replicator dynamics in which half the population plays Demand 1/3 and the other half plays Demand 2/3. A group of Demand 1/2 immigrants will survive and reproduce in the population if their frequency in the population is greater than  $\frac{1}{7}$ .<sup>17</sup> In any such population with this 'critical mass' of Demand 1/2 immigrants, the population will eventually be taken over by the 'fair' strategy.<sup>18</sup>

## 4.2 Constructing a group-selection explanation

The arguments presented so far, coupled with the behavior of the two models in the previous sections, suggest a method for pursuing a robust explanation for the evolution of fairness.

My suggestion is that the evolutionary models can be adapted to provide a group-level explanation, whereby the evolution of fairness is attributed to the benefits such behavior imparts on the group. Although group-level explanations have been recently given a sustained defense by Sober and Wilson ([1998]), such explanations are usually met with suspicion. Skyrms himself suggests that group-level explanations of this sort are 'misguided', although he moderates that comment in an endnote ([1996], pp. 67, 121). As another example, Jonathon Elster characterizes such group-level explanations as 'just-so stories' ([1989], p. 148). For him, the very ease with which we can construct group-level explanations makes them unconvincing.

However, I think that we should not take these objections to be more substantive than they really are. For these authors, the problem with most group-level explanations is that they often do not demonstrate a plausible mechanism which allows group-selection for the behavior to outweigh the

<sup>17</sup> The Demand 1/2 players begin to take over the population when their expected payoff rises above that of the Demand 2/3 players. Let  $x_1$ ,  $x_2$ , and  $x_3$  be the population shares of Demand 1/3, Demand 1/2, and Demand 2/3 respectively. Since the population that will be invaded is comprised of Demand 1/3 and Demand 2/3 in equal proportions, let  $x_1 = x_3$ . Because  $x_1 + x_2 + x_3 = 1$ , we have  $x_1 = \frac{1}{2} - \frac{x_2}{2}$ . The expected payoff of Demand 1/2 is greater than that of Demand 2/3 when  $\frac{1}{2}(x_1 + x_2) > \frac{2}{3}x_1$ . Substituting for  $x_1$ , we get  $\frac{1}{2}(\frac{1}{2} - \frac{x_2}{2} + x_2) > \frac{2}{3}(\frac{1}{2} - \frac{x_2}{2})$ , which simplifies to  $x_2 > \frac{1}{7}$ .

<sup>18</sup> This discussion mirrors some game-theoretic definitions of stability. One standard definition of a stable strategy is one which no mutant strategy can invade, provided that the number of mutants which enter the population at any one time is relatively small.

individual-selection against it.<sup>19</sup> So if we are to meet the objection, we must posit a specific mechanism which operated at the level of the group.<sup>20</sup> That mechanism must encourage the ‘fair’ behaviors to evolve, even when such behaviors may be selected *against* at the level of the individual.<sup>21</sup>

Fair behaviors are often the ones which raise a population’s efficiency, where ‘efficiency’ is understood as average fitness. My claim in the previous sections has been that some population structures encourage the spread of such efficiency-raising behaviors. This was suggested in the discussion of the Haystack Model, when it became clear that populations of fair-dealers will have a higher efficiency, and therefore a higher growth rate, than other populations. In the population structure of the Island Model, that higher growth rate meant that we’d expect to see a larger proportion of migrants who play the intuitively ‘fair’ strategy. When the migrants tend to be fair-dealers, we see the ‘fair’ strategy spread through the entire metapopulation. Thus, the Island Model shows that when the metapopulation is comprised of relatively isolated populations which send out migrants, efficiency-raising behaviors will tend to spread.

There are many other population structures which tend to spread efficiency-raising behaviors. Colonization, inter-group competition for resources, differential group extinction, and many other plausible structures may play such a role. By examining our close relatives among the non-human primates, we can infer which population structures were present in our own evolutionary history.<sup>22</sup> We may then adapt our game-theoretic models in such a way as to represent the evolutionary impact of these population structures. Although this explanatory strategy will not necessarily allow us to derive probabilities for the evolution of fairness norms, it *will* allow us to identify historically real population structures and demonstrate that they encourage the evolution of fairness.

<sup>19</sup> In this respect, these authors’ worries exactly parallel those of Gould and Lewontin, who warn against being overeager to construct adaptationist explanations in evolutionary biology ([1979]). Gould and Lewontin are quite correct when they point out that traits may evolve for reasons having little or nothing to do with any adaptive benefits they may confer on the organism. But of course, this does not mean that adaptationist explanations are *never* appropriate. We merely have to be careful, when proposing adaptationist explanations, that there is a plausible mechanism which actually selected for that trait.

<sup>20</sup> Of course, this is not the only objection to group-level explanations. Jonathon Elster, for example, has a series of arguments against them ([1989], pp. 147–9). Unfortunately, there is not enough room here to discuss them.

<sup>21</sup> The approach that I advocate here has been anticipated to some extent by Robert Boyd and Peter Richerson, who discuss a dual model of evolution, where mechanisms of biological and cultural evolution are allowed to act simultaneously on the evolution of a species ([1976]). Like the present analysis, Boyd and Richerson emphasize the simulation of specific evolutionary mechanisms, and would allow both group-selection and individual-selection mechanisms to operate to varying degrees ([1976], p. 261).

<sup>22</sup> See (Sober [1988]) for a detailed discussion of how cladistics may be used to infer such facts about our evolutionary ancestors.

It is important to note that this explanatory strategy has a virtue which we haven't yet addressed. I said above that fair behaviors are often the ones which raise the population's efficiency—'often', but not 'always'. Since there are cases where fairness does not correlate with efficiency, group-level selection, as a working hypothesis, can explain why fairness sometimes does *not* evolve. We should not be too optimistic about human nature. Human interactions are often unfair. Thus, a good account of the evolution of fairness should have something to say about unfairness as well.

Consider the so-called 'Ultimatum Game' (sometimes called the 'Proposer-Disposer' game), which has been studied extensively by economists, as well as by Skyrms. This game is played by two people: the Proposer and the Disposer. As in Divide-the-Cake, there is a surplus to be divided between them. The Proposer makes some offer as to how the surplus is to be divided, such as 'I'll take two-thirds and you take one-third'. The Disposer has two options: accept or reject. If the Disposer accepts, then the surplus is divided exactly according to the Proposer's offer. If the Disposer rejects, then neither gets anything.

Just as we saw in Divide-the-Cake, there is an intuitively obvious 'fair' division of the surplus in which Proposer and Disposer each get half—and in experimental studies (even across different cultures) this is approximately what happens (Güth, Schmittberger, and Schwarze [1982]; Roth [1985]). But in the vast majority of evolutionary models, this is *not* what happens. Instead, the most *unfair* division soon sweeps through the population in which the Proposer offers to take as much as she possibly can and the Disposer accepts anything. This happens because no matter what offer the Proposer makes, the Disposer is better off accepting—for if the Disposer rejects, then she gets nothing at all. Given that the Disposer will accept anything, the Proposer is better off making her proposals as greedy as possible.<sup>23</sup>

For our purposes, the important point about the Ultimatum Game is that in it, there is no correlation between fairness and group efficiency—so long as the Disposer accepts the offer, the entire surplus is used. Thus, even if the

<sup>23</sup> Many hypotheses have been offered to account for the disparity between the standard evolutionary models of the Ultimatum Game and the experimental results. The leading explanations seem to be that the test subjects place intrinsic value on fairness, or that they want to be perceived as fair by their peers. It is unclear what significance, if any, this disparity has for the project of explaining the evolution of fairness norms. After all, the agents in our evolutionary models cannot be guided by norms of fairness, since we are attempting to explain the origins of such norms. Thus, the evolutionary models which I have considered in this paper cannot be falsified by these experimental results, provided the test subjects are, in fact, guided by norms of fairness. Rather, I suspect that the lesson we learn from the Ultimatum Game is that, to the extent that games like it describe the interactions of our evolutionary ancestors, we should expect a propensity towards *unfairness* to have evolved. See Davis and Holt ([1993], p. 269) and Thaler ([1988]) for surveys of research on the Ultimatum Game, and Rabin ([1990]) for a detailed theory of how to model agents who are guided by norms of fairness.

population structure favors populations with high efficiency, we should not expect anything like fairness to evolve.

So what the group-selection models seem to suggest is this. Often, fairness correlates with efficiency. When it does, and the appropriate population structures are present, group-selection will lead fairness to spread through the global population. If fairness does not correlate with efficiency, as in the Ultimatum Game, all bets are off—groups of fair-dealers will not necessarily out-compete other groups and fairness will probably not spread through the population. In cases where it is unclear how game-playing strategies impact the group's efficiency, an appropriate group-selection model can help us to predict or explain how certain behaviors, including fairness, are established.

## 5 Conclusion

Given the novelty of Skyrms's use of evolutionary game theory in explaining the evolution of the social contract, it is not surprising that discussions of his work tend to focus on methodological issues. The purpose of this paper has been to raise some of these issues and suggest ways to resolve them.

The most important methodological problem with an evolutionary game theory approach is that the models are highly sensitive to the kind of process which they are intended to analyze. An important example of this sensitivity is that models of biological reproduction (e.g. the replicator dynamics) have significantly different behavior from models of learning processes (e.g. the aspiration-imitation model). However, this robustness failure does not stand in the way of pursuing Skyrms's project. It just shows that we must use the models to simulate specific evolutionary mechanisms, and not make general claims based on the behavior of a small number of models.

As a working hypothesis, it is not implausible to think that the explanation for the emergence of norms of fairness has a great deal to do with the benefits such norms impart to the group. We may superimpose a group-selection model onto an evolutionary game-theoretic model as in Section 4. This will allow us to study how group-selection mechanisms favor the spread of norms of fairness. When we understand which group-selection mechanisms favor the spread of fairness norms, we will have an explanation of the evolution of fairness.

There is another benefit to pursuing group-selection models. In his commentary ([1999]), Philip Kitcher suggests that the next line of research in this area should be modeling coalition-formation. His interest in modeling coalitions is motivated by a desire to ensure that the game-theoretic models are really explanatory. For, as Kitcher points out, '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' ([1999], p. 222). He goes on to suggest that if we look to the pre-historic record, we'll find that 'our ancestors lived in larger social groups, comparable to those of present-day chimpanzees and bonobos' ([1999], p. 224) who form coalitions and alliances. Since coalitions are a particular kind of group, the first step toward understanding the dynamics of coalitions may be to examine simple group-selection models. Additionally, we should look at other sorts of population structures which were common among our evolutionary ancestors. If the evolutionary models show that those structures favor the evolution of fairness, we will have taken a significant step in explaining the origins of fairness.

### Acknowledgements

This paper has improved a great deal from comments by Sara Rachel Chant, Malcolm Forster, Daniel Hausman, Brian Skyrms, Elliott Sober, David Sloan Wilson, and two anonymous referees from this journal. I am also very grateful to the Department of Philosophy at Frostburg State University—and especially to Thomas Mappes—for valuable support and hospitality over the past year.

*Department of Philosophy  
University of Wisconsin-Madison  
Madison, WI 53706, USA  
zjernst@students.wisc.edu*

### References

- Bergstrom, T. and Stark, O. [1993]: 'How Altruism Can Prevail in an Evolutionary Environment', *The American Economic Review*, **83**, pp. 149–55.
- Boyd, R. and Richerson, P. [1976]: 'A Simple Dual Inheritance Model of the Conflict Between Social and Biological Evolution', *Zygon*, **11**, pp. 254–62.
- D'Arms, J., Batterman, R., and Górný, K. [1998]: 'Game Theoretic Explanations and the Evolution of Justice', *Philosophy of Science*, **65**, pp. 76–102.
- Davis, D. and Holt, C. [1993]: *Experimental Economics*, Princeton, NJ: Princeton University Press.
- Dawkins, R. [1999]: *The Extended Phenotype*, Oxford: Oxford University Press.
- Deevey, E. [1972]: 'The Equilibrium Population', in R. Lazarus (ed.), *Readings in Population*, London: Macmillan, pp. 2–16.
- Elster, J. [1989]: *The Cement of Society: A Study of Social Order*, Cambridge and New York: Cambridge University Press.
- Gould, S. and Lewontin, R. [1979]: 'The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme', in E. Sober (ed.), *Conceptual Issues in Evolutionary Biology*, 1984, Cambridge, MA: MIT Press, pp. 252–70.

- Güth, W., Schmittberger, R., and Schwarze, B. [1982]: 'An Experimental Analysis of Ultimatum Bargaining', *Journal of Economic Behavior and Organization*, **3**, 367–88.
- Kitcher, P. [1999]: 'Games Social Animals Play: Commentary on Brian Skyrms's Evolution of the Social Contract', *Philosophy and Phenomenological Research*, **59**, pp. 221–28.
- Rabin, M. [1990]: 'Incorporating Fairness Into Game Theory', *American Economic Review*, **83**, pp. 1281–302.
- Roth, A. [1985]: *Game-Theoretic Models of Bargaining: Introduction and Overview*, New York: Cambridge University Press.
- Samuelson, L. [1997]: *Evolutionary Games and Equilibrium Selection*, Cambridge, MA: MIT Press.
- Skyrms, B. [1994a]: 'Darwin Meets the Logic of Decision: Correlation in Evolutionary Game Theory', *Philosophy of Science*, **61**, pp. 503–28.
- Skyrms, B. [1994b]: 'Sex and Justice', *Journal of Philosophy*, **91**, pp. 305–20.
- Skyrms, B. [1996]: *Evolution of the Social Contract*, Cambridge and New York: Cambridge University Press.
- Skyrms, B. [1999]: 'Reply to Critics', *Philosophy and Phenomenological Research*, **59**, pp. 243–54.
- Smith, Maynard J. [1976]: 'Group Selection', *Quarterly Review of Biology*, **51**, pp. 277–83.
- Sober, E. [1988]: *Reconstructing the Past: Parsimony, Evolution, and Inference*, Cambridge, MA: MIT Press.
- Sober, E. [1994]: 'The Primacy of Truth-telling and the Evolution of Lying', in E. Sober (ed.), *From a Biological Point of View: Essays in Evolutionary Philosophy*, New York: Cambridge University Press, pp. 71–92.
- Sober, E. and Wilson, D. S. [1998]: *Unto Others: The Evolution and Psychology of Unselfish Behavior*, Cambridge, MA: Harvard University Press.
- Tennant, N. [1999]: 'Sex and the Evolution of Fair-Dealing', *Philosophy of Science*, **66**, pp. 391–414.
- Thaler, R. [1988]: 'Anomalies: The Ultimatum Game', *Journal of Economic Perspectives*, **2**, pp. 195–206.
- Weibull, J. [1995]: *Evolutionary Game Theory*, Cambridge, MA: MIT Press.
- Wilson, E.O. [1975]: *Sociobiology: The New Synthesis*, Cambridge, MA: Harvard University Press.
- Wright, S. [1943]: 'Isolation by Distance', *Genetics*, **28**, pp. 114–38.
- Wright, S. [1969]: *The Theory of Gene Frequencies*, Vol. 2, *Evolution and the Genetics of Populations: A Treatise in Three Volumes*, Chicago: University of Chicago Press.