SEX AND JUSTICE*

> Some have not hesitated to attribute to men in that state of nature the concept of just and unjust, without bothering to show that they must have had such a concept, or even that it would be useful to them.
>
> —Jean-Jacques Rousseau

In 1710, there appeared in the *Philosophical Transactions of the Royal Society of London* a note entitled "An argument for Divine Providence, taken from the constant Regularity observ'd in the Births of both Sexes." The author, Dr. John Arbuthnot,[1] was identified as "Physitian in Ordinary to Her Majesty, and Fellow of the College of Physitians and the Royal Society." Arbuthnot was not only the Queen's physician. He had a keen enough interest in the emerging theory of probability to have translated the first textbook on probability, Christian Huygens's *De Ratiociniis in Ludo Aleae,* into English—and to have extended the treatment to a few games of chance not considered by Huygens.

Arbuthnot argued the balance between the numbers of the men and women was a mark of divine providence "for by this means it is provided that the Species shall never fail, since every Male shall have its Female, and of a Proportionable Age." The argument is not simply from approximate equality of the number of sexes at birth. Arbuthnot notes the males suffer a greater mortality than females, so that exact equality of numbers at birth would lead to a deficiency of males at reproductive age. A closer look at birth statistics shows that "To repair that loss, provident Nature, by the disposal of its wise Creator, brings forth more Males than Females; and that in almost constant proportion." Arbuthnot supports the claim with a table of christenings in London from 1629–1710 which shows a regular excess of males and with a calculation to show that the probability of getting such a regular excess of males by chance alone was exceedingly small. (The calculation has been repeated throughout

* I would like to thank Alan Gibbard, Bill Harper, Richard Jeffrey, and Barbara Mellers for comments on an earlier draft of this paper. It was completed while the author was a Fellow at the Center for Advanced Study in the Behavioral Sciences. I am grateful for financial support provided by the National Science Foundation, the Andrew Mellon Foundation, and the University of California President's Research Fellowship in the Humanities.

[1] *Philosophical Transactions of the Royal Society of London,* XXVII (1710): 186–90.

0022-362X/94/9106/305–20

the history of probability[2] with larger data sets, and with the conclusion that the male-biased sex ratio at birth in humans is real.) Arbuthnot encapsulates his conclusion in this scholium:

> From hence it follows that Polygamy is contrary to the Law of Nature and Justice, and to the Propagation of Human Race; for where Males and Females are in equal number, if one Man takes Twenty Wives, Nineteen Men must live in Celibacy, which is repugnant to the Design of Nature; nor is it probable that Twenty Women will be so well impregnated by one Man as by Twenty (*op. cit.*, p. 189).

Arbuthnot's note raises two important questions. The fundamental question—which emerges in full force in the scholium—asks why the sex ratio should be anywhere near equality. The answer leads to a more subtle puzzle: Why there should be a slight excess of males? Arbuthnot's answer to the fundamental question is that the creator favors monogamy, and this leads to his answer to the second question. Given the excess mortality of males—for other reasons in the divine plan—a slight excess of males at birth is required to provide for monogamy. Statistical verification of the excess of males—for which there is no plausible alternative explanation—is taken as confirmation of the theory.

The reasoning seems to me somewhat better than commentators make it out to be, but it runs into difficulties when confronted with a wider range of biological data. The sex ratio of mammals in general, even harem-forming species, is close to 1/2. In some such species, twenty females are well impregnated by one male. A significant proportion of males never breed and appear to serve no useful function. What did the creator have in mind when he made antelope and elephant seals?

If theology does not offer a ready answer to such questions, does biology do any better? In 1871, C. Darwin[3] could not give an affirmative answer:

> In no case, as far as we can see, would an inherited tendency to produce both sexes in equal numbers or to produce one sex in excess, be a direct advantage or disadvantage to certain individuals more than to others; for instance, an individual with a tendency to produce more males than females would not succeed better in the battle for life than an individual with an opposite tendency; and therefore a tendency of this kind could not be gained through natural selection. . . . I formerly thought that when a tendency to produce the two sexes in equal num-

[2] In this regard, see S. Stigler, *The History of Statistics: The Measurement of Uncertainty before 1900* (Cambridge: Harvard, 1986).

[3] *The Descent of Man, and Selection in Relation to Sex* (London: Murray, 1871; 2nd rev. ed., New York: Appelton, 1898).

bers was advantageous to the species, it would follow from natural selection, but I now see that the whole problem is so intricate that it is safer to leave its solution for the future (*ibid.*, p. 263).

I. THE PROBLEM OF JUSTICE

Here we start with a very simple problem; we are to divide a chocolate cake between us. Neither of us has any special claim as against the other. Our positions are entirely symmetric. The cake is a windfall for us, and it is up to us to divide it. But if we cannot agree how to divide it, the cake will spoil and we will get nothing. What we ought to do seems obvious. We should share alike.

One might imagine some preliminary haggling: "How about ⅔ for me, ⅓ for you? No, I'll take 60% and you get 40% . . ."; but in the end each of us has a bottom line. We focus on the bottom line, and simplify even more by considering a model game.[4] Each of us writes a final claim to a percentage of the cake on a piece of paper, folds it, and hands it to a referee. If the claims total more than 100%, the referee eats the cake. Otherwise, we get what we claim. (We may suppose that if we claim less than 100% the referee gets the difference.)

What will people do, when given this problem? I expect that we would all give the same answer—almost everyone will claim half the cake. In fact, the experiment has been done. R. V. Nydegger and G. Owen[5] asked subjects to divide a dollar among themselves. There were no surprises. All agreed to a 50–50 split. The experiment is not widely discussed because it is not thought of as an anomaly. The results are just what everyone would have expected. It is this uncontroversial rule of fair division to which I want to direct attention.

Experimenters have even found that this rule of fair division is often generalized to other games where it may be thought of as an anomaly: to ultimatum games—where one player gets to propose a division and the other has to take it or get nothing; and even to dictator games—where one player simply gets to decide how the cake is divided.[6] There is some controversy about the strength of the

[4] Due to John Nash; see his "The Bargaining Problem," *Econometrica,* XVIII (1950): 155–62.

[5] "Two-Person Bargaining, An Experimental Test of the Nash Axioms," *International Journal of Game Theory,* III (1974): 239–50.

[6] There is a large literature, to which I shall give only a few references: O. Bartos, "Negotiation and Justice," *Contributions to Experimental Economics,* VII (1978): 103–26; R. Selten, "The Equity Principle in Economic Behavior," in *Decision Theory and Social Ethics,* H. Gottinger and W. Leinfellner, eds. (Cambridge: Reidel, 1978), pp. 289–301; W. Güth, R. Schmittberger, and B. Schwarze, "An Experimental Analysis of Ultimatum Bargaining," *Journal of Economic Behavior and Organization,* III (1982): 367–88; D. Kahneman, J. Knetsch, and R. Thaler, "Fairness and the Assumptions of Economics," *Journal of Business,* XLIX (1986):

generalization, but that is not important here. We are interested in the strength of the fair-division rule in the original game, and any generalization to these other situations is an indication of its robustness in the context of interest.

We think we know the right answer to the original problem, but why is it right? In what sense is it right? Let us see whether informed rational self-interest will give us an answer. If I want to get as much as possible, the best claim for me to write down depends on what you write down. I do not want the total to go over 100% so that we get nothing, but I do not want the total to be less than 100% either. Likewise, your optimum claim depends on what I write down. We have two interacting optimization problems. We want a solution to our problem to consist of solutions to each optimization problem which are in equilibrium.

We have an *equilibrium in informed rational self-interest* if each of our claims is optimal given the other's claim. In other words, given my claim you could not do better by changing yours and given your claim I could do no better by changing mine. This equilibrium is the central equilibrium concept in the theory of games. It was used already by Cournot, but is usually called a *Nash equilibrium* after John Nash,[7] who showed that such equilibria exist in great generality. Such an equilibrium would be even more compelling if it were not only true that one could not gain by unilaterally deviating from it, but also that on such a deviation one would definitely do worse than one would have done at equilibrium. An equilibrium with this additional stability property is a *strict Nash equilibrium.*

If we each claim half of the cake, we are at a strict Nash equilibrium. If one of us had claimed less, he would have got less. If one of us had claimed more, the claims would have exceeded 100% and he would have got nothing. There are, however, many other strict Nash equilibria as well. Suppose that you claim ⅔ of the cake and I claim ⅓. Then we are again at a strict Nash equilibrium for the same reason. If either of us had claimed more, we would both have got nothing, if either of us had claimed less, he would have got less. In fact, every pair of positive[8] claims that total 100% is a strict Nash

---

S285–300; Thaler, "Anomalies: The Ultimatum Game," *Journal of Economic Perspectives,* II (1988): 195–206; and more generally all the papers in *Psychological Perspectives on Justice,* B. Mellers and J. Baron, eds. (New York: Cambridge, 1993). I shall discuss ultimatum games in a separate essay on "Justice and Commitment."

[7] "Non-Cooperative Games," *Annals of Mathematics,* LIV (1951): 286–95.

[8] If I claim nothing and you claim 100%, we are still at a Nash equilibrium, but not a strict one. For if I were to deviate unilaterally I could not do worse, but I could also not do better.

equilibrium. There is a profusion of strict equilibrium solutions to our problem of dividing the cake, but we want to say that only one of them is *just*. Equilibrium in informed rational self-interest, even when strictly construed, does not explain our conception of justice.

Justice is blind, but justice is not completely blind. She is not ignorant. She is not foolish. She is informed and rational, but her interest—in some sense to be made clear—is not self-interest. Much of the history of ethics consists of attempts to pin down this idea. J. Harsanyi[9] and John Rawls[10] construe just rules or procedures as those which would be got by rational choice behind what Rawls calls a "veil of ignorance": "Somehow we must nullify the effects of specific contingencies which put men at odds and tempt them to exploit social and natural circumstances to their own advantage. In order to do this I assume that parties are situated behind a veil of ignorance."[11] Exactly what the veil is supposed to hide is a surprisingly delicate question, which I shall not pursue here. Abstracting from these complexities, suppose that you and I are supposed to decide how to divide the cake between individuals *A* and *B*, under the condition that a referee will later decide whether you are *A* and I am *B* or conversely. We are supposed to make a rational choice under this veil of ignorance.

Well, who is the referee and how will she choose? I would like to know in order to make my rational choice. In fact, I do not know how to make a rational choice unless I have some knowledge, or some beliefs, or some degrees of belief about this question. If the referee likes me, I might favor 99% for *A*, 1% for *B*; or 99% for *B*, 1% for *A* (I do not care which) on the theory that fate will smile upon me. If the referee hates me, I shall favor equal shares.

It might be natural to say: "Don't worry about such things. They have nothing to do with justice. The referee will flip a fair coin." This is essentially Harsanyi's position. Now, if all I care about is expected amount of cake—if I am neither risk averse nor a risk seeker—I shall judge every combination of portions of cake between *A* and *B* which uses up all the cake to be optimal. 99% for *A* and 1% for *B* is just as good as 50%-50%, as far as I am concerned. The situation is the same for you. The Harsanyi-Rawls veil of ignorance has not helped with this problem (though it would with others). We are left with all the strict Nash equilibria of the bargaining game (plus the 100%-0% divisions).

[9] "Cardinal Utility in Welfare Economics and the Theory of Risk Taking," *Journal of Political Economy*, LXI (1953): 343–5.

[10] "Justice as Fairness" this JOURNAL, LIV, 22 (October 24, 1957): 653–62.

[11] *A Theory of Justice* (Cambridge: Harvard, 1971), p. 36.

Rawls does not have the referee flip the coin. We do not know anything at all about Ms. Fortuna. In my ignorance, he argues, I should act as if she does not like me. So should you. We should follow the decision rule of maximizing minimum gain. Then we shall both agree on the 50%-50% split. This gets us the desired conclusion, but on what basis? Why should we both be paranoid? After all, if there is an unequal division between *A* and *B*, Fortuna can not very well decide against both of us. This discussion could, obviously, be continued.[12] But having introduced the problem of explaining our conception of justice, I would like to pause in this discussion and return to the problem of sex ratios.

II. EVOLUTION AND SEX RATIOS

R. A. Fisher, in his great book *The Genetical Theory of Natural Selection,*[13] saw the fundamental answer to Darwin's puzzle about the evolution of sex ratios and at the same time laid the foundation for game-theoretic thinking in the theory of evolution. Let us assume, with Darwin, that the inherited tendency to produce both sexes in equal numbers, or to produce one sex in excess, does not affect the expected number of children of an individual with that tendency, and let us assume random mating in the population. Fisher pointed out that the inherited tendency can nevertheless affect the expected number of grandchildren.

In the species under consideration, every child has one female and one male parent and gets half its genes from each. Suppose there were a preponderance of females in the population. Then males would have more children on average than females and would contribute more genes to the next generation. An individual who carried a tendency to produce more males would have a higher expected number of grandchildren than the population average, and that genetically based tendency would spread through the population. Likewise, in a population with a preponderance of males, a genetic tendency to produce more females would spread. There is an evolutionary feedback that tends to stabilize at equal proportions of males and females.

Notice that this argument remains good even if a large proportion of males never get to breed. If only half the males get to breed, then males that breed are twice as valuable in terms of reproductive fitness. Producing a male offspring is like buying a lottery ticket on a breeding male. Probability ½ of twice as much yields the same expected reproductive value. The argument is general. Even if 90% of

[12] See Rawls, *A Theory of Justice,* pp. 152ff.; and Harsanyi, "Can the Maximin Principle Serve as a Basis for Morality?" *American Political Science Review,* LXIX (1975): 594–606.

[13] New York: Oxford, 1930.

the males were eaten before having a chance to breed—as is the case with domestic cattle—evolutionary pressures will still drive the sex ratio to unity.

With this treatment of sex ratio, Fisher introduced strategic—essentially game-theoretic—thinking into the theory of evolution. What sex-ratio propensity is optimal for an individual depends on what sex-ratio propensities are used by the other members of the population. A tendency to produce mostly males would have high fitness in a population that produced mostly females but a low fitness in a population that produced mostly males. The tendency to produce both sexes in equal numbers is an *equilibrium* in the sense that it is optimal relative to a population where everyone has it.

We now have a dynamic explanation of the general fact that the proportions of the sexes in mammals are approximately equal. But what about Arbuthnot's problem? Why are they not exactly equal in man? Arbuthnot's argument that the excess of males in the human population cannot simply be due to sampling error has been strengthened by subsequent studies. Fisher has an answer to this problem as well. The simplified argument that I have given so far assumes that the parental cost of producing and rearing a male is equal to that of producing and rearing a female. To take an extreme case, if a parent using the same amount of resources could produce either two males or one female, and the expected reproductive fitness through a male were more than ½ of that through a female, it would pay to produce the two males. Where the costs of producing and rearing different sexes are unequal, the evolutionary feedback leads to a propensity for equal parental investment in both sexes, rather than to equal proportions of the sexes.

The way Fisher applies this to humans depends on the fact that here the sex ratio changes during the time of parental care. At conception, the ratio of males to females is perhaps as high as 120 to 100. But males experience greater mortality during parental care, with males and females being in about equal proportion at maturity, and females being in the majority later. The correct period to count as the period of parental care is not entirely clear, since parents may care for grandchildren as well as children. Because of the higher mortality of males, the average parental expenditure for a male at the end of parental care will be higher than that for a female, but the expected parental expenditure for a male at birth should be lower. Then it is consistent with the evolutionary argument that there should be an excess of males at conception and birth which changes to an excess of females at the end of the period of parental care. Fisher remarks: "The actual sex-ratio in man seems to fulfill these conditions quite closely" (*op. cit.*, p. 159).

III. JUSTICE: AN EVOLUTIONARY FABLE

How would evolution affect strategies in the game of dividing a cake? We start by building an evolutionary model. Individuals, paired at random from a large population, play the bargaining game of section I. The cake represents a quantity of Darwinian fitness—expected number of offspring—which can be divided and transferred. Individuals reproduce, on average, according to their fitness and pass along their strategies to their offspring. In this simple model, individuals have strategies programmed in, and the strategies replicate themselves in accord with, the evolutionary fitness that they receive in the bargaining interactions.

Notice that in this setting it is the strategies that come to the fore, while the individuals that implement them on various occasions recede from view. Although the episodes that drive evolution here are a series of two-person games, the payoffs are determined by what strategy is played against what strategy. The identity of the individuals playing is unimportant, and is continually shifting. This is the *Darwinian veil of ignorance.* It has striking consequences for the evolution of justice.

Suppose that we have a population of individuals demanding 60% of the cake. Meeting each other, they get nothing. If anyone were to demand a positive amount less than 40%, she would get that amount and thus do better than the population average; likewise, for any population of individuals that demand more than 50% (and less than 100%). Suppose we have a population demanding 30%. Anyone demanding a bit more will do better than the population average; likewise, for any amount less than 50%. This means that the only strategies[14] that can be equilibrium strategies under the Darwinian veil of ignorance are demand 50% and demand 100%.

The strategy demand 100% is an equilibrium, but an unstable one. In a population where everyone demands 100%, everyone gets nothing; and if a mutant popped up who made a different demand against 100%ers, she would also get nothing. But suppose that a small proportion of modest mutants arose who demanded, for example, 45%. Most of the time they would be paired with 100%ers and get nothing, but some of the time they would be paired with each other and get 45%. On average, their payoff would be higher than that of the population, and they would increase.

On the other hand, demand 50% is a stable equilibrium. In a population where everyone demands half of the cake, any mutant who demanded anything different would get less than the population average. Demanding half of the cake is an evolutionarily stable

[14] I am talking about pure strategies here.

strategy in the sense of Maynard Smith and G. R. Price,[15] and an attracting dynamical equilibrium of the evolutionary replicator dynamics.[16]

Fair division is thus the *unique evolutionarily stable equilibrium* of the symmetric bargaining game. Its strong stability properties guarantee that it is an attracting equilibrium in the replicator dynamics, but also make the details of that dynamics unimportant. Fair division will be stable in any dynamics with a tendency to increase the proportion (or probability) of strategies with greater payoffs, because any unilateral deviation from fair division results in a strictly worse payoff. For this reason, the Darwinian story can be transposed into the context of cultural evolution where imitation and learning may play an important role in the dynamics.

I have directed attention to symmetric bargaining problems, because it is only in situations where the roles of the players are perceived as symmetric that we have the clear intuition that justice consists in share and share alike. Here, as in the case of sex ratio, it appears that evolutionary dynamics succeeds in giving us an explanation where other approaches fail. Evolution selects from the infinity of equilibria in informed rational self-interest (the Nash equilibria) a unique evolutionarily stable equilibrium which becomes the rule or habit of just division.

IV. POLYMORPHIC PROBLEMS

If we look more deeply into the matter, however, complications arise. In the cases of both sex ratio and dividing the cake, we considered the evolutionary stability of pure strategies. We did not examine the possibility that evolution might not lead to the fixation of a pure strategy, but rather to a polymorphic state of the population where some proportion of the population plays one pure strategy and some proportion of the population plays another.

Consider the matter of sex ratio. Fisher's basic argument was that, if one sex were scarce in the population, evolution would favor production of the other. The stable equilibrium lies at equality of the sexes in the population. This could be because all individuals have the strategy to produce the sexes with equal probability. But it could just as well be true because two quite different strategies are equally represented in the population—one to produce 90% males and one to produce 90% females (or in an infinite number of other polymorphisms). These polymorphic equilibrium states, however, are not in general observed in nature. Why not?

[15] "The Logic of Animal Conflict," *Nature,* CXLVI (1973): 15–8.

[16] P. D. Taylor and L. B. Jonker, "Evolutionarily Stable Strategies and Game Dynamics," *Mathematical Biosciences,* XL (1978): 145–56.

Before attempting to answer that question, let us ask whether there are also polymorphic equilibria in the bargaining game. As soon as you look, you see that they are there in profusion. For example, suppose that half the population claims ⅔ of the cake and half the population claims ⅓. Let us call the first strategy *greedy* and the second *modest*. A greedy individual stands an equal chance of meeting another greedy individual or a modest individual. If she meets another greedy individual she gets nothing since their claims exceed the whole cake, but if she meets a modest individual she gets ⅔. Her average payoff is ⅓. A modest individual, on the other hand, gets a payoff of ⅓ no matter whom she meets.

Let us check and see if this polymorphism is a stable equilibrium. First note that, if the proportion of greedys should rise, then greedys would meet each other more often and the average payoff to greedy would fall below the ⅓ guaranteed to modest. And if the proportion of greedys should fall, the greedys would meet modests more often, and the average payoff to greedy would rise above ⅓. Negative feedback will keep the population proportions of greedy and modest at equality. But what about the invasion of other mutant strategies? Suppose that a supergreedy mutant who demands more than ⅔ arises in this population. This mutant gets payoff of zero and goes extinct. Suppose that a supermodest mutant who demands less than ⅓ arises in the population. This mutant will get what she asks for, which is less than greedy and modest get, so she will also go extinct—though more slowly than supergreedy will. The remaining possibility is that a middle-of-the-road mutant arises who asks for more than modest but less than greedy. A case of special interest is that of the fair-minded mutant who asks for exactly ½. All of these mutants would get nothing when they meet greedy and get less than greedy does when they meet modest. Thus, they will all have an average payoff of less than ⅓ and all—including our fair-minded mutant—will be driven to extinction. This polymorphism has strong stability properties.

This is unhappy news, for the population as well as for the evolution of justice, because our polymorphism is inefficient. Here everyone gets, on average, ⅓ of the cake—while ⅓ of the cake is squandered in greedy encounters. Compare this equilibrium with the pure equilibrium where everyone demands and gets ½ of the cake. In view of both the inefficiency and the strong stability properties of the ⅓–⅔ polymorphism, it appears to be a kind of trap into which the population could fall and from which it could be difficult to escape.

There are lots of such polymorphic traps. For any number, x, between zero and one, there is a polymorphism of the two strategies

demand x, demand 1 − x, which is a stable equilibrium in the same sense and by essentially the same reasoning as in our example. As the greedy end of the polymorphism becomes more greedy and the modest end more modest, the greedys become more numerous and the average fitness of the population decreases. For instance, in the polymorphic equilibrium of ultragreedy individuals demanding 99% of the cake and ultramodest individuals demanding 1%, the ultragreedies have taken over 98/99% of the population and the average payoff has dropped to .01. This disagreeable state is, nevertheless, a strongly stable equilibrium.

The existence of polymorphic traps does not make the situation hopeless, however. As a little experiment, you could suppose that the cake is already cut into ten pieces, and then players can claim any number of pieces. Now we have a tractable finite game, and we can start all the possible strategies off with equal probability and program a computer to evolve the system according to the evolutionary dynamics (the replicator dynamics). If you do this, you will see the most extreme strategies dying off most rapidly, and the strategy of half of the cake eventually taking over the entire population.

We would like to know how probable it is that a population would evolve to the rule of share and share alike, and how probable it is that it will slip into a polymorphic trap. In order to begin to answer these questions, we need to look more closely at the evolutionary dynamics. It is not simply the existence and stability of equilibria that are of interest here, but also what initial population proportions lead to what equilibria. The magnitude of the danger posed by the polymorphic pitfalls depends on the size of their basins of attraction. As an illustration, consider the simpler bargaining game in which there are only three possible strategies: demand ⅓, demand ⅔, demand ½.

The global dynamical picture (under the replicator dynamics) is illustrated in figure 1. Each vertex of the triangle corresponds to 100% of the population playing the corresponding strategy—where S1 = demand ⅓; S2 = demand ⅔; S3 = demand ½. A point in the interior is the point at which the triangle would balance if weights corresponding to the fractions of the population playing the strategies were put at the vertices. There is an unstable polymorphism involving all three strategies where S1 comprises half of the population, S2 a third, and S3 a sixth. There is an attraction toward an equal division of the whole population between S1 and S2, and another toward universality of S3. It is clear that the basin of attraction for S3 (equal division) is substantially larger than that for the attracting polymorphism; but the region that leads to the polymorphism is far from negligible.

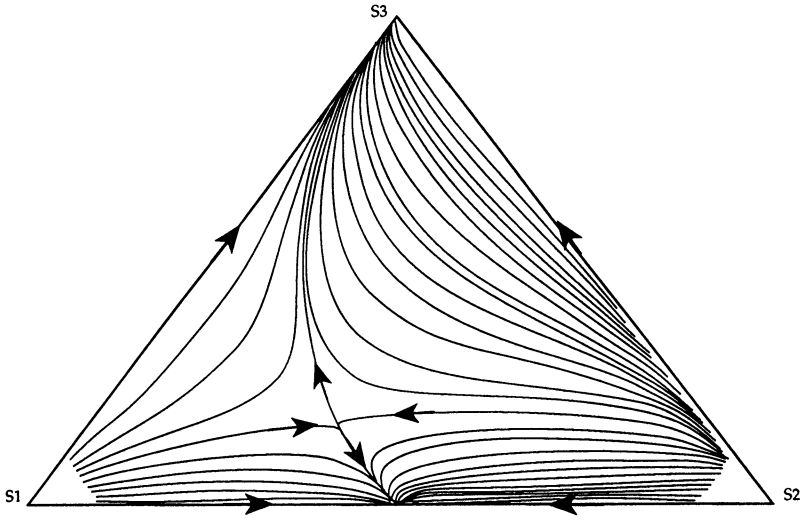

Figure 1

If the basin of attraction of equal division is large relative to that of the polymorphisms, then one can say that justice will evolve from a larger set of initial conditions than will injustice. If chance mutations are added to the dynamic model, this would mean that in the long run, a population would spend most of its time observing the convention of fair division. The latter conclusion—and much more—has recently been established analytically.[17] Still, we might hope for more. Is there some important element that has been left out of our analysis?

V. AVOIDING POLYMORPHIC TRAPS

In some ways, the equilibrium with each individual tending to produce offspring at the 1-to-1 sex ratio is more unstable than the corresponding share-and-share alike equilibrium of the bargaining game. If the population sex ratio were to drift a little to the male side, then the optimum response for an individual would be to produce all females; if it were to drift a little to the female side, then the optimum individual response would be to produce all males. The greater fitness of extreme responses should generate a tendency

[17] See D. Foster and P. Young, "Stochastic Evolutionary Game Dynamics," *Theoretical Population Biology*, XXXVIII (1990): 219–32; H. P. Young, "An Evolutionary Model of Bargaining," *Journal of Economic Theory*, LIX (1993): 145–68, and "The Evolution of Conventions," *Econometrica*, LXI (1993): 57–94; M. Kandori, G. Mailath, and R. Rob, "Learning, Mutation and Long-Run Equilibria in Games," *Econometrica*, LXI (1993): 29–56.

toward polymorphic populations. Such sex-ratio polymorphisms are rarely observed in nature, however.[18] Why not?

There is surprisingly little discussion of this question in the biological literature. One idea, due to J. Verner,[19] is that, if individuals mate within small local groups and the sex ratios of these groups fluctuate, then individuals with a 1-to-1 individual sex ratio will have higher average fitness than those with extreme individual sex ratios —even though the population sex ratio remains at equality. This is because a strategy with, for example, female bias gains less in fluctuations of the local group proportions toward the male than it loses during local group fluctuations toward the female.

Selection for individual sex ratio of 1-to-1 would be even stronger if we assume not only that the differences between the composition of local groups is not simply due to statistical fluctuations, but also that because of the nondispersive nature of the population, like tends to mate with like. If Georgia had a 9-to-1 female-biased sex ratio and Idaho has a 9-to-1 male-based sex ratio, it would not help if the overall sex ratio in the human population were 1-to-1. A mutant with a 1-to-1 sex ratio would prosper in either place.

Let us fix on the general point that it is the assumption of *random mating from the population* which makes the population sex ratio of prime importance, and which gives us as equilibria all the polymorphisms that produce those population proportions. If one drops the assumption of random mating then (1) the analysis becomes more complicated and (2) one of the assumptions of Fisher's original argument for an equal sex ratio has been dropped. In regard to (2), radical departures from random mating can change the predicted sex ratio. Where mating is with siblings, as in certain mites, a strongly female-based sex ratio is both predicted and observed.[20]

At this point, however, I want to abstract from some of the biological complications. Suppose that we are dealing with a case where the predicted sex ratio is near equality, but where there is some

[18] See R. Shaw, "The Theoretical Genetics of the Sex Ratio," *Genetics,* XLIII (1961): 149–63, for a theoretical genetic discussion which treats two reported cases of sex-ratio polymorphisms. One is a case of a population of isopods that have two different color patterns. The different colors had sex ratios of .68 and .32 and were represented in equal numbers in the population.

[19] "Selection for Sex Ratio," *American Naturalist,* XCIX (1965): 419–21. The idea is developed in P. Taylor and A. Sauer, "The Selective Advantage of Sex-ratio Homeostasis," *American Naturalist,* CXVI (1980): 305–10. Also, for critical discussion, see G. C. Williams, "The Question of Adaptive Sex Ratio in Outcrossed Vertebrates," *Proceedings of the Royal Society of London,* B CCV (1979): 567–80.

[20] See W. D. Hamilton, "Extraordinary Sex Ratios," *Science,* CLVI (1967): 477–88; and E. Charnov, *The Theory of Sex Allocation* (Princeton: University Press, 1982).

positive tendency to mate with like individuals. This positive correlation destabilizes the sex-ratio polymorphisms. Will a similar departure from randomness have a similar effect on the polymorphic traps on the road to the evolution of justice?

Let us return to the question of dividing the cake, and replace the assumption of random encounters with one of positive correlation between like strategies. It is evident that in the extreme case of perfect correlation, stable polymorphisms are no longer possible. Strategies that demand more than $1/2$ meet each other and get nothing, strategies that demand less than $1/2$ meet each other and get what they demand. The fittest strategy is that which demands exactly $1/2$ of the cake.

In the real world, both random meeting and perfect correlation are likely to be unrealistic assumptions. The real cases of interest lie in between. For some indication of what is possible, I shall reconsider the case of the greedy-modest polymorphism illustrated in figure 1. Remember that S1 is the modest strategy of demanding $1/3$ of the cake; S2 is the greedy strategy of demanding $2/3$; and S3 is the fair strategy of demanding exactly $1/2$. We now want to see how the dynamical picture varies when we put some positive correlation into the picture. Each type tends to interact more with itself than would be expected with random pairing. The degree of nonrandomness will be governed by a parameter, e. At $e = 0$, we have random encounters. At $e = 1$, we have perfect correlation.[21] Figure 2 shows the dynamics with $e = 1/10$. This small amount of correlation has significantly reduced the basin of attraction of the greedy-modest polymorphism to about $1/3$ the size it was with random encounters. Figure 3 shows the dynamics with $e = 2/10$. There is no longer a stable greedy-modest equilibrium. Fair dealers now have the highest expected fitness everywhere, and any mixed population will evolve to one composed of 100% fair dealers. It is not surprising that correlation has an effect, but it may be surprising so little correlation has such a big effect.

Generally, as correlation increases the basins of attraction of the polymorphic traps decrease and the more inefficient polymorphisms

[21] This is a very simple model used for a quick test of the effects that can be generated by positively correlated encounters. The probability of a strategy meeting itself, $p(S_i|S_i)$, is inflated thus:

$$p(S_i|S_i) = p(S_i) + ep(\text{Not} - S_i)$$

while the probability of strategy S_i meeting a different strategy S_j is deflated:

$$p(S_j|S_i) = p(S_j) - ep(S_j).$$

If $e = 0$ encounters are uncorrelated, if $e = 1$ encounters are perfectly correlated.

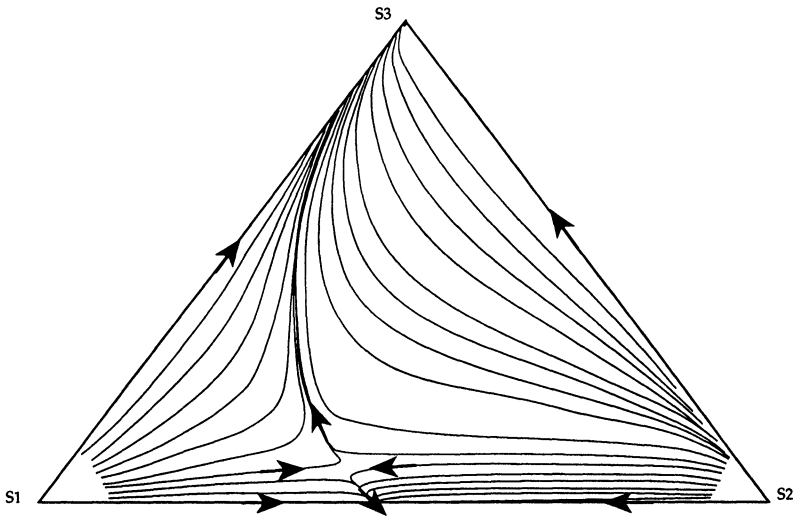

Figure 2

cease to be attractors at all. In the limiting case of perfect correlation, the just population—where everyone respects equity—is the unique stable equilibrium.

VI. THE EVOLUTION OF JUSTICE

Taking stock, what can we say about the origin of the habit of equal division in the problem of dividing the cake? Our evolutionary analy-

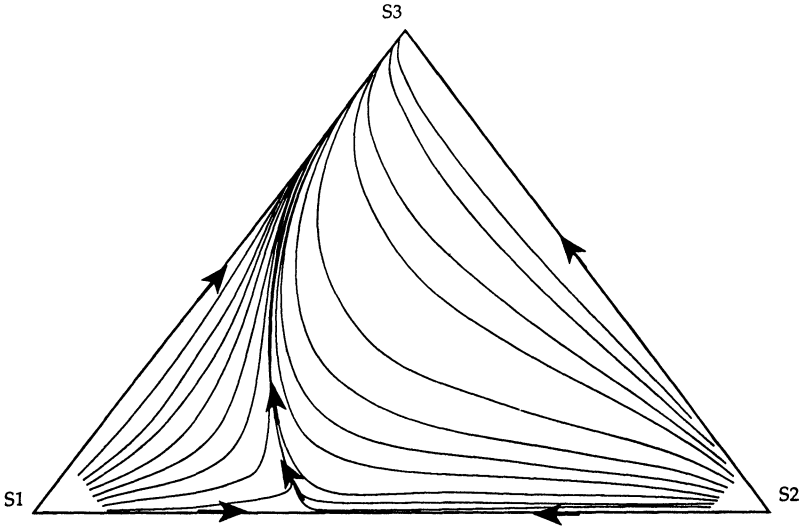

Figure 3

sis does not yield the Panglossian proposition that perfect justice must evolve. But it does show us some things that are missed by other approaches. The concept of equilibrium in informed rational self-interest—the Nash equilibrium concept of classical game theory —left us with an infinite number of pure equilibrium strategies. The evolutionary approach leaves us with one evolutionarily stable pure strategy—the strategy of share and share alike. This selection of a unique equilibrium strategy is a consequence of the evolutionary process proceeding under the Darwinian veil of ignorance. In this way, the evolutionary account makes contact with, and supplements, the veil-of-ignorance theories of Harsanyi and Rawls.

Nevertheless, a closer look at the evolutionary dynamics shows that a population need not evolve to a state where everyone plays the unique evolutionarily stable strategy of fair division. There are stable mixed states of the population, where different proportions of the population use different strategies. These *polymorphic pitfalls* are attractors that may capture a population that starts in a favorable initial state. If there is enough random variation in the evolutionary process, a population caught in a polymorphic pitfall will eventually bounce out of it, and proceed to the fair-division equilibrium. It will also eventually bounce out of the fair-division equilibrium as well, but the amount of time spent at fair division will be large relative to the amount of time spent in polymorphic traps, because of the larger basin of attraction of the fair-division equilibria.

So far, this is the story given by the standard evolutionary game dynamics which assumes random pairing of individuals. If there is some tendency, for whatever reason, for like-minded individuals to interact with each other then the prospects for the evolution of justice are improved. In the extreme case of perfect correlation, a population state of share and share alike becomes a global attractor, and the evolution of justice is assured. (The effects of correlated pairing are of interest in other kinds of interactions as well. I pursue the question of correlation in evolutionary game theory elsewhere.[22])

In a finite population, in a finite time, where there is some random element in evolution and some correlation, we can say roughly 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.

BRIAN SKYRMS

University of California/Irvine

[22] See my "Darwin Meets *The Logic of Decision:* Correlation in Evolutionary Game Theory," in *Philosophy of Science* (forthcoming).