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

Local Fairness

CRISTINA BICCHIERI
Carnegie Mellon University

Social psychologists that study behavior in social dilemmas, as well as experimental game theorists that look at ultimatum and dictator games, report results that are consistently at odds with the predictions of game-theoretic models. Cooperation in one-shot prisoner's dilemmas or punishment in ultimatum games are non-maximizing behaviors, because individuals choose an action that gets them a monetary outcome inferior to what they would get for sure by choosing a different action.

Brian Skyrms's book is a passionate defense of a different approach, one that assumes that agents' choices are simply guided by a fitness criterion, where fitness means reproductive success. Depending on the environment being considered, reproductive success may be measured in terms of number of offspring, market share or by the spread of a given cultural trait or behavior in a population. Since Skyrms's book deals with the evolution of norms (in particular, norms of justice), I shall evaluate the advantages and drawbacks of his approach with respect to the dynamics of social norms only.

By 'social norm' I mean a regular, observable pattern of behavior occurring in a group (or a population) that is supported by mutual expectations of conformity and a conditional preference for conformity (given those expectations) on the part of group members. Though social norms are a central concern in sociology, social psychology and anthropology, accounts of their emergence are scanty, and even scantier are descriptions of how they may become stable, spread to other groups, change or disappear. If we take the results of experiments on, say, ultimatum games to point to the working of norms of fairness, then, Skyrms argues, not only is traditional game theory ill-equipped to account for norm-guided behavior, but it is also conspicuously silent about why such norms, as opposed to others, became widespread.

Note that I am referring here to two very different *explananda*. One thing we want to explain is regularities in human behavior in a host of experimental (and in vivo) circumstances. If rational choice explanations fail, we may invoke norms, social identity effects or cognitive and emotional biases. Suppose we do succeed in explaining certain behavioral regularities as the product of normative pressures. It remains to be explained how and why such

norms have emerged and what makes them so resilient or, in other words, what induces people to conform to them. A central thesis of Skyrms's book is that evolutionary game-theoretic explanations of experimental results are better than the usual game-theoretic ones. His thesis presupposes that we have already concluded that people's cooperative behavior in experimental prisoner's dilemmas, as well as their remarkably consistent punishing behavior in ultimatum games, is norm-driven. It only remains to be explained why we have developed norms of reciprocity and fairness. In what follows I shall briefly review some of the experimental results, and discuss whether norm-based explanations are satisfactory.

The structure of ultimatum games is rather simple (Camerer and Thaler 1995). Two people must split a fixed amount of money M according to the following rules: a proposer (P) offers a division of M to a responder (R). If R accepts the proposed amount, they split the money as proposed; if R refuses, they both get nothing. The outcome of rational, self-interested behavior is obvious: P should not offer R more than the minimum (which may mean just a penny), and R should accept any offer equal or superior to the minimum. In most experiments, P's offer 40–50 percent of the sum, and R's typically reject offers of less than 20 percent. These results are quite robust with respect to variations in the amount of money that is being split, and cultural differences. We know that raising the stake from $10 to $100 does not decrease the frequency of rejections of low offers (those between 10 and 20 dollars), and that in experiments run in Slovenia, Pittsburgh, Israel and Tokyo the modal offers were in the range of 40 to 50 percent (Hoffman et al. 1998; Roth et al. 1991).

The uniformity of respondents' behavior suggests that people do not like being treated unfairly. That is, if subjects perceive an offer of 20 or 30 percent of the money as unfair, they may reject it to "punish" the greedy proposer, even at a cost to themselves. However, if individuals are told that offers are generated by a random device, or they believe the proposer was constrained in her decision, they are willing to accept lower offers. It appears that, if an unequal outcome can be justified, people are willing to put up with very little. We all know from experience that sacrifices are easier to bear if they are shared, or at least appear unavoidable. This knowledge is often exploited to the advantage of one of the parties in bargaining. For example, a common tactic in wage bargaining between unions and management when cuts are needed is for the management to impute losses to market conditions, since sacrifices to make up for losses due to managerial ineptitude are not likely to be accepted.

Lower offers are also accepted whenever responders are informed that the proponent offered a smaller amount to another responder in another ultimatum game. In this case, what seems to matter is the comparison between one's outcome and the outcome of other individuals similarly situated. A

great deal of experimental research on social justice supports the finding that the judgments that people make about whether they are being treated fairly derive not from the actual value of their outcomes, but from comparisons between what they have and what they expected to have. If I have a 10 percent salary rise, but I expected 20 percent, I will feel unjustly treated. If I expected less, I will feel favored. Expected outcomes, in turn, are often determined by social comparison between oneself and similarly situated others. If others get more than I do, I feel deprived. If everyone else is equally deprived, I do not feel so bad. Thus one's sense of what is fair depends on comparing one's outcome with those of other people similarly situated. This phenomenon occurs at a collective level, too. In a well known study done in the sixties, Runciman (1966) found that English manual workers felt more resentment if other manual workers' incomes exceeded their own, but were less concerned about the incomes of non-manual workers. Similarly, philosophy professors are usually not affected by the higher salaries of business school faculty: since the latter have alternative opportunities for non-academic employment, whereas philosophers do not, the salary differential is perceived as market-driven, and thus (at least in the US) not unfair.

In sum, the outcomes of experiments on Ultimatum games show that (i) individuals' assessment of what constitutes fair treatment is influenced by comparisons with similarly situated individuals. Whenever available, such comparisons may mitigate (or exacerbate) the reaction to uneven outcomes. (ii) Intentions matter. We determine the fairness of others' actions according to their motives, not just according to the action taken. One can infer another's motives both by the action chosen, as well as by the actions that were not chosen, but could have been chosen. In the original Ultimatum game, the proposer receives what amounts to a monetary *gift* from the experimenter. As a consequence, P is perceived as having no special right to the money, and is expected to share it with the responder. A norm of fairness is activated that dictates an equal split, and the proposer who is offering little is perceived as mean and unfair, and therefore gets punished. (iii) Norms of fairness are local and context-dependent. By giving participants a reason to expect behavior appropriate to the situation, a context simultaneously gives a clue as to the proposer's intention, whenever the offer is different from what is reasonably expected in that context. Thus R's accept (and presumably expect) less if P's and R's are labeled "sellers" and "buyers", as in this case the context is perceived to be market-like, and thus competitive. In yet another variation, if R is made to believe that P has "earned" the money in one way or another, fairness dictates that P is entitled to the money. Indeed, in such circumstances even very low offers are accepted.[1] Interestingly enough, there

[1] Kahneman et al. (1986) describe how different norms of fairness may apply to different contexts, including those in which unfair behavior is accepted and "excused".

seems to be substantial agreement—at least within a culture's boundaries—about fair allocations or distributions of particular goods. Often, when conflicts arise, they are due to different *interpretations* of the situation eliciting different norms; for example, it is often observed that groups with conflicting interests try to impose interpretations that allow each group to benefit from the application of a particular norm.

The results of Ultimatum games can be better explained by the activation of norms of fairness, as opposed to explanations in terms of psychological propensities such as altruism or fair-mindedness. If generous offers were the result of altruistic motives, we would observe them in a variety of circumstances, as altruism is rather insensitive to considerations of entitlements. Equal division, however, is just one among several possible offers. It is expected when the situation is perceived as one in which the proposer has received a sum of money to be divided with the responder, but the proposer has no particular entitlement to it. Furthermore, whereas altruistic behavior should occur whenever circumstances call for it, the existence of a norm does not imply that it will be followed in every occasion in which the norm would apply. Norms, in other words, are *context-sensitive*. There is ample evidence, especially in the anthropological literature, that norm-compliance is subject to variations and is sensitive to the presence or absence of material or emotional sanctions, as well as to cues about the expectations of other people. I do not want to imply that sanctions are crucial to norm-abiding behavior. They may just reinforce a tendency to obey the norm, and serve the function—together with several other indicators—of focusing individuals' attention on the particular norm that applies to the situation.

A norm-based explanation entails the prediction that proposers will choose an equal split whenever they expect negative consequences from unfair offers, but does not make a definite prediction whenever a deviation from the norm has no cost (and benefits the deviant) or, more generally, when the context presents no cues that would force an agent to *focus* upon the relevant norm. Camerer and Thaler's (1995) discussion of Ultimatum game experiments with asymmetric information is particularly interesting in this respect. In one variant of the experiments discussed, proposers were given 100 chips worth 30 cents each, but the same chips were worth only 10 cents each when owned by the responders. Responders knew the value of the chips for them, but did not know the proposers' value. Proposers knew both values. In this case, an equal division of money would have meant an offer of 75 percent of the chips to R's. Instead, offers were close to 50 percent: an unfair offer, but one that could look fair, and would almost certainly be accepted.

Furthermore, for an explanation in terms of altruism or fair-mindedness to hold, we should observe roughly the same percentage generous offers to occur in all those circumstances in which one of the parties has all the power, i.e. whenever one player receives a sum of money and may give either some, all

or nothing to a second player, who has to accept whatever the first decides. In so-called experimental Dictator games, the modal offer is the game-theoretic equilibrium offer at which the allocator keeps all the money to himself, but a high concentration of offers at equal division can still be observed (Forsythe et al. 1994). In the absence of the sanctioning mechanism provided by the possibility of rejection we observe less generosity, but it is still remarkable that many individuals—in a situation of complete power—are willing to pass some of the money to a stranger. In some one-shot Dictator game experiments, under conditions of anonymity, the allocators offered on average a quarter of the sum (Frey and Bohnet 1995). Again, if generous behavior were due to fair-mindedness, we would expect the percentage of offers to remain constant across experimental conditions. On the contrary, experimental variations in which the players are visually identified, or can communicate with each other, elicit much more generous offers. In fact, whenever players are allowed to look at each other (without communication), or when they can talk to each other, offers of half the money are the norm. However, the circumstances here are very different from those of an Ultimatum game. In the latter, activation of a fairness norm is immediate: P is asked to make an offer (which might be rejected), and is somewhat forced to focus on what is expected of her in such circumstances. The possibility of monetary sanctions, brought about by violation of expectations of a fair share, reinforces norm-abiding behavior. No such sanctions are present in the Dictator game. When the game is one-shot and the other player is anonymous, a fairness norm may have little salience. When players talk to each other, however, or even just look at each other in a face-to-face encounter, there occurs a cognitive and emotional shift of attention. One is forced to consider the possibility of sharing the money, and this consideration might be accompanied by feelings of guilt that sustain norm-congruent behavior. Another example of how fairness norms can be activated by communication are one-shot experiments on social dilemmas where subjects are allowed to communicate. In such circumstances cooperative behavior is pervasive, since a norm of fairness in these contexts implies that all participants equally share the cost of providing a public good (Caporael et al. 1989). In sum, there is abundant evidence that the circumstances of interaction play a crucial role in making a norm more or less salient for the individual decisionmaker, whereas the application of a norm seems to be contingent on the situation.

The question I asked at the outset, whether experimental behavior can be explained as norm-driven, can be answered affirmatively, albeit not without some major qualifications. There is no unique norm of fairness, but several, and the context determines which of them is appropriate. There is continuity between real life and experiments with respect to how 'rights' and 'entitlements', considerations of merit, desert or sheer luck shape our perception of what is fair. Cultures differ in their reliance on different allocative and

distributive rules, since such rules depend on different forms of social organization. Within a given culture, however, there usually is a broad consensus about how different goods and opportunities should be allocated or distributed. The allocation of grants and scholarships is expected to be merit-driven, whereas the allocation of a liver or a kidney should never depend—most people feel—upon the merits or the ability to pay of the recipient. To most people, an auction is an acceptable allocation mechanism if the desired object is, say, a rare miniature; it is unacceptable if it is the last vial of a life-saving medicine.

Is a rational choice model really powerless in explaining the phenomena I have just described? Rational choice should not be equated with particular motives (selfish) or utilities (non-separable, linear in money, etc.). If other people's actions or the circumstances of interaction trigger social values or norms, and the expectations and feelings that support and accompany them, agents' utilities should reflect it. One's utility may not be fixed, but rather determined by the situation one faces. We might be wired to feel and prefer different things according to the context we are in. Whether these propensities and capabilities are the product of social learning or are genetically inherited from our hunter and gatherer ancestors is an open question. What matters to our discussion is the possibility of embedding these considerations in a rational choice model. Rabin's (1993) fairness equilibrium concept is a good example of an attempt to build more sophisticated, powerful rational choice models along these lines. In his model, agents have separable utility functions that take into account other agents' payoffs and the feelings of sympathy or hostility, loyalty or mistrust elicited by the circumstances of interaction. As a result, his model predicts rejections in Ultimatum games whenever the offer is perceived as "mean", as well as cooperation in social dilemmas whenever reciprocation and trust are elicited.[2]

Questions about the emergence of macro-level phenomena such as social norms are best answered by micro-explanations. To be convincing, such explanations need to construct models of the micro-to-macro process that make use of the micro-level data we possess about individual behavior. To make sense of the data, models will need more cognitive detail and greater psychological sophistication. For example, I have recently shown how the combination of certain common and well documented cognitive biases and conformist preferences leads (rational) individuals to develop and subsequently stick to social norms they dislike (Bicchieri and Fukui 1999). Unpopular and inefficient social norms are quite common, but it is not sufficient to say they are the "unintended effects" of social interactions to explain their existence.

---

2 In Rabin's model, there is a trade-off between feelings and monetary payoffs: as the latter rise, feelings become less important. This is a reasonable assumption, though different people will have different thresholds.

We have to provide a detailed mechanism that links individual actions to the macro-level phenomena they generate and that is consistent with the data we have about individuals' preferences and cognitive biases.

Skyrms's replicator dynamics model should not be conceived as an alternative explanation of phenomena such as the emergence of fairness norms. An evolutionary explanation cannot tell us why new norms emerge, or account for their locality and context-dependence. It does not explain how learning occurs and transfers across contexts, so that we not only learn rules and norms, but are also able to generalize them to new cases. Furthermore, an evolutionary model presupposes that individuals only care about other individuals' actions and how they affect one's payoffs. But both real life and experimental knowledge point to the fact that we do care about other people's payoffs, as well as their intentions. A macro-explanation however becomes necessary when we move from emergence to dynamics, in particular population dynamics. The continuous stability of a social norm, its spread outside the boundaries of the community in which it originated, as well as its gradual or sudden disappearance are part of what I call a norm's dynamics. Whereas the dynamics of small-group norms can be accounted for in terms of micro-models, large, anonymous groups present us with very different problems. Strategic behavior makes no sense in these contexts, nor may we suppose that people care about other people's outcomes and intentions, since they are not directly observable nor inferable. Here adaptive, unsophisticated behavior may work to the individual's advantage.

There are circumstances in which we may think of norms as institutions that compete for survival. This often happens when different cultures and societies interact. When less economically developed societies come in contact with more advanced ones, new modes of behavior filter in, and old ways get discarded. For example, norms against hoarding, insurance, interest and middleman activities have been abandoned by societies that experienced rapid economic growth. Adherence to such norms puts local firms at a disadvantage in international markets, and imitation of more successful competitors leads to the sometimes very rapid displacement of norms that had lasted unchallenged for centuries. Can we tell a similar story for norms of fairness? According to Skyrms's account, we know that under certain conditions a norm of equal division will thrive. One such condition is correlation of encounters, so that individuals tend to cluster and interact with similarly situated others. A common culture is a form of clustering: we all share the same norms, and expect to interact with people that will not surprise us by displaying odd manners. Within a culture, we can aim at good micro-explanations of why some norms survive, others die out. Across cultures, the picture gets more complicated. The norms of revenge typical of Sicilian culture have survived in the American Mafia, but they are restricted to regulating behavior among Mafia members; they are never applied to the interactions of this

group with society at large. The American Mafiosi have succeeded in preserving an extreme form of clustering, which allows their norms continued existence. American Indians were not so successful, perhaps because the forms of social organization they had developed were not well-suited at shielding at least some of their customs from the "invasion" of different ones. Norms of fairness may share the same fate. The challenge we face is precisely one of integrating more sophisticated micro-explanations about how local norms have developed and why they persist or change within a given culture with a macro-model describing what may happen when cultures, and societies, interact.

References

C. Bicchieri and Y. Fukui (1999), "The great illusion: Ignorance, informational cascades and the persistence of unpopular norms". *Business Ethics Quarterly* 9: 127–55.

C. Camerer and R. Thaler (1995), "Anomalies : Ultimatums, dictators and manners". *Journal of Economic Perspectives* 9: 209–19.

L. Caporael, R. Dawes, J. Orbell and A. van de Kragt (1989), "Selfishness examined: Cooperation in the absence of egoistic incentives". *Behavioral and Brain Sciences* 12: 683–739.

B. Frey and I. Bohnet (1995), "Institutions affect fairness: Experimental investigations". *Journal of Institutional and Theoretical Economics* 151/2: 286–303.

E. Hoffman, K. McCabe and V. Smith (1998), "On expectations and the monetary stakes in Ultimatum games". *International Journal of Game Theory* (in press).

D. Kahneman, J. Knetsch and R. Thaler (1986), "Fairness as a constraint on profit seeking: Entitlements in the market". *American Economic Review* 76: 728–41.

M. Rabin (1993), "Incorporating fairness into game theory and economics". *American Economic Review* 83: 1281–302.

A. Roth, V. Prasnikar, S. Zamir and M. Okuno-Fujiwara (1991), "Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh and Tokyo: An experimental study". *American Economic Review* 81: 1068–95.

W. Runciman (1966), *Relative deprivation and social justice*. London: Routledge & Kegan Paul.