MODERN PIONEERS IN PSYCHOLOGICAL SCIENCE: AN APS-PSYCHOLOGY PRESS SERIES

This series celebrates the careers and contributions of a generation of pioneers in psychological science. Based on the proceedings of day-long Festschrift events at the annual meeting of the Association for Psychological Science, each volume commemorates the research and life of an exceptionally influential scientist. These books document the professional and personal milestones that have shaped the frontiers of progress across a variety of areas, from theoretical discoveries to innovative applications and from experimental psychology to clinical research. The unifying element among the individuals and books in this series is a commitment to science as the key to understanding and improving the human condition.

PUBLISHED TITLES

1. *Psychological Clinical Science: Papers in Honor of Richard M. McFall*, edited by Teresa A. Treat, Richard R. Bootzin, and Timothy B. Baker (2007).
2. *Rationality and Social Responsibility: Essays in Honor of Robyn Mason Dawes*, edited by Joachim I. Krueger (2008).

RATIONALITY AND SOCIAL RESPONSIBILITY

Essays in Honor of Robyn Mason Dawes

Edited by

Joachim I. Krueger

aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychology Press
Taylor & Francis Group
New York London

9

HOW EXPECTATIONS AFFECT BEHAVIOR

Fairness Preferences or Fairness Norms?

Cristina Bicchieri

University of Pennsylvania

Since its origin, philosophy has been concerned with fairness: how to define it, how to justify our intuitions about it, and how to lend consistency to the multiplicity of meanings fairness usually takes. For the philosopher, what is at stake is the normativity of our moral judgments, what can possibly ground their "ought" claims. In a dialogue that has lasted more than 15 years, I, the philosopher, and Robyn Dawes, the psychologist, have explored the why and how of many behaviors that we would normally call ethical: cooperation and reciprocity, fairness, benevolence, and altruism. We have had many discussions about how much such behavior is sensitive to the decision context, and the crucial role expectations pay in our assessment of a given situation.

Understanding how people form fairness judgments, the cognitive dynamics involved in the process, and what drives fair behavior on one occasion and dampens it in another are important steps that any philosopher should take in the direction of building better normative theories. Naturalizing ethics does not mean reducing what ought to be done to what is done: This would be a trivial naturalistic fallacy misstep. What instead needs to be done is to build our normative theories upon the solid foundation of what we know individuals *can* do, and this is a whole different project. I embarked on this project long ago by trying to show that our ethical norms are just collectively defined and supported social norms. Some such norms are more entrenched than

others, but the cognitive processes underlying norm-following, and the biases we all face in filtering and processing the social information that will ultimately decide whether or not we act in a prosocial way, are essentially the same. Without knowledge of such cognitive processes, and the behaviors they engender, ethics is condemned to remain an abstract and fairly useless endeavor.

In what follows I will concentrate upon some experimental results that show what appears to be individuals' disposition to behave in a fair manner in a variety of circumstances. One common explanation is that individuals have a *preference* for fairness. The alternative explanation I propose is that, in the right kind of circumstances, individuals *obey fairness norms*. To say that we obey fairness norms differs from assuming that we have a preference for fairness (Bicchieri 2000, 2006). To follow a fairness norm, we must have the right kind of expectations. We must expect others to follow the norm too and believe that there is a generalized expectation that we will obey the norm in the present circumstances. The preference to obey a norm is *conditional* upon such expectations.[1] Take away some of the expectations, and behavior will significantly change. A conditional preference will thus be stable under certain conditions, but a change in the relevant conditions may induce a predictable preference shift. The predictions of a norm-based theory are thus testable and quite different, at least in some critical instances, from the predictions of theories that postulate a social preference for fairness.

When economists postulate fairness preferences, they make two related, important assumptions. The first is that what matters to an agent is the final distribution, not the way the distribution came about (Falk, Fehr, & Fischbacher, 1999): this is a *consequentialist* assumption. The second assumption is that preferences are stable. Both assumptions are easy to test. When falsified, however, it is less clear who the culprit is. For example, if a person has a stable preference for fair outcomes, we would expect his or her cross-situational behavior to be consistent and insensitive to the circumstances surrounding the specific distributive situation. Whether you are the proposer in an ultimatum or a dictator game should not matter to your choice of how much money to give to a responder. Similarly, information about *who* the proposer is—a real person or a random device—should not have an effect on one's propensity to accept or reject its offer. What is observed instead is cross-situational inconsistency. The reason for this inconsistency is not obvious. It is possible that people do care about how a distribution came about and that the process itself matters. For example, we might accept an unequal share of the pie if it comes from a lottery but reject it if it results from an auction. Preferences could still be assumed to be stable, but in this case what we prefer is a combination of goods and processes to distribute and allocate those goods. On the other hand, preferences may be highly context dependent. Change the context, or the context's description, and there is a noticeable preference shift. In the latter case, however, making any prediction would require a mapping from contexts to preferences. No such mapping has ever been provided.

In what follows I will examine two of the most common games studied by experimental economists. Ultimatum and dictator games come in many flavors and variants, but the simplest, bare versions of both games are in some sense ideal, because they offer a very simplified allocation problem. The good to be allocated (or divided) is money, and the situation is such that most familiar contextual clues are removed. The results of such experiments consistently defy the predictions of traditional rational choice models. Agents are clearly not solely concerned with their monetary payoffs: They care about what other agents get and how they get it. The big challenge has been to enrich traditional rational choice models in such a way that they can explain (and predict) behavior that is not just motivated by material incentives in a variety of realistic contexts. I will compare one of the most interesting and influential new models with my norm-based approach and show that the hypothesis that people obey fairness norms offers a more satisfactory explanation for the phenomena we observe. Where my predictions differ from those of the alternative, social preference model, the data seem to vindicate my model. However, we need many more experiments to test the effects that manipulating expectations (and thus norm compliance) has on behavior.

THE ULTIMATUM GAME

In 1982 Guth, Schmittberger, and Schwarze published a seminal study in which they asked subjects to play what is now known as an ultimatum bargaining game. Their goal was to test the predictions of game theory about equilibrium behavior. Their results instead showed that subjects consistently deviate from what game theory predicts. To understand what game theory predicts, and why, let us look at a typical ultimatum game (Figure 9.1).

The structure of this game is fairly simple. Two people must split a fixed amount of money M according to the following rules: The proposer (P) moves first and offers a division of M to the responder (R), where the offer can range between M and zero. The responder has a binary choice in each case: to accept the offer or to reject it. If the offer is accepted, the proposer receives $M - x$, and the responder receives x,

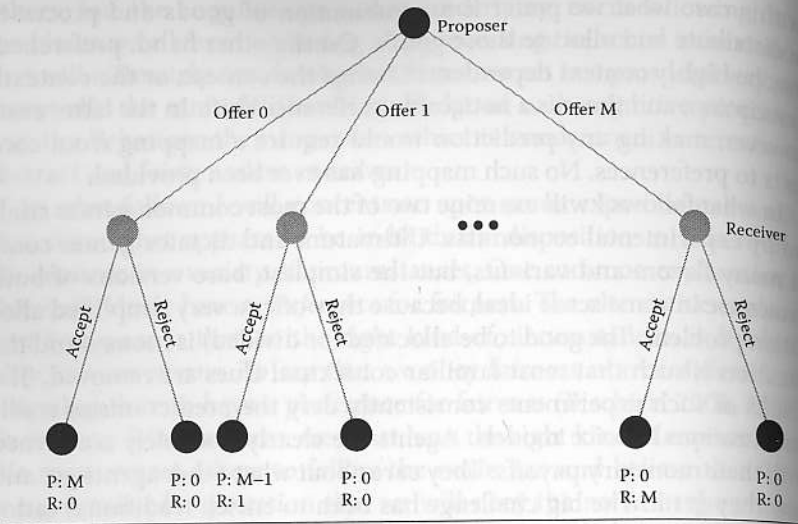

Figure 9.1 Ultimatum game.

where x is the offer amount. If the offer is rejected, each player receives nothing. If rationality is common knowledge, the proposer knows that the responder will always accept any amount greater than zero, because accept dominates reject for *any* offer greater than zero. Hence P should offer the minimum amount guaranteed to be accepted, and R will accept it. For example, if M = \$10 and the minimum available amount is 1 cent, the proposer should offer it, and the offer should be accepted, leaving the proposer with \$9.99.

Experiments find, however, that nobody offers 1 cent or even 1 dollar. Note that such experiments are always one-shot and anonymous. That is, subjects play the game only once with an anonymous partner and are guaranteed that their choice will not be disclosed. The absence of repetition is important to distinguish between generous behavior that is dictated by a rational, selfish calculation and genuine generosity. If an ultimatum game is repeated with the same partner, or if players suspect that future partners will know of their past behavior, it may be perfectly rational for players who are only interested in their material payoff to give generously if they expect to be on the receiving side at a future time. On the other hand, a receiver who might accept the minimum in a one-shot game might want to reject a low offer at the beginning of a repeated game, in the hope of convincing future proposers to offer more.

In the United States, as well as in a number of other countries, the modal and median offers in one-shot experimental games are 40 to

50% of the total amount, and the mean offers are 30 to 40%. Offers below 20% are rejected about half the time.[2] These results are robust with respect to variations in the amount of money that is being split and cultural differences (Camerer, 2003). For example, 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) and that in experiments run in Slovenia, Pittsburgh, Israel, and Tokyo the modal offers were in the range of 40 to 50% (Hoffman, McCabe, & Smith, 1998; Roth, Prasnikar, Okuno-Fujiwara, & Zamir, 1991).

If by rationality we mean that subjects maximize expected utility *and* that they only value their monetary outcomes, then we must conclude that a subject who rejects a nonzero offer is acting irrationally. However, individuals' behavior across games suggests that money is not the sole consideration, and instead there is a concern for fairness, so much so that subjects are prepared to punish at a cost to themselves those who behave in inequitable ways. A concern for fairness is just one example of a more general fact about human behavior: We are often motivated by a host of factors, of which monetary incentives are one, and often not the most important. We act out of love, envy, spite, generosity, desire to imitate, sympathy, or hatred, to name just a few of the passions and desires that move us to act. When faced with different possible distributions, we usually care about how we fare with respect to others, how the distribution came about, who implemented it, and why. Experiment after experiment has demonstrated that individuals care about others' payoffs, that they may want to spend resources to increase or decrease such payoffs, and that what they perceive to be the (good or bad) intentions of those they interact with weigh in their decisions. Unfortunately, the default utility function in game theory is a narrowly selfish one: It is selfish because it depicts people who care only about their own outcomes, and it is narrow because motivations like altruism, benevolence, guilt, envy, or hatred are kept out of the picture. Such motives, however, can and should be incorporated into a utility function, and economists have recently started to develop richer, more complex models of human behavior that try to explain what we have always known: People care about other people's outcomes. Thus, a better way to explain what is observed in experiments (and real life) is to provide a richer definition of rationality: People still maximize their utilities, but the arguments of their utility functions include other people's utilities.

The obvious risk of such models is their "ad hocness": One may easily explain any data by adjusting the utility function to reflect what looks like envy or altruism or a preference for equal shares. What we need

are utility functions that are general enough to subsume many different experimental phenomena and specific enough to make falsifiable predictions. In what follows I will look at some possible explanations for the generous distributions we observe in ultimatum games and test these explanations against some interesting variations of the game. Such testing is not always easy to conduct. The problem is that we still have quite rudimentary theories of how motives affect behavior, and to test a hypothesis about what sort of motives induce us to act one way or another, we have to be very specific in defining such motives and the ways in which they influence our choices. Let me clarify this statement with an example.

Observing the results of ultimatum games, someone might argue that subjects in the proposer's role are behaving altruistically. Others would deny that, saying that people like to give because of the "warm glow" their actions induce in them (Andreoni, 1990), and yet others would say that what we observe is just benevolence—nothing else. To make sense at all, such concepts need to be made as specific as possible, and operational. Take, for example, a distribution (x_1, x_2) of money between two people. Being an altruist would mean that 1's utility is an increasing function of 2's utility, that is, $U_1 = f(x_2)$ and $\delta U_1/\delta x_2 > 0$. Thus, a true altruist would not care about his own share, but he would only care about how much the other gets (and the more, the better). A proposer who is a pure altruist would "donate" all the money to the responder, provided he believes the responder only cares about money. Being benevolent instead means that one cares about one's own payoff *and* the other's, that is, $U_1 = f(x_1, x_2)$. In this case, the first partial derivatives of $U_1 = f(x_1, x_2)$ with respect to x_1, x_2 are strictly positive, meaning that the utility of a benevolent player 1 increases as the utility of player 2 increases. Depending on a player's degree of benevolence, the proposers will turn out to be more or less generous, but a benevolent attitude on the part of the proposers might explain, prima facie, the results of experimental ultimatum games.

The results of typical ultimatum games eliminate the pure altruist hypothesis, because people almost never give more than 50%, but do not eliminate the benevolence hypothesis. If benevolence is a stable character disposition, however, we would expect a certain behavioral stability or consistency in any situation in which a benevolent proposer has to offer a division of money to an anonymous responder.

A variant of the ultimatum game is the dictator game, in which the proposers receive a sum of money from the experimenter and decide to split the money any way they choose; the proposer's decision is final in that the responder cannot reject whatever is offered. If we hypothesize that the ultimatum game results reveal that a certain percentage of the population has a benevolent disposition, we should expect to observe roughly the same percentage of generous offers in all those circumstances in which one of the parties, the proposer, is all powerful. In most of the experiments, however, the modal offer is one in which the proposer keeps all the money, and in double-blind experiments 64% of the participants give nothing. Still, it must be mentioned that although the most frequent offer is zero, the mean allocation is 20% (Forsythe, Horowitz, Savin, & Sefton, 1994). These results suggest that people are not totally selfish, but it would be hard to argue they are benevolent unless we are prepared to presume that benevolence is a changeable disposition, as mutable as the circumstances that we encounter.

SOCIAL PREFERENCES

Altruism and benevolence are just two examples of *social preferences*. By *social preference* I refer to how people rank different allocations of material payoffs to self and others. If we stay with the ultimatum game as an example, we can think of other, slightly more complex ways to explain the results we discussed before. The uniformity of responders' behavior suggests that people do not like being treated unfairly. That is, if subjects perceive an offer of 20 or 30% of the money as unfair, they may reject it to punish the greedy proposer, even at a cost to themselves. It is important to emphasize that these experiments were all one-shot, which means the participants were fairly sure of not meeting again; therefore, punishing behavior cannot be motivated as an attempt to convince the other party to be more generous the next time around. Similarly, proposers could not be generous because they were expecting reciprocating behavior in future interactions. One possibility is to assume that both proposers and responders are showing a preference for fair outcomes or an aversion to inequality. We can thus try to explain the experimental results with a traditional rational choice model, where the agents' preferences take into account the payoffs of others.

In models of inequality aversion, players prefer both more money and more equal allocations. Though there are several models of inequality aversion, perhaps the best known and most extensively tested is the model of Fehr and Schmidt (1999). This model intends to capture the idea that people may be uneasy, to a certain extent, about the presence of inequality, even if they benefit from the unequal distribution. Given a group of L persons, the Fehr–Schmidt utility function of person i is

$$U_i(x_1,...,x_L)=x_i-\frac{\alpha_i}{L-1}\sum_j \max(x_j-x_i,0)-\frac{\beta_i}{L-1}\sum_j \max(x_i-x_j,0)$$

where x_j denotes the material payoff person j gets. α_i is a parameter that measures how much player i dislikes disadvantageous inequality (an envy weight), and β_i measures how much i dislikes advantageous inequality (a guilt weight).[3] One constraint on the parameters is that $0<\beta_i<\alpha_i$, which indicates that people dislike advantageous inequality less than disadvantageous inequality. The other constraint is $\beta_i<1$, so that an agent does not suffer terrible guilt when he or she is in a relatively good position. For example, a player would prefer getting more without affecting other people's payoffs even though that results in an increase of the inequality.

Applying the model to the game in Figure 9.1, the utility function is simplified to

$$U_i(x_1,x_2)=x_i-\begin{cases}\alpha_i(x_{3-i}-x_i) & \text{if } x_{3-i}\geq x_i\\ \beta_i(x_i-x_{3-i}) & \text{if } x_{3-i}<x_i\end{cases}\qquad i=1,2$$

Obviously, if the responder rejects the offer, both utility functions are equal to zero, that is, $U_{1reject}=U_{2reject}=0$. If the responder accepts an offer of x, the utility functions are as follows:

$$U_{1accept}(x)=\begin{cases}(1+\alpha_1)M-(1+2\alpha_1)x & \text{if } x\geq M/2\\ (1-\beta_1)M-(1-2\beta_1)x & \text{if } x<M/2\end{cases}$$

$$U_{2accept}(x)=\begin{cases}(1+2\alpha_2)x-\alpha_2 M & \text{if } x<M/2\\ (1-2\beta_1)x+\beta_2 M & \text{if } x\geq M/2\end{cases}$$

The responder should accept the offer if and only if $U_{2accept}(x) > U_{2reject} = 0$. Solving for x we get the *threshold for acceptance*: $x > \alpha_2 M/(1+2\alpha_2)$. Evidently, if α_2 is close to zero, which indicates that player 2 (R) does not care much about being treated unfairly, the responder will accept very stingy offers. On the other hand, if α_2 is sufficiently big, the offer has to be close to half to be accepted. In any event, the threshold is not higher than $M/2$, which means that hyper-fair offers (more than half) are not necessary for the sake of acceptance.

Note that for the proposer, the utility function is monotonically decreasing in x when $x \geq M/2$. Hence, a rational proposer will not offer more than half of the money. Suppose $x \leq M/2$; two cases are possible depending on the value of β_1. If $\beta_1 > 1/2$, that is, if the proposer feels sufficiently guilty about treating others unfairly, the utility is monotonically increasing in x, and the best choice is to offer $M/2$. However, if $\beta_1 < 1/2$, the utility is monotonically decreasing in x, and hence the best offer for the proposer is the minimum one that would be accepted, that is, (a little bit more than) $\alpha_2 M/(1 + 2_z)$. Last, if $\beta_1 = 1/2$, it does not matter how much the proposer offers, as long as it is between $\alpha_2 M/(1 + 2\alpha_2)$ and $M/2$. Note that the other two parameters, α_1 and β_2, are not identifiable in ultimatum games.

As noted by Fehr and Schmidt, the model allows for the fact that individuals are heterogeneous. Different α's and β's correspond to different types of people. Although the utility functions are common knowledge, the exact values of the parameters are not. The proposer, in most cases, is not sure what type of responder he or she is facing. Along the Bayesian line, the proposer's belief about the type of the responder can be formally represented by a probability distribution P on α_2 and β_2. When $\beta_1 > 1/2$, the proposer's rational choice does not depend on what P is. When $\beta_1 < 1/2$, however, the proposer will seek to maximize the expected utility:

$$EU(x)=P(\alpha_2 M/(1+2\alpha_2)<x)\times((1-\beta_1)M-(1-2\beta_1)x)$$

Therefore, the behavior of a rational proposer in the ultimatum game is determined by the proposer's own type (β_1) and his or her belief about the type of the responder. The experimental data suggest that for many proposers, either β is big ($\beta > 1/2$) or they estimate the responder's α to be large. The choice of the responder is only determined by the responder's type (α_2) and the offer. Small offers are rejected by responders with a positive α.

The positive features of the above-described utility function are that it can rationalize both positive and negative outcomes and that it can explain the observed variability in outcomes with heterogeneous types. One of the major weaknesses of this model, however, is that it has a consequentialist bias: Players only care about final distributions of outcomes, not about how such distributions come about.[4] As we shall see, more recent experiments have established that how a situation is framed matters to an evaluation of outcomes and that the same distribution can be accepted or rejected depending on "irrelevant" information about the players or the circumstances of play. Another difficulty with this approach is that, if we assume the distribution of types to be constant in a given population, we should observe, overall, the same proportion of "fair" outcomes in ultimatum games. Not only does this not happen,

but we observe individual inconsistencies in behavior across different situations in which the monetary outcomes are the same. If we assume, as is usually done in economics, that individual preferences are stable, we would expect similar behaviors across ultimatum games. If instead we conclude that preferences are context dependent, we should provide a mapping from contexts to preferences that indicates in a fairly predictable way how and why a given context or situation changes one's preferences. Of course, different situations may change a player's expectation about another player's envy or guilt parameters, and we could thus explain why a player's behavior may change depending upon how the situation is framed. In the case of Fehr and Schmidt's utility function, however, experimental evidence that I shall discuss later implies that a player's own β (or α) changes value in different situations, yet nothing in their theory explains why one would feel consistently more or less guilty (or envious) depending on the decision context.

NORMS MATTER

Rule-based approaches are not completely new. Guth (1995), for example, interpreted the results of the ultimatum game as showing that people have rules of behavior such as sharing money equally, and they apply them when necessary. The problem with such solutions is that we need a plausible story about how people change their behavior in response to changes in payoffs and framing. If rules are inflexible, but we observe flexible compliance, there must be something wrong with a rule-based approach. Indeed, a common understanding of norms, one that I have tried to dispel in my definition (see Appendix 9.1), is that they are inflexible behavioral rules that one would apply in any circumstance that calls for them. Nothing could be farther from the truth. To be effective, norms have to be *activated* by salient cues.[5] As I explain (Bicchieri, 2000, 2006), a norm may exist, but it may not be followed simply because the relevant expectations are not there or because one might be unaware of being in a situation to which the norm applies.

I have argued that people have *conditional preferences* for conformity to a norm, in that they would prefer to follow it on condition that (1) they expect others to follow it and (2) they believe that, in turn, they are expected by others to abide by the norm (see Appendix 9.1 and Bicchieri, 2006). Both conditions have to be present to generate conformity. Indeed, there is plenty of evidence that manipulating people's expectations has an effect on norm compliance (Cialdini et al., 1990). Thus, I would argue that belief elicitation in experiments is crucial to determine whether a norm will be perceived as relevant and then followed.

We already know, for example, that telling subjects how others have behaved in a similar game has a profound effect on their choices and that allowing people to communicate before playing the game often results in a cooperative outcome.[6]

Ultimatum games are an ideal tool to study fair behavior, because they offer a very simple allocation choice. The good to be allocated is money, and the situation is such that most familiar contextual clues are removed. It is thus possible to introduce in this rarefied environment simple contextual information and control for its effects on the perception of what constitutes a fair division. We know that to be fair means different things in different contexts. In some situations being fair means sharing equally. In others it may mean giving more to the needy or to the deserving. In the simplest context, when there is no reason to differentiate between proposer and responder, an equal split is usually called for, but the salience of the equal split solution is lost if subjects are told that offers are generated by a random device (Blount, 1995) or if it is believed that the proposer was otherwise constrained in his or her decision. In both cases responders are willing to accept lower offers. This phenomenon is well known to consummate bargainers: If an unequal outcome can be credibly justified as a case of *force majeure*, people can be convinced to accept much less than an equal share. Also, variations in the strength of property rights alter the shared expectations of the two players regarding the norm that determines the appropriate division.

In the original ultimatum game, the proposer receives what amounts to a monetary gift from the experimenter. As a consequence, the proposer is perceived as having no special right to the money and is expected (at least in our culture) to share it equally with the responder. Because the fairness norm that is activated in this context dictates an equal split, the proposer who is offering little is perceived as stingy and consequently gets punished. Note that the proposer who was constrained in his or her decision is not seen as being intentionally stingy, because intentions do matter only when the choice is perceived as being freely made. To infer another person's intention or motive, we consider not only the action chosen, but also the actions that were not chosen but, as far as we know, *could* have been chosen.

Since what counts as fair is highly context dependent, a specific context simultaneously gives reasons to expect behavior appropriate to the situation and a clue as to the proposer's intention, especially when the offer is different from what is reasonably expected in that context. Subjects approach resource sharing or, for that matter, any other situation with implicit knowledge structures (scripts) that detail conditions that

are prototypically associated with sharing tasks. Once we have categorized the particular decision task we face, we enact scripts that tell us how people typically behave and what they expect others to do. However, it must be emphasized that people will display expected, appropriate behavior to the extent that crucial environmental cues match those of well-known prototypical scripts. An interesting question to ask is thus under which conditions an equal sharing norm will be violated. I shall discuss this point more extensively later on, but for now let me say that my hypothesis is that a deviation from equal sharing will be mainly due to (1) the presence of appropriate and acceptable justifications for taking more than an equal share or (2) the shift to a very different script that involves different roles and expectations. An example of the second reason is when the proposer is labeled "seller," and the responder, "buyer"; in this case the proposer offers a lower amount than in the control, and responders readily accept (and expect) less than an equal share (Hoffman, McCabe, Shachat, & Smith, 1994). In this case, the interaction is perceived as being market-like, and in a market script it is deemed equitable that a seller earns a higher return than a buyer. An example of the first reason is when the proposer has "earned" the right to the money by, for example, getting a higher score on a general knowledge quiz (Frey & Bohnet, 1995; Hoffman & Spitzer, 1985). In this case the proposer has an available, acceptable justification for getting more than the equal share. Doing better than someone else in a test is a common and reasonable mechanism, at least in our society, for determining differential access to a shared resource. It thus seems appropriate to many proposers to choose equity versus equality in such conditions.

There is continuity between real life and experiments with respect to how rights and entitlements, considerations of merit, need, desert, or sheer luck shape our perception of what is fair and what kind of reasons count as acceptable justifications for violating a fairness norm. Cultures differ in their reliance on different allocative and distributive rules, because such rules depend on different forms of social organization. Within a given culture, however, there usually is a consensus about how different goods and opportunities should be allocated or distributed. Cross-cultural studies of ultimatum and dictator games in 15 small-scale societies show quite convincingly that the behavior displayed in such games was highly correlated with the economic organization and social structure of each society (Henrich, Boyd, Bowles, Fehr, & Camerer, 2004). Furthermore, because experimental play is presumably categorized according to the specific sociocultural patterns of each society, the experimental results showed much greater variability than the results of typical ultimatum and dictator games played in modern

Western (or westernized) societies.[7] These results lend even more support to the hypothesis that social norms, and the accompanying shared expectations, play a crucial role in shaping behavioral responses to experimental games.

A norm-based explanation of the results of experiments with ultimatum and dictator games predicts that whenever proposers are focused upon the relevant expectations, they will behave in a norm-consistent way. In the traditional ultimatum game, the expected cost of not following an equal division rule may be enough to elicit fair behavior. In considering what the responder would accept, the proposer is forced to look at the situation and categorize it as a case in which an equality rule applies. This does not mean the person who follows the norms is in fact fair or casts a high value on equitable behavior. As I make plain in my definition of what it takes to follow an existing norm (Appendix 9.1), if a player assesses a sufficiently high probability to the opponent's following the norm and expects to be punished for noncompliance, that player will prefer to conform to a norm even if he or she has no interest in the norm itself.

The general utility function I introduced in Bicchieri, 2006, can now be applied to the ultimatum game. Let π_i be the payoff function for player i. The norm-based utility function of player i depends on the strategy profile s and is given by

$$U_i(s) = \pi_i(s) - k_i \max_{s_{-j} \in L_{-j}} \max_{m \neq j} \{\pi_m(s_{-j}, N_j(s_{-j})) - \pi_m(s), 0\}$$

where $k_i \geq 0$ is a constant representing i's sensitivity to the relevant norm. Such sensitivity may vary with different norms; for example, a person may be very sensitive to equality and much less so to equity considerations. The first maximum operator takes care of the possibility that the norm instantiation (and violation) might be ambiguous in the sense that a strategy profile instantiates a norm for several players simultaneously (as would be the case, for example, in a social dilemma with three players). The second maximum operator ranges over all the players other than the norm violator. In plain words, the discounting term (multiplied by k_i) is the maximum payoff deduction resulting from all norm violations.

The model is motivated by people's apparent respect (or disregard) for social norms regarding fairness. In the traditional ultimatum game, the norm usually prescribes a fair amount the proposer ought to offer. The norm functions that represent this norm are the following: N_1 is a constant N function, and N_2 is nowhere defined.[8] If the responder (player 2) rejects the offer, the utilities of both players are zero:

$$U_{1reject}(x) = U_{2reject}(x) = 0$$

Given that the proposer (player 1) offers x and the responder accepts, the utilities are the following:

$$U_{1accept}(x) = M - x - k_1 \max(N_1 - x, 0)$$

$$U_{2accept}(x) = x - k_2 \max(N_2 - x, 0)$$

where N_i denotes the amount player i thinks he or she should get or offer according to some social norm applicable to the situation, and k_i is non-negative. Note that k_i measures how much player 1 dislikes deviating from what he or she takes to be the norm. To obey a norm, sensitivity to the norm need not be high. Fear of retaliation may make a proposer with a low k behave according to what fairness dictates but, absent such risk, that player's disregard for the norm will lead him or her to be unfair. For the moment, I assume it is common knowledge that $N_1 = N_2 = N$, which is not too unreasonable in the traditional ultimatum game. Again, the responder should accept the offer if and only if $U_{2accept}(x) > U_{2reject} = 0$, which implies the following *threshold for acceptance*: $x > k_2N/(1 + k_2)$. Notice that an offer larger than the norm dictates is not necessary for the sake of acceptance.

For the proposer, the utility function is decreasing in x when $x \geq N$, so a rational proposer will not offer more than N. Suppose $x \leq N$. If $k_1 > 1$, the utility function is increasing in x, which means the best choice for the proposer is to offer N. If $k_1 < 1$, the utility function is decreasing in x, which implies that the best strategy for the proposer is to offer the least amount that would result in acceptance, that is, (a little bit more than) the threshold $k_2N/(1 + k_2)$. If $k_1 = 1$, it does not matter how much the proposer offers, provided the offer is between $k_2N/(1 + k_2)$ and N.

It should be noted that k_1 plays a very similar role as β_1 in the Fehr-Schmidt model. If we take N to be $M/2$ and k_1 to be $2\beta_1$, the two models agree on what the proposer's utility is. It is equally apparent that k_2 in this model is analogous to α_2 in the Fehr–Schmidt model. There is, however, an important difference between these parameters. The αs and βs in the Fehr–Schmidt model measure people's degree of aversion toward inequality, which is a very different disposition than the one measured by the ks, that is, people's sensitivity to different norms. The latter will usually be a stable disposition, and behavioral changes may thus be caused by changes in focus or in expectations. A theory of norms can explain such changes, whereas a theory of inequity aversion does not. I will come back to this point later.

It is also the case that the proposer's belief about the responder's type figures in the proposer's decision when $k_1 < 1$. The belief can be represented by a joint probability over k_2 and N_2, if the value of N_2 is not common knowledge. The proposer should choose an offer that maximizes the expected utility

$$EU(x) = P(k_2N_2/(1+k_2) < x) \times (M - x - k_1(N_1 - x)$$

As will become clear, an advantage this model has over the Fehr–Schmidt model is that it can explain some variants of the traditional ultimatum game more naturally. However, it shares a problem with the Fehr–Schmidt model: They both entail that fear of rejection is the only reason people offer almost-fair amounts rather than lower sums. This prediction, however, could be easily refuted by a parallel dictator game where rejection is not an option.

VARIATIONS ON THE ULTIMATUM GAME

So far I have only considered the basic ultimatum game, which is not the whole story. A number of interesting variants of the game exist in the literature, to some of which I now apply the two alternative models to see if they can tell reasonable stories about what happens in those experiments.

Ultimatum Game With Asymmetric Information and Payoffs

Kagel, Kim, & Moser (1996) designed an ultimatum game in which the proposer is given a certain amount of chips. The chips are worth either more or less to the proposer than they are to the responder. Each player knows how much a chip is worth to him or her but may or may not know that the chip has a different value to the other player. Participants play an ultimatum game over 10 rounds with changing opponents, and this is public knowledge. The particularly interesting setting is one in which the chips have higher (three times more) values for the proposer, and only the proposer knows it. It turns out that in this case the offer is (very close to) half of the chips and the rejection rate is low. A popular reading of this result is that people merely prefer to *appear* fair, as a really fair person is supposed to offer about 75% of the chips. As Figure 9.2 shows, proposers offered close to 50% of the chips, and very few such offers were rejected.

To analyze this variant formally, we only need a small modification to our original setting. That is, if the responder accepts an offer of x, the proposer actually gets $3(M - x)$, though, to the responder's knowledge, the proposer only gets $M - x$. In the Fehr–Schmidt model, the utility

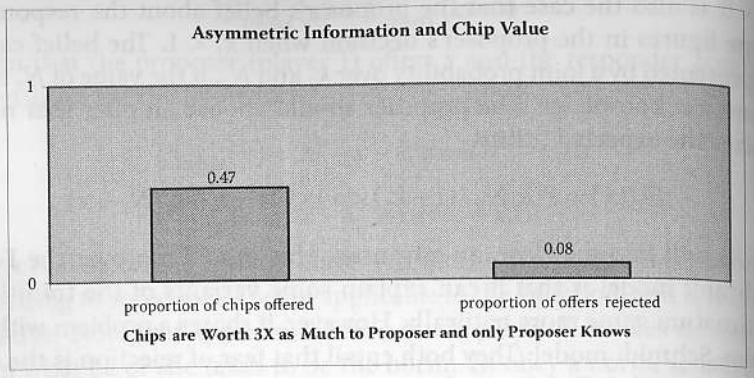

Figure 9.2 Asymmetric information about chips value.

function of player 1 (the proposer), if the offer gets accepted, is now the following:

$$U_{1accept}(x)=\begin{cases}(3+3\alpha_1)M-(3+4\alpha_1)x & \text{if } x\geq 3M/4\\ (3-3\beta_1)M-(3-4\beta_1)x & \text{if } x<3M/4\end{cases}$$

The utility function of the responder upon acceptance does not change, as, to the best of the responder's knowledge, the situation is the same as in the simple ultimatum game. Also, if the responder rejects the offer, both utilities are again zero. It follows that the responder's threshold for acceptance remains the same; he or she accepts the offer if $x > \alpha_2 M/(1 + 2\alpha_2)$. For the proposer, if $\beta_1 > 3/4$, his or her best offer is $3M/4$; otherwise the best offer is the minimum amount above the threshold. An interesting point is that even if a player offers $M/2$ in the simple ultimatum game, which indicates that $\beta_1 > 1/2$, that player may not offer $3M/4$ in this new condition. This prediction is consistent with the observation that almost no one offers 75% of the chips in the real game.

At this point, it seems the Fehr–Schmidt model does not entail a difference in behavior in this new game, but proposers in general do offer more in this new setting than they do in the usual ultimatum game, which naturally leads to the lower rejection rate. Can the Fehr–Schmidt model explain this? One obvious way is to adjust α_2 so that the predicted threshold increases, but there is no reason in this case for the responder to change his or her attitude toward inequality. Another explanation might be that under this new setting, the proposer believes that the responder's distaste for inequality increases, for after all, it is the proposer's belief about α_2 that affects the offer. This move sounds as questionable as the last one, but it does point to a reasonable explanation. Because the proposer is uncertain about the responder's type, the proposer's belief about α_2 should be represented by a nondegenerate probability distribution. The proposer should choose an offer that maximizes his or her expected utility, which in this case is given by the following:

$$EU(x) = P\,(\alpha_2 < x / (M - 2x)) \times ((3 - 3\beta_1)\,M - (3 - 4\beta_1)\,x)$$

The main difference between this expected utility and the one in the simple ultimatum game is that it involves a bigger stake. Hence, it is likely to be maximized at a bigger x unless the distribution (the proposer's belief) over α_2 is sufficiently odd. Thus, the Fehr–Schmidt model can explain the phenomenon in a reasonable way.

If we apply my model to this new setting, again the utility function of player 2 does not change. The utility function of player 1 (the proposer) given acceptance is changed to

$$U_{1accept}(x)=3(M-x)-k_1\max(N_1'-x,0)$$

I use N_1' here to indicate that the proposer's perception of the fair amount, or the proposer's interpretation of the norm, may have changed due to his or her awareness of the informational asymmetry.[9] My model behaves quite similarly to the previous one. Specifically, the responder's threshold for acceptance is still $k_2N_2/(1 + k_2)$. The proposer will and should offer N_1' only if $k_1 > 3$, so people who offer the fair amount in the simple ultimatum game ($k_1 > 1$) may not offer the fair amount in the new setting. That means that even if $N_1' = 3M/4$, the observation that few people offer that amount does not go against my model. The best offer for most people ($k_1 < 3$) is the smallest amount that would be accepted. However, because the proposer is not sure about the responder's type, the proposer will choose an offer to maximize his or her expected utility, and this in general leads to an increase of the offer, given an increase of the stake. Although it is not particularly relevant to the analysis in this case, it is worth noting that N_1' is probably less than $3M/4$ in the situation as thus framed. This point will become crucial in games with obvious framing effects.

Ultimatum Game with Different Alternatives

There is also a very simple twist to the ultimatum game, which turns out to be quite interesting. Falk et al. (2003) introduced a simple ultimatum game where the proposer has only two choices: either offer 2 (and keep 8) or make an alternative offer that varies across treatments

in a way that allows the experimenter to test the effect of reciprocity and inequity aversion on rejection rates. The alternative offers in four treatments are (5,5), (8,2), (2,8), and (10,0). As Figure 9.3 shows, when the (8,2) offer is compared to the (5,5) alternative, the rejection rate is 44.4%, which is much higher than the rejection rates in each of the three alternative treatments. It turns out that the rejection rate depends a lot on what the alternative is. The rejection rate decreases to 27% if the alternative is (2,8), and further decreases to 9% if the alternative is (10,0).[10]

It is hard for the Fehr–Schmidt model to explain these results. In their consequentialist model there does not seem to be any role for the available alternatives to play. As the foregoing analysis shows, the best reply for the responder is acceptance if $x > \alpha_2 M/(1 + 2\alpha_2)$. That is, different alternatives can affect the rejection rate only through their effects on α_2. It is not entirely implausible to say that what could have been otherwise affects one's attitude towards inequality. After all, one's dispositions are shaped by all kinds of environmental or situational factors, to which the "path not taken" seems to belong. Still it sounds quite odd that one's sensitivity to fairness changes as alternatives vary, and in particular, it is not compatible with the assumption of independence of irrelevant alternatives, a common assumption in decision theory.

The norm-based model, by contrast, seems to have an easier time. For one thing, my model can explain the data by telling a story about how the norm's perception might change, and the story, unlike the previous case, can be quite plausible. Recall that my definition of what it takes to follow a norm relies heavily on expectations, both empirical and normative. As I discussed (Bicchieri, 2000, 2006), how we decide and act in a situation depends upon how we interpret, understand, and encode it. Once a situation is categorized as a member of a particular class, a schema (or script) is invoked. Such a script allows us to make

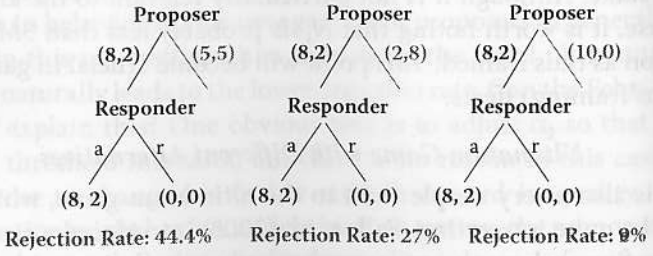

Figure 9.3 Ultimatums with alternatives offers.

inferences about unobservable variables, predict other people's behavior, make causal attributions, and modulate emotional reactions. The script we invoke is the source of both projectible regularities and the legitimacy of our expectations. If, as I argued (Bicchieri, 2006), social norms are embedded into scripts, then the particular way a situation is framed will have a large effect on our expectations about others' behavior and what they expect from us. Thus, a change in the way a situation is framed will induce a change in expectations and have an immediate effect on our focusing (or not focusing) on the norm that has (or has not) been elicited.

As the possible alternatives vary, the player may no longer believe that the same norm applies, and it is quite reasonable to conjecture that different alternatives point the responder to different norms (or lack thereof). In the (8,2), (5,5) situation, players are naturally focused on the equal split. The proposer who could have chosen it but did not is sending a clear message about his disregard for fairness. If the expectation of a fair share is violated, the responder will probably feel outraged, attribute a greedy intention to the proposer, and punish him accordingly. If the alternatives are (8,2) or (2,8), few people would expect a proposer to sacrifice for the responder. In real life, situations like this are decided with a coin toss. In the game context, it is difficult to see that any norm would apply to the situation. This is why 70% of the subjects choose the (8,2) split and only 27% reject it. Finally, the choice of (8,2) when the alternative is (10,0) appears quite nice, and indeed the rejection rate is only 9%. When the alternative for the proposer is to offer the whole stake, there is little reason for the responder to think that the norm is still (50%, 50%) or something close to this. Thus, a natural explanation given by my model is that N_2 changes (or may be empty) as the alternative varies. The results of this experiment tell us that most people do not have selfish material preferences, in which case they would always accept the (8,2) division. They also tell us that people are not simply motivated by a dislike for inequality, for otherwise we would have observed the same rejection rate in all contexts.

Ultimatum Game With Framing

Framing effects, a topic of continuing interest to psychologists and social scientists, have also been investigated in the context of ultimatum games. Hoffman et al. (1994), for example, designed an ultimatum game in which groups of 12 participants were ranked on a scale of 1 to 12 either randomly or by superior performance in answering questions about current events. The top six were assigned to the role of seller and the rest to the role of buyer. They also ran experiments with the stan-

dard ultimatum game instructions, both with random assignments and assignment to the role of proposer by contest. The exchange and contest manipulations elicited significantly lowered offers, but rejection rates were unchanged compared to the standard ultimatum game.[11]

Figure 9.4 shows that the "exchange" framing significantly lowered offers but also that being the winner of a contest in the traditional ultimatum game had an effect on the proposers' offers. Several other experiments have consistently shown that when the proposer is a contest winner (Frey & Bohnet, 1995) or has "earned the right" to that role (Hoffman & Spitzer, 1985), offers are lower than in the traditional ultimatum game. As I suggested before, in the presence of prototypical, acceptable justifications for deviating from equality, subjects will be induced to follow an equity principle. Framing in this case provides salient cues suggesting that an equity rule is appropriate to the situation.

Because, from a formal point of view, these situations are not different from that of a traditional ultimatum game, the previous analysis remains the same. Hence, according to the Fehr–Schmidt model, the framing of the game decreases α_2. In other words, the role of a buyer or the knowledge that the proposer was a superior performer or had simply earned the right to his role lowers the responder's concern for fairness. This change does not sound intuitive and demands some explanation. In addition, the proposer has to *expect* this change in the responder's concern for fairness in order to lower his offer. It is equally, if not more, difficult to see why the framing can lead to different beliefs the proposer has about the responder.

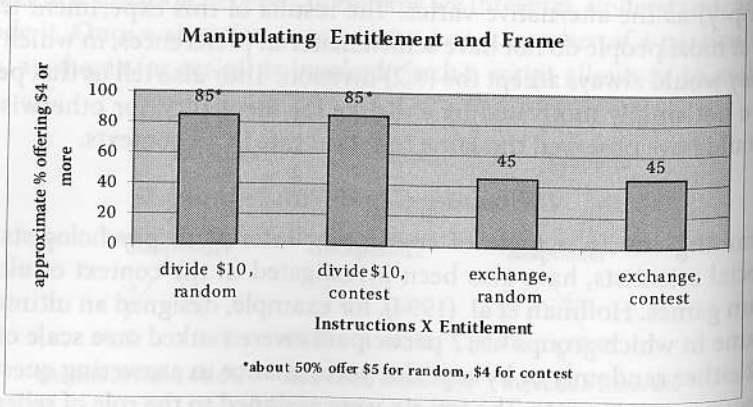

Figure 9.4 Entitlement and framing effects.

In my model, the parameter N plays a vital role again. Although we need more studies about how and to what extent framing affects people's expectations and perception of what norm is being followed, it is intuitively clear that framing, like the examples mentioned above, will change the players' conceptions of what is fair. The exchange framework is likely to elicit a market script where the seller is expected to try to get as much money as possible, whereas the entitlement context has the effect of focusing subjects away from equality in favor of an equity rule. In both cases, what has been manipulated is the perception of the situation and thus the expectations of the players. An individual's sensitivity and concern for norms may be unchanged, but the relevant norm is clearly different from the usual fairness as equality rule.

Dictators With Uncertainty

In a theory of norms, the role of expectations is crucial. Norms and expectations are part of the same package. Focusing people on a norm usually means eliciting certain expectations and, in turn, when people have the right empirical and normative expectations, they will tend to follow the relevant norm. In the traditional ultimatum game, at least in Western societies, the possibility of rejection forces the proposer to focus upon what is expected of him or her.[12] In the absence of information about the responder, and without a history of previous games and results as a guide, equal (or almost equal) shares become a focal point. Eliminate the possibility of rejection, and equality becomes much less compelling: For example, we know that in double-blind dictator games, 64% of the proposers keep all the money. The dictator game is particularly interesting as a testing ground for the study of how norms influence behavior, because it illustrates in a clear manner how sensitive we are to the presence, reminder, or absence of others' expectations. In such a decision context, an equal share seems much less compelling. In fact, in a dictator game there is no prima facie clear-cut behavioral rule to follow and, because of that, we can better examine the role expectations (and their manipulation) play in the emergence of a consensual script and, consequently, a social norm.

An experiment conducted by Dana, Weber, & Kuang (2003) enlightens this point. The basic setting is a dictator game where the allocator has only two options. The game is played in two very different situations. Under the known condition (KC), the payoffs are unambiguous, and the allocator has to choose between option A (6,1) and option B (5,5), where the first number in the pair is the allocator's payoff, and the second number is the receiver's payoff. Under the unrevealed condition (UC), the allocator has to choose between option A (6,?) and option

B (5,?), where the receiver's payoff is 1 with probability 0.5 and 5 with probability 0.5 (Figure 9.5). Before making a choice, however, the allocator is given the option to find out privately at no cost which game is being played and thus know what the receiver's payoff is.

It turns out that 74% of the subjects choose B (5,5) in KC, and 56% choose A (6,?) without revealing the actual payoff matrix in UC. This result, as Dana et al. (2003) point out, stands strongly against the Fehr–Schmidt model. If we take the revealed preference as the actual preference, choosing (5,5) in KC implies that $\beta_1 > 0.2$, while choosing (6,?) without revealing in UC implies that $\beta_1 < 0.2$.[13] Hence, unless a reasonable story can be told about β_1, the model does not fit the data. If a stable preference for fair outcomes is inconsistent with the above results, can a conditional preference for following a norm show greater consistency? Note that, if we were to assume that N_i is fixed in both experiments, a similar change of k would occur in my model, too.[14] However, the norm-based model can offer a natural explanation of the data through an interpretation of N_i. In KC subjects have only two, very clear choices. There is a fair outcome (5,5) and there is an inequitable one (6,1). Choosing (6,1) entails a net loss for the receiver and only a marginal gain for the allocator.

Note that the choice framework focuses subjects on fairness, though, as I mentioned before, the usual dictator game has no such focus. Dana et al.'s (2003) example evokes a related situation (one that we frequently encounter) in which we may choose to give to the poor or otherwise disadvantaged: What is $1 more to the allocator is $4 more to the receiver, mimicking the multiplier effect that money has for a poor person. In this experiment, what is probably activated is a norm of beneficence, and subjects uniformly respond by choosing (5,5). Indeed, when receivers in Dana

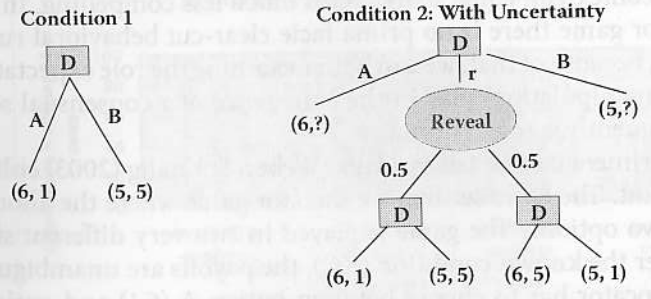

Figure 9.5 Dictator games with and without uncertainty.

et al.'s experiment were asked what they would choose in the allocator's role, they unanimously chose the (5,5) split as the most appropriate.

A natural question to ask is whether we should hold N fixed, thus assuming a variation in people's sensitivity to the norm (k), or if instead what is changing here is the perception of the norm itself. I argue that what changes from the first to the second experiment is the perception that a norm exists and applies to the present situation, as well as expectations about other people's behavior and what their expectations about one's own behavior might be. Recall that in my definition of what it takes for a norm to be followed, a necessary condition is that a sufficient number of people expect others to follow it in the appropriate situations *and* believe they are expected to follow it by a sufficient number of other individuals. People will prefer to follow an existing norm conditionally upon entertaining such expectations. In KC the situation is transparent, and so are the subjects' expectations. If a subject expects others to choose (5,5) and believes he or she is expected so to choose, that subject might prefer to follow the norm (provided the subject's k, which measures one's sensitivity to N, is large enough). In UC, on the contrary, there is uncertainty as to what the receiver might be getting. To pursue the analogy with charitable giving further, in UC there is uncertainty about the multiplier ("am I giving to a needy person or not?") and thus there is the opportunity for *norm evasion*: The player can avoid activating the norm by not discovering the actual payoff matrix. Though there is no cost to see the payoff matrix, people will opt to not see it in order to avoid having to adhere to a norm that could potentially be disadvantageous. Thus, a person who chooses (5,5) under KC may choose (6,?) under UC with the same degree of concern for norms. Choosing to reveal looks like what moral theorists call a *supererogatory* action. We are not morally obliged to perform such actions, but it is awfully nice if we do. Indeed, I believe few people would expect an allocator to choose to reveal, and similarly I would expect few people would be willing to punish an allocator who chooses to remain in a state of uncertainty.

A very different situation would be one in which the allocator has a clear choice between (6,1) and (5,5) but is told that the prospective receiver does not even know he or she is playing the game. In other words, the binary choice would focus the allocator, as in the KC condition, on a norm of beneficence, but the allocator would also be cued about the absence of a crucial expectation. If the recipient does not expect the allocator to give anything, is there any reason to follow the norm? This is a good example of what I have extensively discussed in Bicchieri (2000, 2006). A norm exists, the subject knows it and knows the norm applies to the situation, but the

subject's preference for following the norm is conditional on having certain empirical and normative expectations (see Appendix 9.1). In our example the normative expectations are missing, because the recipient does not know that a dictator game is being played or know his or her part in it. In this case, I predict that a large majority of allocators will choose (6,1) with a clear conscience. This prediction is different from what a fairness preference model would predict, but it is also at odds with theories of social norms as constraints on action. One such theory is Rabin's (1995) model of moral constraints. Very briefly, Rabin assumes that agents maximize egoistic expected utility, subject to constraints. Thus, our allocator will seek to maximize his or her payoffs but experience disutility if the action taken is in violation of a social norm.

However, if the probability of harming another is sufficiently low, a player may circumvent the norm and act more selfishly. Because in Rabin's model, the norm functions simply as a constraint, beliefs about others' expectations play no role in a player's decision to act. Because the (6,1) choice does harm the recipient, Rabin's model should predict that the number of subjects who choose (6,1) is the same as in the KC of Dana et al.'s (2003) experiment. In my model, however, the choices in the second experiment will be significantly different from the choices we have observed in Dana et al.'s KC condition.

To summarize, the norm-based model explains the behavioral changes observed in the above experiments as due to a (potentially measurable) change in expectations. An individual's propensity to follow a given norm would remain fixed, as would the individual's preferences. However, because preferences in my model are conditional upon expectations, a change in expectations will have a major, predictable effect on behavior.

APPENDIX 9.1: CONDITIONS FOR A SOCIAL NORM TO EXIST

Let R be a behavioral rule for situations of type S, where S can be represented as a mixed-motive game. We say that R is a social norm in a population P if there exists a sufficiently large subset $P_{cf} \subseteq P$ such that for each individual $i \in P_{cf}$:

1. **Contingency:** i knows that a rule R exists and applies to situations of type S.
2. **Conditional preference:** i prefers to conform to R in situations of type S on the condition that:
   a. **Empirical expectations:** i believes that a sufficiently large subset of P conforms to R in situations of type S, and either:
      b. **Normative expectations:** i believes that a sufficiently large subset of P expects i to conform to R in situations of type S;
      b' **Normative expectations with sanctions:** i believes that a sufficiently large subset of P expects i to conform to R in situations of type S, prefers i to conform, and may sanction behavior.

A social norm R is followed by population P if there exists a sufficiently large subset $P_f \subseteq P_{cf}$ such that, for each individual $i \in P_f$, conditions 2a and either 2b or 2b' are met for i, and, as a result, i prefers to conform to R in situations of type S.

REFERENCES

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *Economic Journal, 100*, 464–477.

Bicchieri, C. (2000). Words and deeds: A focus theory of norms. In J. Nida-Rumelin & W. Spohn (Eds.), *Rationality, rules and structure.* Dordecht, The Netherlands: Kluwer Academic.

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms.* Cambridge, UK: Cambridge University Press.

Blount, S. (1995). When social outcomes aren't fair: The effect of causal attributions on preferences. *Organizational Behavior and Human Decision Processes, 63*: 131–144.

Camerer, C. (2003). *Behavioral game theory: Experiments on strategic interaction.* Princeton, NJ: Princeton University Press.

Camerer, C., Loewenstein, G., & Rabin, M. (2004). *Advances in behavioral economics.* Princeton, NJ: Princeton University Press.

Camerer, C., & Thaler , R. H. (1995). Anomalies: Ultimatums, dictators, and manners. *Journal of Economic Perspectives, 9(2)*:209–219.

Cameron, L. (1995). Raising the stakes in the ultimatum game: Experimental evidence from Indonesia. Princeton Department of Economics, Industrial Relations Sections. Working Paper, 345.

Cialdini, R., Kallgren, C., et al. (1990). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. *Advances in Experimental Social Psychology, 24*, 201–234.

Dana, J., Weber, R., & Kuang, J. X. (2003). Exploiting moral wriggle room: Behavior inconsistent with a preference for fair outcomes. Carnegie Mellon Behavioral Decision Research. Working Paper, 349.

Falk, A., Fehr, E., & Fischbacher, U. (2003). Testing theories of fairness—Intentions matter. Institute for Empirical Research in Economics, University of Zürich. Working Paper, 63.

Fehr, E., & Gachter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives, 14*(3), 159–181.

Fehr, E., & Schmidt, K. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics, 114*, 817–868.

Forsythe, R., Horowitz, J.L, Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior, 6*, 347–369.

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

Frey, B., & Bohnet, I. (1997). Identification in democratic society. *Journal of Socio-Economics, 26*, 25–38.

Guth, W. (1995). On ultimatum bargaining experiments: A personal review. *Journal of Economic Behavior and Organization, 27*, 329-344.

Guth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum games. *Journal of Economic Behavior and Organization, 3*, 367–388.

Henrich, J., Boyd, R., Bowles, S., Fehr, H. G. E., & Camerer, C. (Eds.). (2004). *Foundations of human sociality: Ethnography and experiments in 15 small-scale societies*. Oxford, UK: Oxford University Press.

Hoffman, E., McCabe, K. A., Shachat, K., & Smith, V. (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior, 7*, 346–380.

Hoffman, E., McCabe, K. A., & Smith, V. (1998). Behavioral foundations of reciprocity: Experimental economics and evolutionary psychology. *Economic Inquiry, 36*, 335–352.

Hoffman, E., & Spitzer, M. (1985). Entitlements, rights, and fairness: An experimental examination of subjects' concept of distributive justice. *Journal of Legal Studies, 2*, 259–297.

Kagel, J. H., Kim, C., & Moser, D. (1996). Fairness in ultimatum games with asymmetric information and asymmetric payoffs. *Games and Economic Behavior, 13*, 100–110.

Rabin, M. (1995). Moral preferences, moral constraints, and self-serving biases. University of California at Berkeley, Department of Economics. Working Paper, 95-241.

Roth, A. E., Prasnikar, V., Okuno-Fujiwara, M., & Zamir, S. (1991). Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An experimental study. *American Economic Review, 81*, 1068–1095.

NOTES

1. The conditions for following a norm are formally described in Chapter 1 of Bicchieri, 2006, and in Appendix 1 here.
2. Guth et al. (1982) were the first to observe that the most common offer by proposers was to give half of the sum to the responder. The mean offer was 37% of the original allocation. In a replication of their experiments, they allowed subjects to think about their decision for one week. The mean offer was 32% of the sum, which is still very high.
3. The term max $(x_j - x_i, 0)$ denotes the maximum of $x_j - x_i$ and 0; it measures the extent to which there is disadvantageous inequality between *i* and *j*.
4. This is a separability of utility assumption: What matters to a player in a game is the payoff at a terminal node. The way in which that node was reached and the possible alternative paths that were not taken are irrelevant to an assessment of the player's utility at that node. Utilities of terminal node payoffs are thus separable from the path through the tree and from payoffs on unchosen branches.
5. Cues that activate, or bring to mind, a norm may involve a direct statement or reminder of the norm, observing others' behavior, similarity of the present situation to others in which the norm was used, as well as how often or how recently one has used the norm.
6. I discuss these results and the relevant literature in Bicchieri, 2006, Chapter 4.
7. In some groups, rejections were extremely rare, even when offers were very low, whereas in other groups "hyper-fair" offers were frequently rejected, pointing to very different (but interculturally shared) interpretations of the experimental situation.
8. Intuitively, N_2 should proscribe rejection of fair (or hyper-fair) offers. The incorporation of this consideration, however, will not make a difference in the formal analysis.
9. It is important to note that since norms are very dependent on expectations, informational asymmetries will almost certainly affect norm-following behaviors.
10. Note that 30% of the subjects proposed (8,2) when the alternative was (5,5), 70% proposed (8,2) when the alternative was (2,8), and 100% proposed (8,2) when the alternative was (10,0). Each player played four games, presented in random order, in the same role.
11. Rejections remained low throughout, about 10%. All rejections were on offers of $2 or $3 in the exchange instructions. There was no rejection in the contest entitlement/divide $10, and there was 5% rejection of the $3 and $4 offers in the random assignment/divide $10.

12. I do not want to imply that sanctions are crucial to norm following. 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.
13. In KC choosing option B implies that $U_1(5,5) > U_1(6,1)$, or $5 - \alpha_1(0) > 6 - \beta_1(5)$. Hence, $5 > 6 - 5 - \beta_1$ and therefore $\beta_1 > 0.2$. In UC not revealing and choosing option A implies that $U_1(6, (.5(5), .5(1))) > U_1(.5(5,5), .5(6,5))$, since revealing will lead to one of the two "nice" choices with equal probability. We thus get $6 - .3(\beta_1) > 2.5 + .5(6 - \beta_1)$, which implies that $\beta_1 < 0.2$.
14. According to my model, if we keep N_i constant, choosing option B in KC means that $U_1(5,5) > U_1(6,1)$, and hence $5 > 6 - k_1(4)$. It follows that $k_1 > 0.25$. In UC not revealing and choosing option A implies that $U_1(6, (.5(5), .5(1))) > U_1(.5(5,5), .5(6,5))$, and hence $6 - k_1(2) > 5.5$, which implies that $k_1 < 0.25$.