Appendix B Is Homo Sociologicus Strategic?

The analysis in this book accepts the notion that people tend to respect the socially expected and normative sanctions (Chapter 5). It also rests on a particular notion of rationality, maintaining that when institutions generate behavior, socially articulated and disseminated rules regarding the situation span the domain that people understand and within which they can act rationally. Are these two premises consistent with each other? Is it appropriate to consider individuals as strategic while recognizing that social and normative considerations influence behavior? Or should we model people as *homo sociologicus,* as passive rule followers?[1] Specifically, is it appropriate to model individuals who have such social and normative inclinations as rational decision makers when they are guided by socially articulated and disseminated rules?[2] Do they have stable preferences regarding outcomes? Are they motivated by the consequences of their actions? In other words, do they act strategically? This appendix presents evidence to support the claim that, although people have social and normative propensities, it is nevertheless appropriate and necessary to consider them as rational in the above sense.

Experimental game theory is a promising analytical framework to address these questions, particularly because participants share common knowledge of the rules of the game and many experiments were explicitly designed to reveal individuals' social and normative inclinations. These experiments provide three ways to address the foregoing questions: considering whether nonrational explanations better fit the data, testing whether the observed behavior is consistent with some well-behaved preference ordering, and using experimental results to determine whether people are motivated by consequences and behave strategically. The

[1] For a survey of psychological evidence indicating that individuals do not always have stable preferences, see Rabin (1998), who identifies two main reasons for this to be the case. First, people have difficulty evaluating their own preferences: they do not always accurately predict their own future preferences or even accurately assess the well-being they have experienced from past choices. Second, research on framing effects, preference reversals, and related phenomena reveal that people may prefer option x to y when the choice is elicited one way but prefer y to x when the choice is elicited another way. The first issue is more relevant to what people want to exchange and less relevant to the issue here, namely, how institutions enable actions. The second issue is consistent with the argument made here—that institutions frame the context within which individuals choose actions.

[2] The literature on the issue of rationality is immense. For a recent discussion and survey, see Mantzavinos (2001, pp. 50-4).

evidence is inconsistent with nonrational accounts, consistent with a well-behaved preference ordering, and reflective of consequential and strategic behavior.[3]

Consider, for example, the ultimatum game, which has been used to study altruism. In this game, a proposer suggests a potential division of a fixed amount of money. The responder can either agree to the proposal (in which case the amount is divided accordingly) or disagree (in which case both get nothing). If players are motivated only by self-interest and money income, the unique subgame perfect equilibrium is one in which the proposer makes the smallest possible offer, which the responder accepts. Numerous experiments conducted in different countries with different monetary amounts and different experimental procedures reveal that this is not what actually happens. Fehr and Schmidt (1999) report that in 71 percent of the cases, the proposers offered between 40 and 50 percent of the total to the other player. Moreover, individuals often reject low offers, revealing that they prefer that both parties end up with nothing to receiving what they perceive as an inadequate allocation.

Although this evidence is usually considered as reflecting altruism or aversion to inequality, a nonrational explanation for this behavior has been advanced. Roth and Erev (1995) and Binmore, Gale, and Samuelson (1995) try to explain the existence of fair offers and the rejection of low offers in this game using an irrational learning model.[4] The central idea is that proposers and responders have distinct incentives to learn. The rejection of low offers is not costly for responders, who are irrational in terms of slowly learning, rather than deducting, not to reject them. In contrast, rejections are very costly to proposers, who therefore quickly learn to avoid making low offers. Hence behavior may not converge to the subgame perfect equilibrium in which the lowest possible offers are made.

The validity of such learning arguments with respect to simple games such as the ultimatum game seems doubtful. Furthermore, in many studies (as discussed later), proposers do anticipate responders' reactions.[5]

---

[3] See E. Hoffman, (1996b); Hoffman et al. (1996a, 1996b); Fehr and Schmidt (2001), Henrich et al. (2001), and Falk and Fischbacher (2000).

[4] Another possible explanation is that because most real-life interactions are repeated, subjects in laboratory mistake one-time games for repeated games. Even if this is true, it cannot account for many of the results reported, such as the tendency to cooperate when interactions are anonymous and behavior is known to be of short duration.

[5] The merit of an alternative theory—that individuals act in a one-shot game as they do in repeated games—is discussed later.

Are altruistic individuals rational? Using the dictator game, Andreoni and Miller (2002) demonstrate that behavior exhibiting social preference is consistent with a well-behaved preference ordering. The dictator game resembles the ultimatum game, except that the proposer acts as a dictator who can divide the fixed amount in any way he pleases (including assigning the full amount to himself). Andreoni and Miller constructed dictator game experiments in which they manipulated the "exchange rate" between what the dictator gives and the other player receives. For every dollar the dictator gave up, the other person received an amount smaller than, equal to, or greater than one dollar. Changing the dictator's budget constraint in this way enabled the behavior of the same individual to be examined under different constraints. It is therefore possible to test whether behavior satisfies the necessary and sufficient conditions required for the existence of well-behaved preferences.[6]

The results were unambiguous, leading the authors to conclude that preferences are predictable and well behaved at the aggregate level and that individuals exhibit a significant degree of rationally altruistic behavior. Indeed, more than 98 percent of the subjects made choices that were consistent with utility maximization. It is possible to capture altruistic choices with quasi-concave utility functions for individuals; altruism reflects rational behavior, given the underlying preferences.[7] Furthermore, Andreoni and Miller found that a model capturing the preference revealed in one experiment consistently accounts for behavior in other experiments.[8]

Many experiments reveal that individuals respond as postulated in game theory to the strategic environment in which they interact.[9] In hundreds of double-auction experiments, prices

[6] Specifically, they have examined whether individuals reveal a preference ordering that satisfies the generalized axiom of revealed preference (GARP). A is directly revealed as preferred to B if B was in the choice set when A was chosen. If A is directly revealed as preferred to B, B is directly revealed as preferred to C and Y is directly revealed as preferred to Z, then A is indirectly revealed as preferred to Z. The GARP is as follows: A is indirectly revealed as preferred to B, then A is not strictly within the budget set when B is chosen, that is, B is not strictly directly revealed as preferred to A. Satisfying GARP is both a necessary and sufficient condition for the existence of well-behaved preferences, given linear budget constraints.

[7] As they note, however, their analysis did not explore the influence of the changing environment—the rules of the game, level of anonymity, the gender or age of the participants, or the framing of the decision—on the preference ordering.

[8] Fehr and Schmidt (1999) report similar results.

[9] Ostrom (1998), however, argues that "what is clearly the case from experimental evidence is that players do not use backward induction in their decisionmaking plans in an experimental laboratory" (p. 5). The context of these words, however, suggests that what she might have had in mind is that the results are inconsistent with backward induction in finitely repeated games, under the assumption that people are motivated only by self-interest. Ostrom cites Rapoport (1997) and McKelvey and Palfery (1992) to

and quantities quickly converged to the competitive equilibrium predicted by standard self-interest theory.[10] In the case of the two games discussed previously, Forsythe et al. (1994) hypothesize that if people are motivated only by altruism (or aversion to inequality), the outcome in both games should be the same. However, individuals could also be reciprocators—conditional cooperators who are willing to take materially costly actions that raise or lower others' payoffs depending on others' past actions and their perceived intentions. In particular, if people are willing to punish others for what they consider to be unfairly low offers and if the proposers anticipate this through backward induction, higher offers should be made in the ultimatum game than in the dictator game. In fact, offers were significantly higher in the ultimatum game, suggesting that many proposers do apply backward induction. In a ten dollar dictator game, 21 percent of the proposers gave the other player nothing, and 21 percent gave the other an equal share. In a ten dollar ultimatum game, however, all proposers offered the responder something, and 75 percent offered at least an equal amount.

Similar results are reported in cross-country analysis. Henrich et al. (2001) conducted experiments in 15 very different settings, ranging from modern urban to hunter-gatherer societies. They concluded that in all of these societies, individuals exhibited stable preferences and behavior motivated by consequences. In each society, people by and large correctly anticipated the responses of others.[11]

Fehr and Schmidt (1999) also found evidence of backward induction. They report that in twelve public good games without punishment, in which free-riding is a dominant strategy, average and median contributions in the first period were 40–60 percent of the endowment, but 73 percent of participants contributed nothing in the last period.

Fehr and Gächter (2000) conducted experiments with an extended public good game in which individuals have the option to participate in the (costly) punishment of others after contributions are made. They found that the behavior of reciprocators who are willing to punish

---

support her position. But Rapoport's analysis is not concerned with rejecting backward induction. His focus and main conclusion regard the importance of the framing effect on behavior, captured by information about the order of play (p. 133). He notes that the order of moves influences equilibrium selection. McKelvey and Palfery (1992) examine the centipede game, which is problematic as far as backward induction is concerned, as Fudenberg and Tirole (1991, pp. 96-100) note. They conclude that a game of incomplete information based on reputation explains their data.

[10] See surveys in Davis and Holt (1993) and Hagel and Roth (1995).

[11] See also Roth et al. (1991).

free-riding is anticipated by at least some potential free riders and that the expectation that free-riding will be punished prevents its occurrence from the beginning. Individuals who deviated more from the average contribution were punished more severely, and they responded to this punishment by increasing their contributions. Some individuals inflicted punishments to generate an increase in average contributions and were successful in achieving this.

Fehr and Fischbacher (2001) explicitly tested the ability of individuals to understand the strategic difference between one-shot and repeated games. The evidence indicates that, by and large, they understand it very well. Fehr and Fischbacher ran two sets of ultimatum game experiments. In both experiments, subjects played the game ten times, each time with a different opponent. In one set of experiments, the proposers knew nothing about the past behavior of their current responders. In the other set a "reputation" condition was imposed, as past behavior of the responders was made known. In theory, if individuals understand the distinction between one-shot and repeated interactions, responders would be motivated to build up reputations for "toughness" and rejection of low offers. Hence the acceptance threshold (the lowest acceptable offer for the responder) should increase. Slightly more than 80 percent of the responders increased their acceptance thresholds under the reputation condition.[12]

Gächter and Falk (2002) found behavior consistent with the insight of incomplete information models—that is, individuals act "as if" they are of a particular type in order to cause others to identify them as such. They examined behavior in gift exchange games in which the proposer offers a wage to the responder, which the responder can accept or reject. If the responder rejects the offer, both players receive a zero payoff. If he accepts, he is paid the offered wage but has to make a costly "effort" choice. Clearly, if the responder maximizes only his monetary payoff, his best response is always to accept any offer and to choose the lowest possible effort level.

Gächter and Falk studied two versions of this game. In the one-shot experiment, the parties were informed that they would never play against each other again. In the repeated-game experiment, the parties knew that they would play ten times. Reciprocity, or a significant and positive wage-effort relationship, was found in both experiments. Consistent with game-theoretic analysis, reciprocity and incentives provided by repeated interactions seemed to complement

[12] For similar results in gift exchange games, see Gächter and Falk (2002). Their findings undermine the suggestion that individuals exhibit dispositional social preferences because they mistake the one-shot laboratory experiments with repeated, real-life situations.

each other. The positive wage-effort relationship was steeper and effort levels higher in the repeated-game treatment. About half of the individuals who revealed themselves as selfish in the last period by providing the selfish amount of labor imitated the reciprocators in all other periods of the repeated-game experiment.[13] Individuals act "as if" they are of a particular type in order to cause others to identify them as such.

Experimental evidence thus lends support to the claim that individuals are rational, in the sense of having stable preferences and being motivated by the consequences of their actions. They behave strategically, trying to anticipate others' responses to their actions, adjusting their responses to others' actions, and using backward induction.[14] The experimental evidence reinforces the view of the great sociologist, Talcott Parsons, that "action remains rational in the sense that it comprises the quasi-intentional pursuit of gratification by reasoning humans who balance complex and multifaceted evaluative criteria" (DiMaggio and Powell 1991a, p. 17).

These experimental results fit well with recent empirical findings in institutional sociology. In facing new situations, individuals actively seek to improve their lot. DiMaggio and Powell note, for example, that "early adopters of organizational innovations are commonly driven by a desire to improve performance" (1991b, p. 65). At the same time, they emphasize the importance of mimetic behavior: individuals mimic the behavior of others in situations in which institutions generate behavior. This response is consistent with the argument developed in Chapter 7 that individuals with social propensities act rationally when facing a new situation, but that once an institutionalized equilibrium behavior establishes itself, each individual best acts mimetically.

Individuals have the propensity to respond to social and moral considerations. Yet, as the experimental and sociological evidence indicates, even such individuals have stable preferences regarding outcomes and act strategically to achieve them.

[13] For similar findings, see Fischbacher, Gächter, and Fehr (2001).

[14] Lindbeck (1997) elaborates on why it is appropriate to assume that individuals act rationally given the values they have internalized.