Routledge
Taylor & Francis Group

Journal of Economic Methodology

ISSN: 1350-178X (Print) 1469-9427 (Online) Journal homepage: http://www.tandfonline.com/loi/rjec20

Behavioral experiments: how and what can we learn about human behavior

Ana C. Santos

To cite this article: Ana C. Santos (2009) Behavioral experiments: how and what can we learn about human behavior, Journal of Economic Methodology, 16:1, 71-88, DOI: 10.1080/13501780802684278

To link to this article: https://doi.org/10.1080/13501780802684278

Journal of Economic Methodology,
Vol. 16, No. 1, March 2009, 71–88

Routledge
Taylor & Francis Group

Behavioral experiments: how and what can we learn about human behavior

Ana C. Santos*

Centre for Social Studies, School of Economics, University of Coimbra, Portugal

(Received 14 July 2007; accepted 12 May 2008)

This paper addresses the experimental trade-off between the exercise of control over the actions of the experimental participants and the potential to provide understanding about human behavior. Control is a requirement of the experimental method to produce pertinent and intelligible results for scientific inquiry. But the more control is exercised the more the experimental results are the outcome of economists' actions. Economic experiments must therefore achieve a difficult balance. They must elicit intelligible behavior while ensuring that the actions of those taking part in the experiment are not determined by the design set-up and the rules of the experiment. The paper puts forward criteria for the analysis of the level of participation of experimental subjects in economics experiments and applies them to several experiments to illustrate the relevance of assessing the agency of experimental participants when deriving inferences about human behavior.

Keywords: experimental economics; ultimatum game; fairness; individual motivations

1 Introduction

Experimenting in the human sciences is informed by the trade-off between the exercise of control over the actions of the experimental participants and the potential to provide understanding about human behavior. Control is a requirement of the experimental method to produce pertinent and intelligible results for scientific inquiry. But the more control is exercised the more the experimental results are the outcome of economists' actions. Economic experiments must therefore achieve a difficult balance. They must elicit intelligible behavior while ensuring that the actions of those taking part in the experiment are not determined by the design set-up and the rules of the experiment.

The present paper puts forward a framework for the analysis of the level of human agency in economic experiments. The framework aims at evaluating the extent to which the behavioral patterns observed in the laboratory are to be attributed to participants' traits or instead to economists' actions. The goal is to contribute to a more informed use of the results of experimental economics, especially so by non-experimental economists who might be more prone to a hasty use of experimental results. This is, indeed, an ever-pressing issue given that experimental economics is now an established field of economics whose results are fueling, in a more systematic way, the theoretical work of the discipline. The recently

*Email: anasantos@ces.uc.pt

ISSN 1350-178X print/ISSN 1469-9427 online

DOI: 10.1080/13501780802684278
http://www.informaworld.com

acquired scientific allure, especially following the 2002 award of the Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel to Vernon Smith 'for having established laboratory experiments as a tool in empirical economic analysis',[1] increases the risk of an uncritical use of 'intuitive' and 'sexy' experimental results (cf. Rubinstein 2006, p. 6).

The paper is structured as follows. Section 2 presents the experimental trade-off between control and human agency. Section 3 defines three criteria to assess the level of human agency in economic experiments, or the extent to which an economic experiment can be considered a *behavioral experiment*, that is, an experiment from which inferences can be made about the motivational factors underlying human behavior. The following three sections present the ultimatum game experiment, the one-shot experiment and the market game experiment, respectively, and analyze their levels of human agency. Section 7 derives the main conclusions and final remarks follow in the last section of the paper.

2 Experimental practices and human agency in economics

The trade-off between control and human agency pervades methodological discussions in the discipline and across experimental fields. Mary Morgan, in a comparative assessment between models and experiments, has explicitly noted the importance of posing the question: 'Where is the potential for independent action in the experiment?' The reason is that the exercise of control may jeopardize the relative epistemic superiority of laboratory experiments that stems from the participation of human subjects. These controls 'raise the danger of over-taming the participants in the particular way so that participants are no longer domesticated, but agents whose behaviour is directed by models of the world, models dictated by the economist' (2005, p. 325).

In fact, the high level of control imposed by the procedures of experimental economics has been taken as a reason for dismissing the relevance of the entire experimental enterprise or, at best, its scope. Nikos Siakantaris, when discussing the generalizability of experimental results to non-laboratory situations (i.e. parallelism) noted that the 'better experimental economists do their job in controlling variables, the more they are threatened by a lack of parallelism and hence of the usefulness of their project'. The 'situations of relative isolation', like those obtained by experimental means, 'are the exception rather than the rule in economic life' (2000, pp. 273–274).[2]

On an extensive discussion of the experimental practices of economics and psychology, held in the journal *Behavioral and Brain Sciences*, Hertwig and Ortmann (2001) classified the procedures of experimental economics as 'regulatory' practices which contrasted to those of psychology which were deemed comparatively 'laissez-faire'. As the discussion made clear, the mandatory practices of economics are highly controversial in psychology. They are perceived by the psychologist critic as illegitimate procedures that extricate rational behavior by '"beating subjects over the head" through constant repetition, feedback, complete and detailed information, and anonymity'. Thus, it cannot come as a surprise that 'the subjects act in accordance with standard "economic" (i.e. Nash) theory when the experimental situation is arranged in this manner' (Dawes 1999, p. 23).

The trade-off between control and human agency is not only discussed by outsiders. It also informs the practice and the debate within the community of experimental economists. The scrutiny of experimental results inescapably hinges on control issues and so does controversy in the field. The most common charge is that the experimental economist dismissed some crucial variable that was not adequately controlled for. This is particularly evident when economists fail to elicit rational behavior or to confirm the predictions of economic theory that presuppose rationality. The experimental set-up is then accountable for not having provided subjects the opportunity to behave rationally. Underlying these criticisms is then a demand for new controlled experiments.

2.1 The experimental practices of economics

The 'regulatory' procedures of experimental economics aim at creating 'micro-economic systems' in the laboratory. Following the Nobel Laureate Vernon Smith (1982), who had a major role in establishing the standard procedures for conducting experiments in economics, a micro-economic system is made up of two component parts: the *environment* and the *institution*. The environment specifies each agent's preferences, commodity endowments and technology. The institution defines the set of admissible messages that individuals may use in communication and the rules under which agents communicate and exchange commodities for the purpose of modifying the initial endowments in accordance with their preferences. The environment and the institution are thus the control variables of an economics experiment on the basis of which the behavior observed in the laboratory is interpreted. The institution is easily controlled by defining the admissible messages and the rules of communication and exchange. Controlling the environment is more complicated, however. The environment is controlled by designing a reward structure that satisfies 'the set of sufficient conditions for a valid controlled microeconomic experiment' – nonsatiation, saliency, dominance and privacy (Smith 1982, pp. 931–935). That is, the reward structure must pay a non-negligible amount (nonsatiation) that must be related to the outcomes of individual actions (saliency). Such a reward structure must also be capable of offsetting the subjective costs, or values, associated with the process of making and executing decisions (dominance) and, finally, the rewards earned must not be known to other subjects to force them to be concerned only about their own payoffs (privacy). The precepts are meant to ensure that the experimental problem is understood by the participants as intended by the economists. And this means that the rewards earned by the participants are significant enough to be worth pursuing and that they understand how the rewards are linked to the decisions they make so that they constitute the main motivation. To this end, the problem must be also described in simple, neutral and abstract terms so as to avoid error and the interference of unintended misconceptions about the context of interaction. Deceiving is also prohibited in economics to prevent second-guessing about the purpose of the experiment.[3]

From this brief description it comes out very straightforwardly that the 'regulatory' procedures of experimental economics attempt to induce self-interest by rendering as a salient course of action the pursuit of monetary gains that are contingent on individual performance, and by omitting the impact of individual actions on others. But it should be noted that the inducement of self-interest is

simply a methodological procedure that aims at ensuring that the behavior of the participants can be interpreted by reference to the design of the experiment.

Control is a methodological requirement that addresses the inescapable underdetermination problem that arises whenever attempting to make inferences from data to theory. The goal is to make sure that the initial conditions and the auxiliary assumptions hold in the experimental setting (e.g. ensure the dominance of self-interested motivations) so that the observations can be brought to bear on the hypotheses under scrutiny (e.g. that markets clear), either confirming or disconfirming them.[4] When variables are not controlled, they may interfere with and influence the outcome of the experiment which comprises the interpretation of the experimental results.

It is therefore not surprising that experimental participants conform to the rational choice model (i.e. attempt to maximize experimental payoffs) in environments in which they face simple problems and have the opportunity to learn the incentive structure of the economic situation. But neither should it be surprising that experimental participants do not always behave like rational men or women. Strict adherence to the methodological prescriptions of experimental economics does not ensure that individuals behave in a self-interested manner or that they succeed in taking the course of action that best suits their interests. Subjects' motivations are multiple and their cognitive limitations may prevent them from perceiving and pursuing the 'best' course of action. When participants deviate from the rational model in a controlled experiment they generate 'anomalies', i.e. facts unexplained in the light of theory.[5] They indicate the presence of other-regarding motives, if subjects reveal concern about the well-being of others, or indicate subjects' cognitive limitations that prevent the maximization of the monetary payoffs.

From this it follows that economics experiments generate knowledge about the contexts in which individuals behave in a selfish and rational manner or about those in which they are moved by other-regarding considerations and prone to mistakes. This is eloquently depicted by the experimental economist John Ledyard in the case of the public goods experiments:

> It is possible to provide an environment in which at least 90% of subjects will become selfish Nash Players. Heterogeneous payoffs and resources, complete and detailed information, particularly about the heterogeneity, anonymity from others and the experimenter, repetition and experience, and low marginal payoffs will all cause a reduction in rates of contribution, especially with small numbers. Add unanimity to the mechanism and rates will go to zero. It is possible to extinguish any trace of 'altruism' in the lab. [But] [i]t is [equally] possible to provide an environment in which almost all of the subjects contribute toward the group interest. Homogeneous interest, little or rough information, face-to-face discussions in small groups, no experience, small numbers and high marginal payoffs from contributing will all cause an increase in contributions. (Ledyard quoted in Dawes 1999, p. 23)

The extreme circumstances in which participants behave like rational economic agents show that only under very stringent conditions do subjects become *homo economicus*. Because rational economic men and women are more easily found in lab markets, economists tend to believe that they are also pervasive in real markets where economic agents have the right incentives and sufficient experience to learn how to behave rationally. They then conclude that economic theory generally applies to the contexts where it is supposed to apply – markets – while dismissing the significance of the 'anomalies'. These tend to be found in contexts where economic

theory is not supposed to apply anyway – non-market contexts where economic incentives are weak and opportunities for learning are insufficient. As Ken Binmore eloquently put it:

> [E]conomic theory should only be expected to predict in the laboratory if the following three criteria are satisfied:
>
> - The problem the subjects face is not only 'reasonably' simple in itself, but is framed so it seems simple to the subjects;
> - The incentives provided are 'adequate';
> - The time allowed for trial-and-error adjustment is 'sufficient'. (1999, p. F17)

Whether or not economic theory applies to the laboratory or to real world markets is, of course, an empirical matter. Nonetheless, the belief that under the 'right' conditions individuals behave rationally has led to an overly excessive concentration of resources devoted to eliminating the anomalies of economic theory (Loomes 1991; Loewenstein 1999; Starmer 1999). By improving the incentive structure of the experimental task and by conceding participants enough time to learn it, economists have been able, on some occasions, to reduce the significance of the anomalous behavior.[6] But on the way they seem to neglect the effort undertaken to have experimental participants behaving like rational economic men.

2.2 Human agency in economics experiments

The fact that economists are equally able to generate behavior that conforms to and conflicts with the model of rational economic men does not undermine the experimental enterprise. As mentioned above, economics experiments can fruitfully shed light on what it takes for individuals to become rational economic men and what prevents them from being so. But when experiments are used to study the motivational factors underlying human behavior, the behavioral patterns observed in the lab must be attributed to the participants' values, beliefs, expectations, preferences, and attitudes rather than to the design and rules of the experiment. That is, the experimental set-up must allow room for the participants' agency.

The present paper proposes a framework to evaluate whether the behavioral patterns observed in economic experiments are to be attributed to individual traits or instead to economists' actions. Or to put it in another way, the framework to be presented allows identification of the '*behavioral experiments*' of economics – the experiments from which inferences can be made about individual motivations and behavior. The crucial characteristic of behavioral experiments is the relative high level of agency of experimental participants. Insofar as control is exercised by inducing self-interest and reward-maximizing motives, a high level of human agency entails a higher potential to provide understanding about the factors that trigger other-regarding considerations as well the conditions that affect individual rationality.[7] Given the specificity of the experimental method of economics, the level of 'human agency' is to be assessed at three distinct levels: (1) the range of motivations induced on the experimental subjects; (2) the range of options available to them; and (3) the individual and aggregate outcomes of the actions taken.

To better illustrate how a behavioral experiment is to be identified, the criteria will be applied to a set of experiments conducted by Vesna Prasnikar and Alvin Roth

(1992) that aimed to eradicate the 'anomalous' egalitarian partitions of income observed in the ultimatum game experiment and thereby demonstrate that individuals remain rational agents when self-interest conflicts with other-regarding considerations. These experiments are particularly suitable for the present purposes because their reduced level of human agency does not allow the deriving of claims about the motives of human behavior. Nonetheless, the economists were adamant in concluding that self-interest prevails – subjects follow the strategic opportunities of the games. The ultimatum game experiment is a behavioral experiment. But its high level of human agency produces ambiguous results.

3 How to evaluate 'human agency' in economics experiments

The analysis of the results of behavioral experiments requires special care. On the one hand, if experimenters over-tame subjects' motivations and actions, the experimental results convey the actions of the economists rather than the behavioral attributes of the experimental participants. On the other hand, if experimenters fail to control the laboratory context they will produce ambiguous results. Behavioral experiments must therefore achieve a difficult balance between experimental control and human agency. They must be carefully designed so that the results they generate can be made to bear on subjects' motivational attributes. Given the characteristics of the experimental method described above, it is easy to see that behavioral experiments have a potential to generate knowledge about the motivational factors that obstruct the manifestation of self-interested behavior.

Deriving inferences from the observed behavior of the participants in economics experiments to the underlying motivations is not a straightforward matter. The reason is that there is no univocal relation between a given action and a corresponding motivational factor. The same action may be caused by various and divergent motives.[8] For instance, in the ultimatum game (to be described below) the equal division of a fixed amount of money between two subjects may be explained by reasons other than a preference for equality. It can be explained by the fear of retaliation for violating an expectation or the desire to please others by doing what one is expected to do. It can also be explained by the desire of simply doing the right thing under the circumstances or by the fear of experiencing feelings of guilt for not having done so, and so forth. This is why the interpretation of behavioral experiments often requires the conduction of follow-up experiments that attempt to evaluate the relative contribution of the various motivational factors in operation.

The appraisal of the level of 'human agency' in an economics experiment requires assessing: (1) the range of motivations induced on the experimental subjects; (2) the range of options available to them; and (3) the individual and aggregate outcomes of the actions taken. First, if the experiment aims to derive conclusions about motivational factors it must elicit the relevant set of motives so that they can be manifested in the laboratory. To give an example, an experiment in which dominance and privacy prevail cannot purportedly trigger concerns about the impact of one's actions on the well-being of others. The participant will be simply ignorant of the circumstances of others and the impact of her actions on them. Second, insofar as subjects' motivations in economics experiments are manifested via

the actions taken, the range of options available to experimental subjects must enable their expression. To continue with the same example, other-regarding concerns can only be manifested in the laboratory if the range of options available includes actions that can improve the well-being of others, say the possibility of dividing the gains of the experiment with others. Finally, it should be investigated whether the available actions can bring about the intended outcomes, i.e. those that would be obtained from the relevant motives under scrutiny. If not, the experiment may not elicit those motivations to their full extent. For instance, if the proposal of an equal division of the experimental earnings is not effective in producing this outcome, then subjects may make a different offer and a preference for equality may not manifest itself in the laboratory.

If the purpose of the experiment is to produce knowledge about the motivational factors underlying human behavior, it must elicit the relevant set of motives, it must allow them to be manifested via the actions of the participants and it must render these actions effective in producing intended outcomes. An adequate behavioral experiment is one that allows the expression of human motivations by suitable actions that bring about intended outcomes. To put it yet another way, the proposed criteria assess the strength of inductive inferences from the evidence produced by the experiment (e) to hypothesized conjectures about the motivational factors that account for the observed patterns (H). Now following the eliminative inductive account synthesized by Francesco Guala (2005), the framework put forward assesses whether an inference from e to H comes from a test situation or set-up such that 'the observation of e would be probable if H were true, but unlikely if it were false' (p. 136). It investigates if the observed patterns rather than being explained by individual attributes are instead explained by the experimental set-up. Once it is established that evidence conveys the expression of subjects' attributes, the next step would then consist of discriminating among alternative hypotheses that may entail the same piece of evidence.

The next sections analyze the ultimatum game and two follow-up experiments – the best-shot and the market game experiments – which were purposely designed to infer the motives (H) underlying the fair partitions (e) observed in the ultimatum game. It will be shown that by failing to meet the criteria of behavioral experiments, these experiments also failed to meet their stated goals.

4 The ultimatum game experiment

The ultimatum game experiment (UG), first conducted by Güth, Schmittberger, and Schwarze (1982), is a two-round game that consists of the partition of a fixed amount of money between two subjects, who remain anonymous throughout the game. In the first round, player 1, also known as the 'proposer', proposes a division of this amount between the two. In the second round, player 2, the 'responder', decides whether or not to accept the offer. If she accepts, each receives accordingly, otherwise both receive nothing. Under these conditions a very asymmetric distribution should follow. The game-theoretical prediction, assuming that each subject is rational and nonsatiated with money and that he gets utility from the outcome of the negotiation, is that the proposer should receive the bulk of the fixed amount. He will always prefer the alternative that yields the higher payoff and therefore he will offer the smallest positive payoff. Player 1 will expect that this offer

will be accepted by the responder, who accepts any positive offer rather than rejecting it and earning nothing. The experimental results, however, diverge from the theoretical prediction. They were 'anomalous'. Not only did proposers make more generous offers, but the responders also refused positive payoffs by opting for no rewards.

Because various interpretations were available, a wide range of experiments were subsequently carried out to investigate the role of potentially relevant factors. The results, however, turned out to be fairly robust to varying conditions. The original results of the UG were not significantly affected by subjects' experience, the stakes involved, anonymity between subjects and towards the experimenter, and various demographic factors.[9] The results of the UG have since become well-accepted stylized facts, as summarized here: (1) there are almost no offers below 10% or above 50% of the amount to be distributed; (2) the modal and median offers are 40–50%; (3) the means are around 30–40%; (4) offers of 40–50% are rarely rejected; (5) offers below 20% are rejected about half the time; (6) the rejection rate increases with the decrease of offers.

4.1 Human agency in the UG

It is easy to show the extent to which the ultimatum game is a behavioral experiment. The level of human agency is high and this explains the behavioral patterns that departed from the incentive structure of the experiment. From the brief description given above it is easy to see that this experiment satisfies the three criteria identified above: (1) the range of motives in the ultimatum game are varied; (2) various partitions are feasible, from very asymmetric to the equal division of the fixed amount of money; (3) the decisions of the participants are consequential, i.e. they determine the final outcomes.

The ultimatum game did not constrain or guide in any particular way the motivations of the participants. Self-interested and other-regarding motives may be equally present. The game consists in the division of a sum of money where individuals might possibly want to get the highest share possible. Nonsatiated proposers will then offer a low amount of money which will be accepted by responders, who prefer accepting any positive amount rather than earning nothing. Because the more the proposer keeps for himself, the less remains for the other player, other-regarding considerations might interfere as well. If proposers care about the amount received by the other player, they will make generous offers. By the same token, ungenerous offers will be rejected by the responders that feel aggravated with uneven partitions of income.

The eventual presence of other-regarding motives is strictly related with the impossibility of satisfying the precept of 'privacy'. Privacy is not guaranteed in this experiment because the payoffs of the players are the very object of the decision problem. Because each player knows the payoff of the player they are interacting with, intersubjective considerations cannot be avoided. Subjects are aware that their decisions have an impact on others and that the decisions of others affect them. These intersubjective considerations are further magnified by the fact that the UG consists of a bilateral interaction. In sum, in the UG the payoffs are common knowledge and depend upon the details of the interactive process with another individual.

4.2 Motivations, behavior and context

Rather than being determined by the experimental set-up and rules, in the UG the selection of a given partition depends on how subjects perceive the particular context of interaction and the suitable courses of action in it. Specifically, the experimental results express proposers' views on what an adequate proposal is under the circumstances, as well as responders' views on the adequacy of the offer made which determines their decisions to accept or reject it.

Güth et al. (1982) suggest that the UG creates a situation in which the exploitation of the position of advantage by the proposer is unacceptable. This is the case because the bargaining situation consists of a game between two opponents who would stand in an equal position were it not for the arbitrary allocation of the different roles between them. Under these circumstances, the 50:50 split is the salient distribution. Proposers offer generous partitions and responders reject low offers to punish what is perceived as the exploitation of an undeserved position of advantage. According to the authors, an asymmetric relation would be more acceptable in a market context such as the consumer markets of industrialized countries where 'buyers … might be used to have less strategic power' (1982, p. 369).

Follow-up experiments have subsequently attempted to identify the factors that affect conformity to the norm that prescribes the equal division. Deviations from this norm were observed when the proposer earned the right to make the offer (e.g. by answering to a quiz), which seemed to justify keeping more than the equal share (Hoffman, McCabe, Shachat, and Smith 1994). In contexts where the 'social distance' between subjects was higher subjects tended to offer less generous proposals which were more easily accepted (Bohnet and Frey 1999), the same results obtained in situations framed as market exchanges (Hoffman et al. 1994). When ungenerous offers are imputed to the intentions of another individual, they are more easily rejected than when they are generated by a random device or a third party who does not benefit from them (Blount 1995; Falk, Fehr, and Fischbacher 2003). The UG carried out in 15 small-scale societies in Papua New Guinea, in the Amazon, and in Africa also revealed significant cross-cultural differences (Henrich et al. 2001, 2004), suggesting that sharing norms closer to equal divisions is correlated with cultures with more cooperative patterns of social interactions. Thus, this series of experiments has shown the various ways by which other-regarding considerations interact with the social context. Contexts where social proximity is high and individuals stand in a fairly symmetric relation more easily trigger other-regarding concerns.

It could be argued, then, that the experimental context still determines subjects' motivations and actions. The behavioral diversity observed in these experiments denies it, however. The patterns observed in behavioral experiments are the outcome of both subjects' attributes and the social context recreated in the laboratory. Given the high level of human agency, the ultimatum game experiment *qua* behavioral experiment triggers various motives that account for the various patterns of behavior.

4.3 From actions to motivations: fairness or strategy?

A different range of motives may explain the relatively fair partitions of income. We have seen that the UG could have triggered both self-interest and

other-regarding considerations. On first impression, generous offers seem to be the expression of subjects' caring about others' well-being. But they could also express strategic thinking based on proposers' correct anticipation of a high probability of refusal of extremely low offers. Generous proposals could then have been strategic in the sense that they entailed a higher chance of positive gains. Further experiments were then carried out to find out whether the UG results constituted evidence for strategic reasoning or instead an indication of fairness, i.e. the desire to treat others fairly and punish those who do not behave that way, even at a cost to oneself. In the next two sections, it will be shown that discriminating strategic from fairness motives requires the design of an experiment that satisfies the three criteria identified above.[10]

5 The sequential best-shot game

The view that individual behavior conforms to the rational choice model if individuals have time to learn with experience suggested new experiments to test whether considerations of fairness could be displaced by strategic considerations as experience and understanding were acquired. To this end, Prasnikar and Roth (1992) ran a 10-period UG and best-shot (BS) experiments. The UG is a 10-period experiment which followed the standard two-round game. The BS is a 10-period public good experiment in which player 1 first states the quantity q_1 he wants to supply of a public good, after which player 2, knowing q_1, states the quantity q_2 he is willing to supply. The amount of the public good is then given by the maximum of these two quantities (q_1 or q_2, the best-shot). The incentive structure of the BS dictates that player 1 chooses $q_1=0$ and player 2 chooses $q_2=4$, which yields a payoff of \$3.70 and \$0.42 respectively (this constitutes an extreme distribution in the order of 8,8/1). As expected, fair shares were observed in the UG while extreme distributions were observed in the BS.

These results were interpreted as providing evidence for strategic behavior in that the behavioral patterns conformed to the incentive structure of both games. While in the UG proposers did better (i.e. their payoffs increased) by deviating from the equilibrium (i.e. by increasing their offers), in the BS players did better (i.e. their payoffs increased) by converging towards the equilibrium (i.e. by decreasing their contributions).

Having shown that subjects' behavior accords well with the incentive structure of the experiments, the authors concluded that subjects' decisions in the UG are primarily strategic. Fairness considerations matter because they have strategic value to subjects which is learned by the losses subjects incur when they make low offers. Of course, this interpretation only succeeds in explaining the behavior of player 1. Because subjects changed game partners between periods, no strategic interpretation can explain the rejection of positive offers by player 2.

On closer inspection, one easily demonstrates that the claim that strategic considerations override fairness is not supported by the experimental results. The BS is not a proper behavioral experiment. The BS tipped the scale in favor of strategic behavior by constraining the attainment of the equal split and thence the manifestation of fairness. In this experiment, the equal contribution to the public good was inconsequential because subjects could not share the cost of public provision. To any positive offer by player 1, the equal provision by player 2 would

worsen his payoff while it would not improve the payoff of player 1. Under these circumstances, the best response of player 2 is always to provide a zero quantity of the public good. Player 1, who had the first-mover advantage, chooses not to contribute and thereby benefit himself from the unequal distribution.[11] Because the cost of the public good could not be shared between the two players, the central issue the BS raised concerned the selection of the beneficiary of the unequal distribution. Thus, the unequal distribution of income observed in the BS is most adequately attributed to a feature of the experimental design, rather than to subjects' motivations.

In the BS experiment subjects could have had fairness concerns and therefore they could have chosen to share the burden of providing the public good if it improved the situation of them both. But the structure of the game inflicted an unnecessary cost on one of the players. Nonetheless, there is some evidence for fairness concerns given that some players 2 did resist unequal distributions by opting for a zero provision of the public good when player 1 did not contribute. But the structure of the game ultimately led players 2 to accept the unequal outcome by being the sole providers of the public good. This means that this experiment does not satisfy the third criterion of a behavioral experiment. In the BS subjects could have had fairness concerns and they could have chosen the equal distribution of income. The problem is that the attainment of the equal distribution is too demanding in that it requires on the part of one subject willingness to support an unnecessary cost for the provision of the public good. The claim that strategic considerations overrode fairness considerations is therefore not warranted. The best-shot experiment is not an adequate behavioral experiment.

6 The sequential market game experiment

The inefficiency of the BS experiment was identified by Güth and Tietz who explicitly noted that '[i]f sharing the burden of providing the public good is impossible, fairness considerations cannot be applied' (1990, p. 428). In response, Prasnikar and Roth designed a 10-period market game experiment (MG) with extreme equilibrium predictions but in which equality is compatible with efficiency (i.e. the total payoff is always distributed between the two players). In this market nine buyers compete for the acquisition of one unit of a good provided by a single seller. The buyers tender bids up to \$10 (the redemption value of the good) and the seller either accepts or rejects the highest bid (if several, one of them is selected randomly with equal probability). If the seller accepts it, he or she receives the corresponding amount and the successful buyer the difference between the \$10 and the bid price. The other buyers do not earn anything. If the seller rejects it, all players receive zero earnings. The equilibrium prediction is that buyers bid approximately the maximum amount which is accepted by the seller. This is the case because buyers do not have the chance of getting the market commodity with lower bids and the seller never rejects the highest bid. The theoretical prediction is thus the extreme distribution which is also an efficient outcome.

The results of the MG converged to the equilibrium prediction and were also consistent with a positive relation between the equilibrium trajectory and the incentive structure of the experiment. That is, buyers improved their earnings by converging to the equilibrium and so the experimenters concluded:

> Taken together, these results suggest that although equilibrium predictions may need to be modified to take into account nonmonetary aspects of players' preferences (e.g., in the ultimatum games), nevertheless, even when equilibrium yields very unequal payoffs, strategic considerations are not *displaced* by considerations of equity. On the contrary, the best-shot and market games show that whether equilibrium will be observed depends on the off-the-equilibrium-path behavior, which responds to the off-equilibrium-path incentives. (Prasnikar and Roth 1992, p. 886 emphasis in original)

But again the results are not as self-evident as Prasnikar and Roth suggest. It is not the case that in the face of conflicting motives, individuals opt for self-interest. Even though the equal split is an efficient outcome, the design of the market game experiment simply prevented it. This means that the MG experiment did not satisfy the third criterion for 'human agency'. In the MG buyers could have had fairness motives that could result in the choice of a more balanced distribution of income. However, the $5 bid (or one close to it) could not succeed in the market. Hence, this is not an adequate behavioral experiment from which inferences about motivational factors can be derived.

The problem was that sellers could only choose between accepting the highest proposal or rejecting it. They could not select fair partitions even if they were proposed. The implementation of the equal distribution would require that buyers all had a preference for the equal division or that they could coordinate and agree not to tender bids above $5. But this was excluded at the outset by the design of the experiment because subjects could not communicate with one another. In fact, they did not even know with whom they were interacting. In the absence of communication, the attainment of the equal partition would require that *all* subjects shared the belief that this was the legitimate bid and that they *all* believed that others had this expectation too. Only then could they be confident that the $5 bid would be the winning bid and thereby determine the equal split of the payoffs. But this is not a credible expectation. As a wide range of experiments have shown (including the UG) individuals' expectations, beliefs and preferences are not homogeneous.

The implementation of the equal distribution of income between the winning buyer and the seller would require coordination among the buyers because none of them in isolation had sufficient bargaining power to impose it. In contrast to the UG, where the proposer could define the income distribution, in the MG the winning bid was determined by the interaction of nine players who competed among themselves. In the face of competition, buyers were compelled to tender attractive bids to raise their chance of winning. The implication of this is that the behavior of buyers cannot reflect subjects' views on what an adequate partition of income is between two individuals. It is more plausibly affected by the competition among subjects and their desire to be the winning bidder or their wish to prevent others from being so. Thus, the competition between subjects might have triggered other motivational factors not present in other experiments. Rather than raising the chance of being the winning bidder and earning an incredibly small amount of money, the escalation of bids can be explained by feelings of resentment and retaliation towards other players. The fact that buyers were willing to bid the maximum amount, and earn nothing by so doing, suggests just that. Finally, the framing of the game as a market situation may have also changed the perceptions of the experimental setting rendering inequality between buyers and sellers more acceptable, especially so when individuals compete for a scarce good. In contrast, the UG and the BS experiments consist in the resolution of a distributive problem where

obvious other-regarding considerations emerge. To conclude, not only does the MG render the attainment of equal payoffs unfeasible, the outcomes actually arrived at might be explained by a different range of motives than fairness or strategic considerations. Quite perplexing, this is recognized by the experimenters:

> Consider a hypothetical buyer whose preference for equality is such that his very first choice outcome would be to have all buyers submit identical bids of $5 (or $1), and who bids accordingly in the first two rounds. When he sees how high the actual transaction price is, he becomes annoyed with the other buyers, and (with the same motivation that would have caused him to express his displeasure by rejecting too small an offer if he were a seller in the ultimatum game) he decides to become the high bidder in round 3, in order to deprive other buyers of the benefits of what he sees as their unreasonable behavior. The point in considering such a hypothetical buyer is to observe that in *this* game his nonmonetary preferences cause him to behave in a manner indistinguishable from an income maximizer, while in the ultimatum game his preferences lead away from the equilibrium predicted for income maximizers. (Prasnikar and Roth 1992, p. 885, emphasis in original)

The experimenters overlook the fact that the extreme distributions of income are to be explained by experimental design that renders the equal split simply unattainable. Therefore, the assertion that strategic motives displace fairness considerations in the market game experiment is not warranted. Fair actions were simply inconsequential. Nonetheless, and similarly to the BS, there is some evidence that these motives were also present in this experiment. The $5 bid was the modal bid in four out of 20 market rounds (including rounds 7 and 10 in which players had already acquired enough experience to understand that these could not be winning bids). This indicates that at times individuals like to express their preferences even when it is inconsequential. In the MG they could do that because conditions 1 and 2 for 'human agency' were satisfied. Because criterion 3 was not, it is plausible that the manifestation of fairness motives via the choice of the equal split was severely constrained.

7 Motivations, actions, outcomes, and theories

The application of the three criteria to the ultimatum game and other follow-up experiments demonstrates the importance of assessing the level of 'human agency' when deriving inferences about the motives underlying human behavior. The elimination of other-regarding behavior by taming subjects' motivations and by limiting the range of available options and outcomes is not very compelling. It only shows how a particular pattern of behavior can be eradicated in particular circumstances.

The fact that economic experiments involve the participation of human subjects does not warrant the conclusion that the observed behavior is the result of subjects' attributes. This has to be carefully scrutinized. This scrutiny requires assessing the range of motives present, whether subjects are given the opportunity to express them, and whether the actual consequences of subjects' actions adequately convey subjects' motives.

We have seen that neither the BS nor the MG experiments are adequate designs to confront fairness with strategic motives. In these experiments, unequal income distributions are to be attributed to the institutions of the micro-economic system that rendered the equal partition an inefficient or unattainable outcome. In contrast,

in the UG fair distributions were both efficient and feasible. Experiments have also shown that what fairness amounts to depends on the context of interaction. The context provides clues that help subjects interpret the social situation and thereby select the most adequate course of action in it. Conformity to this norm might be explained by various motives – doing the right thing, doing what is expected, avoiding social ostracism, and so forth. Discriminating among the various underlying motives, in turn, requires the conduction of other behavioral experiments.

The analysis undertaken here therefore stresses that the relationship between motivations, actions, and outcomes must be scrutinized before inferences be made about the motives of individual behavior. This is particularly relevant given that the results of experiments are now the raw material for the construction of new theories of human behavior, which detaches the experimental results from the processes that produced them.

The theory of fairness developed by Fehr and Schmidt (1999) is representative of that. This theory, which is based on the experiments reviewed above (and others), represents an attempt to organize within a 'coherent framework' competitive and cooperative behaviors both of which can be explained by 'self-centered' inequity aversion, i.e. an aversion towards inequitable distributions of payoffs that is stronger when these distributions are unfavorable to the individual. This theory then predicts the prevailing behavior in equilibrium via the interplay between the micro-economic institution and the distribution of inequity aversion in the population. If the majority of the individuals within the population are selfish then the prediction will approximate traditional theory, but if in the population individuals care a lot about equity, then the theory tends to predict egalitarian outcomes. When applying the model to the MG (with both proposer and responder competition), the authors conclude:

> The crucial observation in this game is that no single player can enforce an equitable outcome. Given that there will be inequality anyway, each proposer has a strong incentive to outbid his competitors in order to turn part of the inequality to his advantage and to increase his own monetary payoff. A similar force is at work in the market game with responder competition. As long as there is at least one responder who accepts everything, no other responder can prevent an inequitable outcome. Therefore, even very inequity-averse responders try to turn part of the unavoidable inequality into inequality to their advantage by accepting low offers. *It is, thus, the impossibility of preventing inequitable outcomes by individual players that renders inequity aversion unimportant in equilibrium*. (Fehr and Schmidt 1999, p. 834, emphasis added)

Fehr and Schmidt's reading of the results seems to support the view that in heterogeneous populations, fairness is relevant when inequality-averse individuals have sufficient bargaining power to inflict costs on selfish individuals. Fair behavior is then enforced on selfish individuals because it is in their self-interest to behave fairly. In the ultimatum game responders had such power because they could impose a cost on proposers by rejecting extremely low offers, which they could avoid by making more generous offers. In the MG this possibility is not available.[12] This reading is reinforced when the authors, extrapolating these results to real world environments, venture that 'this suggests that fairness plays a smaller role in most of market goods than in labor markets' (p. 835). While in labor markets, where workers have some discretion over their work level and thence '[b]y varying their effort they can exert a direct impact on the relative material payoff of the employer',

'[c]onsumers, in contrast, have no similar option available' (p. 835). But from this it does not follow that markets are ruled solely by strategic considerations. The fact that fairness considerations were ineffective in experimental settings in which fair actions were not available or were inconsequential cannot be overlooked. Whether in the lab or in the real world, claims about motives require assessment of the relationship between motivations, actions, and outcomes.

The foregoing analysis of the level of human agency of economics experiments shows that laboratory market contexts tend to trigger and legitimize self-interested behavior. Non-market contexts where social proximity is high tend to trigger other-regarding concerns and raise the expectation that social interactions should be guided by shared norms of conduct. It also shows that laboratory markets are contexts that induce self-interest and income-maximizing behavior by constraining the motivations and the actions of the experimental subjects, as well as the resulting outcomes. The market experiments are particularly successful in doing that by:

(1) framing the experimental situation as a social context in which self-interest and income-maximization is the salient norm of conduct;
(2) reducing social proximity between subjects and thereby inhibiting other-regarding considerations;
(3) reducing the range of options available to experimental subjects;
(4) rendering individual preferences, motivations, and actions irrelevant to the resulting outcomes.[13]

The exercise carried out here therefore suggests caution when using experimental markets to draw general conclusions about the motives underlying human behavior. The extreme experimental conditions in which behavior solely guided by self-interest is observed recommends that. In real world environments the variety of human motivations and the range of options available to individuals are much larger. This is also the case of real world markets in which individuals may be desirous of and capable of acting fairly.

8 Final remarks

This paper provides a cautionary warning about the use of the results of experiments to make claims about the motivational factors underlying human behavior. Such claims can only be obtained from behavioral experiments that establish a tight correlation between motives, actions, and outcomes. When any of these partial relations does not hold, the inferences about individual values, beliefs, expectations, preferences, and attitudes are not warranted. An appropriate behavioral experiment is one that allows the expression of human motivations by suitable actions that bring about intended outcomes.

It may seem that lying beneath the exercise carried out here is the presupposition that human beings are instrumentally rational in the sense that their actions are always targeted at particular intended outcomes, whichever these may be. Such presupposition may be problematic when applied to behavioral experiments that aim at uncovering the various and heterogeneous motivational factors underlying human behavior, which should also encompass actions that possess an intrinsic value in themselves because doing them is the envisaged reward. They should also include

beliefs, expectations, and preferences about the processes whereby the final outcomes come about.

Though the exercise is built on the analysis of the relation between motives, actions, and outcomes, the instrumentalist reading is not implied by it. This is an inescapable feature of the experimental method. Inferences from economic experiments are inevitably based on the observed actions of the participants and resulting outcomes. This is, in effect, presented as a strength of the experimental method as compared with other empirical methods. Experiments allow observing how people behave when confronted with a concrete situation of interest, which was purposely prepared to that end, as opposed to assessing what they say they would do in such and such circumstances. At any rate, the analysis of the potential motives in confrontation, the actions available, the actions taken, and their actual outcomes provide valuable information about the various motives of human action. The analysis of the follow-up experiments to the ultimatum game indeed showed that at points individuals like to express their values, beliefs and preferences even when it is inconsequential to do so. The point is simply that in experimental economics the motives underlying human behavior are most clearly expressed by actions that have material consequences to oneself or to others.

Acknowledgements

I thank Arthur Schram and the audience of the EIPE Research Seminar (held in Rotterdam, November, 13, 2006), João Rodrigues, John B. Davis and two anonymous referees, for their helpful comments and suggestions. Usual disclaimers apply.

Notes

1. *http://nobelprize.org/nobel_prizes/economics/laureates/2002/press.html*.
2. Siakantaris' analysis of the trade-off between control and parallelism takes parallelism as a requirement for the validity of economics experiments. The analysis undertaken here is less demanding. As will become apparent, the view to be presented considers that economics experiments can provide understanding about human behavior regardless of whether experimental results apply elsewhere.
3. Even though these were originally meant as sufficient (not necessary) conditions for controlling an economic experiment, they have become the standard procedures of experimental economics. As will be shown below, complete conformity to these prescriptions depends on the purpose of the experiment (especially the privacy condition).
4. For a more detailed account of the underdetermination problem in the context of experimental economics see Guala (2005, chap. 3).
5. The growing accumulation of empirical results that cannot be explained 'by assuming that agents have stable, well-defined preferences and make rational choices consistent with those preferences in markets that (eventually) clear' was the motivations for the column 'anomalies' in the *Journal of Economic Perspectives* (e.g. Thaler 1988).
6. In an extensive survey Camerer and Hogarth (1999) show that the capacity of financial incentives and learning opportunities to induce rational behavior is not conclusive.
7. The *behavioral experiments* contrast with the *technological experiments* of economics. While the former provide room for human agency so as to produce knowledge about individual motivations and behavior, the latter control the actions of the individuals so as to neutralize the impact of individual idiosyncrasies on to the performance of the micro-economic system. For a more thorough characterization of these experiments see (reference omitted).

8. This constitutes a particular instance of the problem of underdetermination. The hypothesized conjectures about the motivations of the experimental participants are underdetermined by the behavioral patterns produced by economics experiments.
9. For a recent survey of the main results of the ultimatum games see Camerer (2003, chap. 2).
10. Other explanations were available and other experiments were conducted to investigate whether the experimental results conveyed manifestations of altruism, aversion to inequality, reciprocity, etc. (see Camerer 2003, chap. 2). For the purposes of the present argument it suffices that I focus on the strategic and fairness explanations.
11. When player 1 chooses $q_1=4$ we will always get \$0.42, whether player 2 chooses $q_2=0$ or $q_2=4$. Though $q_2=0$ yields an income of \$3.7 and $q_2=4$ yields an income of \$0.42 to player 2. The equal contribution ($q_i=q_1=q_2$) that yielded positive payoffs ($w_i=w_1=w_2$) were: (q_i; w_i)=(4, \$0.42) (5, \$0.4), (3, \$0.39), (6, \$0.33), (2, \$0.31); (7, \$0.21), (1, \$0.18), (8, \$0.04).
12. This result has been replicated in other experiments such as the public goods experiments in which the possibility of punishing open to individuals holding social preferences is effective in forcing cooperative behavior on selfish rational individuals (cf. Fehr and Gächter 2000).
13. See Guala (2001) for an excellent illustration of such a market experiment.

References

Binmore, K. (1999), 'Why Experiment in Economics?' *Economic Journal*, 109, F16–F24.

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.

Bohnet, I., and Frey, B.S. (1999), 'Social Distance and Other-Regarding Behavior in Dictator Games: Comment', *American Economic Review*, 89, 335–339.

Camerer, C.F. (2003), *Behavioral Game Theory: Experiments in Strategic Interaction*, Princeton: Princeton University Press.

Camerer, C.F., and Hogarth, R.M. (1999), 'The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework', *Journal of Risk and Uncertainty*, 19, 7–42.

Dawes, R.M. (1999), 'Experimental Demand, Clear Incentives, Both, or Neither?' in *Games and Human Behavior: Essays in Honor of Amnon Rapoport*, eds. D.V. Budescu, I. Erev, and R. Zwick, Mahwah, NJ: Laurence Erlbaum Associates, pp. 21–28.

Falk, A., Fehr, E., and Fischbacher, U. (2003), 'On the Nature of Fair Behavior', *Economic Inquiry*, 41, 20–26.

Fehr, E., and Schmidt, K.M. (1999), 'A Theory of Fairness, Competition, and Cooperation', *Quarterly Journal of Economics*, 114, 817–868.

Fehr, E., and Gächter, S. (2000), 'Cooperation and Punishment in Public Goods Experiments', *American Economic Review*, 90, 980–994.

Guala, F. (2001), 'Building Economic Machines: The FCC Auctions', *Studies in History and Philosophy of Science*, 32, 453–477.

——— (2005), *The Methodology of Experimental Economics*, New York: Cambridge University Press.

Güth, W., and Tietz, R. (1990), 'Ultimatum Bargaining Behavior: A Survey and Comparison of Experimental Results', *Journal of Economic Psychology*, 11, 417–449.

Güth, W., Schmittberger, R., and Schwarze, B. (1982), 'An Experimental Analysis of Ultimatum Bargaining', *Journal of Economic Behavior and Organization*, 3, 367–388.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., and McElreath, R. (2001), 'In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies', *American Economic Review*, 91, 73–78.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., and Gintis, H. (eds.) (2004), *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*, Oxford: Oxford University Press.

Hertwig, R., and Ortmann, A. (2001), 'Experimental Practices in Economics: a Methodological Challenge for Psychologists?' *Behavioral and Brain Sciences*, 24, 383–451.

Hoffman, E., McCabe, K., Shachat, K., and Smith, V.L. (1994), 'Preferences, Property Rights and Anonymity in Bargaining Games', *Games and Economic Behavior*, 7, 346–380.

Loewenstein, G. (1999), 'Experimental Economics From the Vantage-Point of Behavioural Economics', *Economic Journal*, 109, F25–F34.

Loomes, G. (1991), 'Experimental Methods in Economics', in *Companion to Contemporary Economic Thought*, eds. D. Greenaway, M. Bleaney, and I. Stewart, London: Routledge, pp. 593–613.

Morgan, M. (2005), 'Experiments Versus Models: New Phenomena, Inference and Surprise', *Journal of Economic Methodology*, 12, 317–329.

Prasnikar, V., and Roth, A.E. (1992), 'Considerations of Fairness and Strategy: Experimental Data From Sequential Games', *Quarterly Journal of Economics*, 107, 865–888.

Rubinstein, A. (2006), 'Discussion of Behavioral Economics'. *http://arielrubinstein.tau.ac.il/papers/behavioral-economics.pdf*.

Santos, A.C. (2007), 'The 'materials' of experimental economics: technological versus behavioral experiments', *Journal of Economic Methodology*, 14, 311–337.

Siakantaris, N. (2000), 'Experimental Economics Under the Microscope', *Cambridge Journal of Economics*, 24, 267–281.

Smith, V.L. (1982), 'Microeconomic Systems as an Experimental Science', *American Economic Review*, 72, 923–955.

Starmer, C. (1999), 'Experiments in Economics: Should We Trust the Dismal Scientists in White Coats?' *Journal of Economic Methodology*, 6, 1–30.

Thaler, R.H. (1988), 'Anomalies: The Ultimatum Game', *Journal of Economic Perspectives*, 2, 195–206.