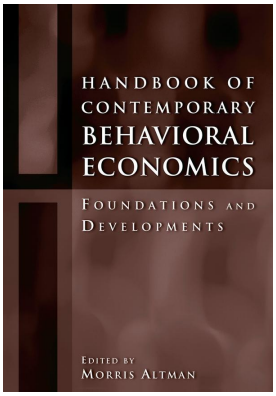

Handbook of Contemporary Behavioral Economics Foundations and Developments

Morris Altman

Experiments and Behavioral Economics

Publication details
https://test.routledgehandbooks.com/doi/10.4324/9781315703879.ch22

Robert J. Oxoby

Published online on: 15 Jul 2006

CHAPTER 22

EXPERIMENTS AND BEHAVIORAL ECONOMICS

ROBERT J. OXOBY

Experimental methods are now considered an important part of economic research. This should come as no surprise: for a field so closely aligned with psychology in its interest in individual behavior, experimental methods are a natural (and some would argue necessary) tool. Concurrently, the "second wave" of research in behavioral economics (Rabin 1998, 2002) has brought recognition to the value of incorporating psychological insights into economic theory. The implications of these insights are becoming increasingly important in enriching (and invigorating) economic theory and informing policy debates.

As a result, economists have been actively using experimental methods, the traditional methodology of psychologists. Following the reasoning of others (e.g., Lazear 2000), the strength of economists' theoretical methodology provides the opportunity and ability to pursue research questions traditionally considered outside the purview of economics. Indeed, methodological individualism and mathematical formalism provide economics with an advantage over other social sciences in tractably identifying the assumptions that underlie human behavior. These advantages make economics, in many ways, an ideal realm for experimental methods. The clear definition of assumptions provides researchers with formal refutable hypotheses that can be directly tested in laboratory environments.

That said, the rapid growth in the application of experimental methods in economics and the increasing focus on behavioral issues brings a strong need to reevaluate experimental methodology as applied in economics (and other disciplines, for that matter). In this chapter, we review some of the basic elements of the experimental methods employed in economics and critically examine how economists conduct experiments. Our attention here is on how the research conducted by behavioral economists may be compromised by some of the experimental methods currently employed in economics. Thus our intent is not to develop a manual of how to conduct an experiment (interested readers are referred to Friedman and Sunder 1994; Davis and Holt 1993; Aronson, Wilson, and Brewer 1998). Rather, we raise a series of issues that economists (whether theorists, experimentalists, or policy makers) should bear in mind regarding the application of experimental methods in economics, particularly when exploring behavioral aspects of decision making. Specifically, we focus our attention on the issues of validity and realism as applied to the use of experiments in research in behavioral economics.

EXPERIMENTAL METHODS IN ECONOMICS

While there does not appear to be a well-specified set of professional standards for conducting economics experiments, there is general agreement on the necessary components for a good ex-

periment (for example, see Davis and Holt 1993; Friedman and Sunder 1994; Roth 1988; Smith 1987). Violating these guides may result in experiments conducted in "dirty test tubes" (Binmore 1999) and results that inadequately test the hypotheses in question.

First and foremost, participants in an economics experiment must face adequate incentives. Given economists' focus on the application of cost-benefit analysis in decision making, the provision of adequate and salient incentives is a necessary condition for observing economic decision making in the laboratory. Second, most economists agree that the problem faced by participants in an experiment must not be too complex and must be framed in a manner simple enough for participants to understand. Third, if we are interested in decision making, experiments must allow participants to make good, effective decisions. Thus, deception is inappropriate (and potentially damaging) for economics experiments.[1] Finally, many experimental economists believe that time for trial and error must be allowed for participants to learn the workings of an experiment (i.e., how to "play the game"). Many of the experimental games used in economics are abstract or foreign to day-to-day decision making. Thus, repetition might be in order to allow participants to make trial-and-error adjustments. Given these guidelines, our interest is in how these aspects of an economics experiment influence the validity and realism of experimental results for research questions in behavioral economics.

In the discussion of validity and realism in experiments, it is useful to have an example to illustrate various concepts. Throughout this chapter we will make use of the ultimatum game as an example. In the ultimatum game, a proposer is allocated an endowment ω of which she must choose an amount $x \in [0,\omega]$ to offer a responder. The responder can then either accept or reject the offer. If the offer is accepted, the responder receives a payoff of x and the proposer receives a payoff of $\omega - x$. If the offer is rejected, each participant receives a payoff of zero.

As economics folklore has it, the ultimatum game was first proposed to Werner Güth (Güth, Schmittberger, and Schwarze 1982) by Reinhardt Selten as an example of a game in which there would be consistent deviations from the subgame perfect Nash equilibrium (Selten 1975). Given preferences over own wealth, subgame perfection implies that the responder will accept any nonnegative offer and, given this, the proposer will choose $x = 0$. On the other hand, ultimatum game experiments indicate that responders typically reject offers of less than 30 percent of the endowment, and proposers offer between 30 and 50 percent of the endowment. This game has been widely studied and experimental results are strikingly robust across incentive amounts, cultures, and elicitation methods (Henrich et al. 2001; Oxoby and McLeish 2003; Roth et al. 1991; Slonim and Roth 1998; see Camerer 2003 for a thorough review of this literature and results). As a result, the ultimatum game is often used in the motivation of theoretic models of fairness, reciprocity, and other forms of concern for others (e.g., Bolton and Ockenfels 2000; Charness and Rabin 2002; Fehr and Schmidt 1999).

EXPERIMENTAL VALIDITY

One of the primary advantages of experiments is the degree of control one obtains in identifying the causal relationships between dependent and independent variables. Ideally, one would like to conduct experiments in the field (i.e., natural or field experiments; see Harrison and List 2004) in which individuals make real decisions. However, experiments in the field are plagued by various forms of heterogeneity and "noise" that reduce one's ability to infer causal relationships. Economists, perhaps more than other social scientists, recognize the important trade-offs that exist between experimental control and outside realism.

In this section, we focus on these trade-offs by examining the types of validity experiments can provide for behavioral economists. Cook and Campbell (1979) identify three types of validity that may be used to interpret experimental results: internal validity, external validity, and construct validity.[2]

Internal Validity

Internal validity refers to the structure of an experiment itself and the degree with which one may infer causal relationships from the results. Internal validity asks the question, "To what extent are the independent (treatment) variables the sole source of the distribution of dependent variable?" The key in assessing internal validity is to examine the experiment to identify aspects of the decision environment, beyond the treatment variable, that could influence the experimental results. A good experiment makes use of the ability to observe behavior and decision making in a controlled environment, controlling the variation between experimental treatments to ensure that participants receive the same stimuli and experience the same conditions. As a result, the differences in observed behavior can be attributed to the differences participants encounter in the experimental treatments (i.e., the independent variables). The internal validity of an experiment is often questioned when there is noise in the experimental protocol or there are uncontrolled stimuli affecting participants' decisions in the experiment.

As an example, consider an experimental ultimatum game in which internal validity is compromised. A growing literature has examined the extent to which the threat of negative reciprocity in the ultimatum game (i.e., responders rejecting strictly positive offers) is subject to found-money effects.[3] Consider an ultimatum game experiment designed to identify the extent to which the distribution of offers is subject to the origin of the endowment used in bargaining. Thus the treatment variable is the source of the endowment used in bargaining. In the control treatment, participants play the ultimatum game following standard protocols (e.g., Güth, Schmittberger, and Schwarze 1982) in which the endowment is determined and provided by the experimenter. In the second treatment, the source of the endowment is altered. We will consider two potential sources for the endowment. In treatment T1, the endowment is provided not by the experimenter but rather by the proposers. That is, individuals assigned the role of proposer must provide an endowment from their own resources when they arrive at the experiment (cf. Clark 2002). In treatment T2, proposers must earn the endowment by engaging in some task.[4]

Consider the comparisons of experimental results between the control treatment and either treatment T1 or T2. Which of the treatment sessions (T1 or T2) provides stronger evidence of how robust behavior in the ultimatum game is to found-money effects? That is, is there greater internal validity in a comparison of the results from the control against results from treatment T1 or against results from treatment T2*?* Many would answer that there is greater internal validity in comparing the results from the control treatment with those from treatment T2, since between session T2 and the control session only one aspect of the decision environment has been altered (the mechanism used to allocate the endowments) and the experimenter can accurately observe how the endowment was determined. A similar difference exists between the control session and session T1, but the relationship between the source of the endowment and behavior is muddied, as the experimenter has no control or information regarding the determination or source of the money participants bring to the experiment—it could have been earned through the participants' employment, received as a gift, or unexpectedly found. Note that the latter two cases are examples of "found money" and precisely what the experimenter is trying to avoid in having participants provide their own endowments.

The key to obtaining internal validity is taking advantage of an experiment's ability to eliminate confounding factors that affect behavior and limit the differences between treatments to only one (or a selected number) of independent variables. In this way, the experimenter can neatly identify the effect of the independent variable(s) on decision making in the absence of confounds presented by other mitigating factors.

In addition to correctly choosing the independent variables in an experiment, a critical tool for achieving internal validity is random assignment. That is, if individuals are randomly assigned to each treatment in an experiment, then ex ante heterogeneity among the population of participants is controlled for insofar as there are no other factors (e.g., age, gender, level of education) that may directly differ between the treatments. Thus, given a properly designed experiment in which only the independent variables differ across treatments, random assignment solves the problem of internal validity. For example, there is ample evidence that individuals' personal and demographic characteristics have a strong influence on behavior. Eckel and Grossman (1998) find that women donate almost twice as much as men in anonymous dictator games. Similarly, Carter and Irons (1991) and Kahneman, Knetsch, and Thaler (1986) find that economics and business majors offer significantly less in ultimatum games.[5] Random assignment implies that the populations of participants in each treatment have similar distributions of personal characteristics (e.g., gender, education). Thus these (potentially unobservable) characteristics do not account for differences in the distribution of results between treatments.

In economics experiments, random assignment is often only partially implemented. In ideal circumstances, participants in an experiment would be assigned to different treatments and participate in the experiment at the same time. Thus, the population characteristics of the subject pool, differences in the communication of instructions, and temporal events that may affect participants in a similar manner (e.g., returning from a long weekend, lunch) are controlled.[6] While there may be no reason to think that these (seemingly minor) events could have an effect on behavior, neither is there any a priori reason to think that they will not affect behavior.[7]

In economics, we often observe comparisons between experimental results conducted at different times, with different participant pools, and administered by different experimenters. In such cases there is always the potential that the results may be attributable to events occurring in the period between the sessions, differences between the participant pools, or differences in the characteristics of the administrating experimenter (e.g., personality or demographic differences). To the reader seeking to inform theory or develop policy based on experimental results, one should always cautiously ask how much of the experiment's results may be attributable to such differences. For behavioral economists who are interested in using experiments to elucidate and explore psychological phenomena and processes in economic contexts, there is a rich literature demonstrating how these factors can influence participants' behavior in experiments (see Aronson, Wilson, and Brewer 1998).

Note that there may often be practical reasons for conducting the treatments of an experiment at different times or different places. For example, different treatments may require different continuances or facilities, which precludes conducting treatments at the same time. In such circumstances, the importance of random assignment in the initial phases of the experiment (i.e., recruiting) is heightened. This, along with the collection of demographic information to analyze the results for fixed effects, can help strengthen the internal validity of such experiments.

As a final note, psychologists typically regard within-subject designs as preferable to between-subject designs when it comes to maintaining internal validity. In within-subject designs, each subject participates in the experiment under each treatment. As such, each participant serves as her own control, thereby identifying individual-level differences that might otherwise be treated

as errors in the analysis of a between-subject design. In many economics experiments, however, particularly when money is used as an incentive and where income effects may engender different types of behaviors, within-subject designs may actually introduce greater confounds. In such environments there is an implicit trade-off between the internal validity obtained from within-subject design and the internal validity obtained by controlling for wealth effects or other economic phenomena affecting decision making.

External Validity

While issues of internal validity may challenge the causal relationship inferred from an experiment's results, *external validity* addresses the extent to which the causal relationship identified in the experimental setting can be generalized to other contexts, places, times, and people (e.g., Andersen et al. 2004). Questions of external validity often revolve around the context or participant pool used in the experiment. More subtly, external validity refers to the particular causal relationship gleaned from an experiment and the extent to which this relationship is robust in other environments. For example, experiments of ersatz labor markets conducted with university students may be subject to the criticism of the subject pool involved, the characteristics of which may or may not be representative of the population actively involved in the labor market (Sears 1986). As such, the results from the experiment may not translate into policy that can be implemented in real labor markets.

A particularly difficult challenge to the external validity of experiments in behavioral economics is that of context. Economists are very wary of establishing context in their experiments. In public goods games, instructions typically avoid use of the words "public good," and labor market experiments refer to the artificial employers and employees as "type A" and "type B" participants. However, most of the decisions people make are viewed by the decision maker as being within a given context and accompanied by a particular history that influences the understanding of events. For experimentalists, establishing a little bit of context can go a long way: there are strong differences between the way participants play the "community game" and the "Wall Street game" even when these two games are identical variations of the prisoner's dilemma game (Loewenstein 1999).[8] Given the influential work of Kahneman and Tversky on framing effects and reference dependency (Kahneman and Tversky 1979, 1988), it is clear that contextual issues play an important role in determining individual and group decision making.

In some sense, the problem of context is particularly difficult for behavioral economists. Many of the very insights they seek to incorporate into economics (ideas of fairness, emotions, reciprocity) are founded on the contextual aspects of a decision environment. For example, while experiments with the ultimatum game have led to advances regarding theories of fairness and reciprocity (e.g., Charness and Rabin 2002; Dufwenberg and Kirchsteiger 2004), the game itself is usually conducted in the absence of any context. Rather, participants are assigned roles and no "story" is given as to why the proposer/responder relationship develops or exists. As such, participants look for a decision-making strategy to employ in this environment. Although participants may also look for such strategies when a context is established by the experimenter, the absence of context concedes control over interpreted context to the participant, thereby reducing the experimenter's control in the laboratory. The tension created between self-interest and rules of thumb such as 50-50 may explain the observed results of offers ranging from 30 to 40 percent and rejection of offers below 30 percent.

While there is little doubt of the robustness of ultimatum game results (Camerer 2003), let us consider how important context is in this experiment. First, we may think behavior in this game

will be strongly influenced by norms (e.g., 50-50). As such, when playing this game without context, participants opt to implement a commonly understood norm of behavior. However, if the game is repeated, one may think of a context endogenously arising (at least in the minds of participants) and influencing behavior. For example, Binmore and colleagues (1993) find evolution toward the theoretic prediction in the distribution of offers and acceptance rates in a repeated ultimatum game.[9] This evolution may be evidence of the import of context: once participants have had an opportunity to experience the game, a new context may develop in which a new norm may come into being. The fact that we do not observe evolution to the subgame perfect Nash equilibrium should not come as a surprise: norms are strikingly robust, and when a norm is adhered to by a majority, transgressing it may be difficult.[10] Norms evolve slowly but systematically. The fact that we observe any evolution in the experimental environment developed by Binmore and colleagues (1993) should be taken as evidence that repetition can change the context of an experiment, thereby changing the rule of thumb or norm employed by participants.

A more important question of external validity arises when one considers the policy implications of an experiment. With the wider acceptance of incorporating behavioral insights into economics, economists conducting behavioral research are increasingly being asked questions that relate to economic policy (Camerer et al. 2003; Thaler and Sunstein 2003). While one might agree that results from experimental ultimatum games should inform economic theory, it is difficult to say precisely how these results should inform economic policy. Experimental results indicate that individuals take into account the payoffs of others in determining behavior, but how should such a finding influence policy regarding welfare programs, the provision of public goods (e.g., school choice initiatives or the funding of public schools), labor market regulation, or redistributive taxation? This is a trickier question, as individual decision making in the face of economic policy is rife with context. In bargaining environments, individuals are not proposers and responders but employers and employees, unions and firms, parents and children. The context created by these titles alone may significantly change the way in which others' payoffs are incorporated into one's utility function and how reciprocity or kindness are construed.

That said, it is worth asking how important external validity is to behavioral economics. In some sense, research in behavioral economics has been founded on the desire to develop a richer theory of decision making, one building on the neoclassical model but incorporating insights from research in psychology and sociology. Thus, many of the experiments in economics were devised to test existing theory and models rather than to make generalizations that might inform policy debates (e.g., Güth, Schmittberger, and Schwarze 1982). Indeed, much of the "first wave" of behavioral research in economics was characterized as anomalies against existing economic theory (see Thaler 1992). There is a definite benefit in theory testing, and experiments are an effective method toward this end.[11] Further, insofar as theory informs policy, so should experiments help in policy analysis and design. To borrow an analogy from Laver and Shepsle (1996), while experimental analysis and policy analysis are apples and oranges, they are both fruit. As such, one can certainly (although perhaps cautiously) inform policy analysis with the behavioral insights gained from experiments.

Construct Validity

As with external validity, *construct validity* challenges neither the internal consistency of an experiment nor the causal relationship between the dependent and independent variables inferred from the experiment's results. Rather, construct validity explores how these variables are measured in an individual's decision making and looks at the underlying relationship between these

variables. A natural way to think of construct validity is in terms of how the dependent and independent variables are factored into an individual's decision calculus.

As an example of the import of construct validity in the ultimatum game, there have been several papers developing theoretical models explaining the large offers and rejection of strictly positive offers observed in experiments. For example, Fehr and Schmidt (1999) develop a model based on inequity aversion (extended to include efficiency concerns and reciprocity by Charness and Rabin 2002). Bolton and Ockenfels (2000) develop a similar model based on relative payoffs. Rabin (1993) models a "kindness function" that yields cooperation with or punishment of others' acts, and Dufwenberg and Kirchsteiger (2004) extend this to describe reciprocity in extensive form games. Each of these models differs in important ways that implicitly point to different psychological underpinnings of how the variables in the ultimatum game influence decision making. This issue of construct validity in the ultimatum game focuses on which of these models is "correct" in the way in which it characterizes decision making in that environment.

In a similar spirit, Rubinstein (2001) addresses the issue of construct validity using anomalies in intertemporal choice, demonstrating that both quasi-hyperbolic discounting (Laibson 1996) and a procedural decision rule based on canceling similar events (e.g., Tversky 1977) describe the same anomalies. Again, construct validity asks which of these models most accurately captures the fundamental psychological process that is at work in intertemporal decision making.

The construct validity of an experiment can be challenged in several ways. The complexity of the decision environment may compromise the contextual validity of an experiment by muddying the relationship between the treatment (independent) variable and the theoretic variable or issue of interest. Similarly, the context (or lack thereof) of an experiment may distort the extent to which the treatment variable appropriately represents the theoretical process and variables employed in decision making. The key to fostering construct validity is the proper choice of an independent variable and sufficient treatment conditions to allow the experimenter to identify the behavioral insight and process actually at work.[12]

REALISM

Most experimentalists agree that many of the experiments they conduct lack what would be casually referred to as realism. Due to the conditions of an experiment and the desire to control for outside influences on behavior, experiments (save for natural experiments) often lack realism in that the circumstances individuals are encountering are unlikely to arise in the real world (they are often referred to as lacking *mundane realism;* see Aronson, Wilson, and Brewer 1998). From the perspective of behavioral economics, mundane realism may not be the most important aspect of an experiment. Rather, in the interest of bringing psychological insights into the realm of economic analysis, economists conducting experiments should be concerned with *experimental realism* and *psychological realism.*

Experimental realism is often defined as the degree to which the situations constructed in the experiment actively engage participants. On the other hand, psychological realism (as defined by Aronson, Wilson, and Akert 1994) refers to the degree to which the psychological processes occurring in an experiment are comparable with the psychological processes occurring in ordinary decision making.

With respect to experimental realism, economists have often been critical of experiments in psychology and hypothetical studies (e.g., hypothetical contingent valuation studies) in which individuals' behaviors and decision making are not motivated by adequate incentives or deception was employed. In the eyes of economists, the results obtained from experiments with insuf-

ficient incentives may be suspect, as individuals were not able to "put their money where their mouth is" and their decisions had no consequences. In game-theoretic jargon, the behavior observed in these experiments may be only cheap talk and an inadequate reflection of what individuals would do if real incentives or costs were involved. This is not to say that experiments with hypothetical consequences have no value or cannot inform theory and empirical economics; rather, we should not expect these experiments to engage participants in the same way as experiments with real consequences (Binmore 1999; Holt 1995). Similarly, if participants believe they may be deceived in an experiment, they have no reason to try to make an optimal choice. Given that participants may be wary of the decision environment in the experiment, deception may imply that they do not even know how to make an optimal choice in that environment.

With respect to incentives, Smith (1987) presents a conceptual framework with two sufficient conditions for a valid controlled experiment: saliency and nonsatiation. Formally, saliency requires that for a given outcome x, individuals' rewards are linked to the outcome via a function mapping outcomes onto rewards: $\pi = f(x)$. Nonsatiation requires that the utility function defined over rewards be strictly increasing (the utility function is an increasing monotone function): if $\pi > \pi'$ then $u(\pi) > u(\pi')$.

Given these conditions, experimental economists usually insist on the use of adequate incentives in experiments, and these incentives are usually in the form of monetary payments. As argued by Smith, economists should "use a monetary reward function to induce utility value on the abstract accounting outcomes of an experiment" (1987, 245). Thus, the offers and rejections observed in an ultimatum game played with real money are considered more "valid" and a truer reflection of individuals' preferences than those obtained from an ultimatum game played with hypothetical money.

In a large sense, this type of thinking is right. However, as Loewenstein (1999) states, "experimental economists should not deceive themselves into believing that the use of such rewards allows them to control the incentives operating in their experiments." This is particularly true for experiments in behavioral economics. Many times the phenomena we are interested in studying (e.g., other-regarding behavior or decision-making heuristics) are motivated by nonmonetary incentives associated with conformity or maintaining one's self-esteem. Further, many of the decisions we make in real life are not motivated by monetary payments.

There has been active research on the effect of monetary incentives on decision making in experiments. In tests of expected utility theory, Loomes and Beattie (1997) and Loomes (1998) find that providing incentives to participants changes little the extent to which behavior violates the axioms of expected utility. In his experiences conducting experiments, Rubinstein (2001) found little difference between experiments conducted with no money and results published using real money. Similarly, Henrich and colleagues (2001), Oxoby and McLeish (2003), Roth and colleagues (1991), and Slonim and Roth (1998) find striking robustness in ultimatum game results across cultures, sizes of incentive, and elicitation methods. On the other hand, Blumenschein and colleagues (1997), Forsyth and colleagues (1994), Kruse Brown and Thompson (2001), and McClintock and McNeel (1967) find significant effects of incentives in experimental games. Thus, results on the importance of monetary incentives are mixed. The presence of these mixed results is supported by the review of Smith and Walker (1993): in some experiments the size of financial incentives matters little, while in others financial incentives reduce the deviations from theoretic predictions. These findings are consistent with the view expressed by Camerer: "The effect of paying subjects is likely to depend on the task they perform" (1995, 635).

Thus, even with the use of financial rewards, there may be questions regarding the extent to which experimental realism holds in an experiment. It may be not the nature of the monetary incentives per se that influences the realism of an experiment, but the context in which those

incentives are provided. As a striking example, consider the research of Cherry, Frykblom, and Shogren (2002), Oxoby and Spraggon (forthcoming), and Ruffle (1996) on the influence of found-money effects. In these experiments, senders in dictator games allocated significantly more to themselves when they had "earned" the endowment and significantly more to receivers when they perceived the receiver as having "earned" the endowment. This should not be surprising: casual empiricism and research on found-money effects (Arkes et al. 1994; Thaler 1999) suggest that the source of an endowment of money plays a large role in how decisions are made over that money. These results indicate that one potential source of experimental realism is the legitimacy of assets in an experiment. As argued by Cherry, Frykblom, and Shogren, "just as rewards must be salient . . . the assets in a bargain must be legitimate to produce a rational result" (2002, 1220).

These results also point to a potential problem with experiments in behavioral economics regarding psychological realism. Taking the dictator game (or the ultimatum game, for that matter) as an example, there are very few circumstances in which a person may find herself in a real-world situation similar to the dictator game. Thus the game may lack external validity. While this may not be a major concern (see Mook 1983 and the preceding discussion), the fact that the endowments in a standard experiment are delivered by the experimenter may alter the way individuals think in the experiment. The results of Cherry, Frykblom, and Shogren (2002) provide a profound illustration of this: legitimizing assets on the part of dictators resulted in 95 percent support for the theoretic prediction. Thus, the standard dictator or ultimatum game may lack psychological realism in that the type of decision making participants display in the experiment may be very different from that employed in real-world situations.

The problems of psychological realism may be greater for behavioral economists given the standardized use of monetary incentives. There may be strong interactions between nonpecuniary motives and financial motives. Frey (1997) argues that the presence of monetary incentives may undermine or strengthen (depending on the decision-making environment) the intrinsic motivations of individuals. As a result, experiments that use financial rewards may be testing not the actual behavioral phenomena but rather how these phenomena are altered by monetary concerns. To the extent that these monetary concerns are absent in the context-dependent environments individuals encounter, the psychological processes individuals utilize may be different and yield different behaviors. Indeed, the interaction between monetary incentives and personal or social motivations is poorly understood.

One of the more interesting findings along this line of research is that of Gneezy and Rustinchini (2000) and Gneezy (2003), namely, that the effect of incentives is nonmonotonic and that small (inadequate) incentives may result in poorer performance than no incentives at all. As Gneezy (2003) argues, extrinsic motivation (i.e., monetary incentives) might change the way participants perceive an activity and (along the lines of Frey 1997) destroy the intrinsic motivations to act when there is no explicit reward from the activity. Related evidence shows how monetary incentives (more specifically, the structure of those incentives) influences the way individuals make decisions and perceive the behavior of others. Oxoby (2005) finds that the use of a decision-making heuristic (the proportion heuristic from Silvera, Josephs, and Giesler 2001) is heavily influenced by the type of incentive mechanism used to ostensibly motivate behavior. Similarly, Oxoby and Friedrich (2002) find that behavior in a trust game is strongly affected (and in a nonintuitive way) by whether the money used in bargaining was earned using joint or relative performance evaluations (i.e., team or tournament-style contracts).[13] Given that the psychological processes employed by experimental participants may be influenced by an experiment's constructs, caution should be used when interpreting these results as directly testing the psychological processes utilized in decision making taking place beyond the laboratory.

With respect to the use of deception in experiments, economists typically view deception as taboo (Hey 1998; McDaniel and Starmer 1998). First, deception dilutes the perceived incentives individuals face, thus compromising experimental realism. This can occur even with the hint of deception, thus making it important that deception *never* be employed lest it taint the pool of potential participants.[14] Second, and perhaps more important, we cannot expect individuals to make "normal" decisions when they believe they may be being deceived. Casually, we know that we make different types of decisions when we think we may have been misled; we should expect the same from participants in our experiments. If we are interested in studying decision making, the experiments we employ must give participants accurate (although maybe not all) the information necessary for engaging in good decision making. The presence of deception significantly changes the behavior of participants, confounding the inferred relation between the independent and dependent variables and compromising the psychological realism of the experiment.

CONCLUSION

Research in behavioral economics is founded on an interdisciplinary approach to understanding human behavior. As such, interested researchers should make use of all the available methodologies in their pursuits. The benefits of incorporating these methods have yielded a richer description of economic man and have provided researchers with greater insights into human decision making. In turn, these gains allow policy makers to design economic and social policies grounded in a more accurate theory of individual decision making.

For those interested in understanding psychological phenomena, experiments are an invaluable tool when brought together with the economic methodology used to understand behavior. However, the application of experimental methods in economics poses particular challenges, particularly for behavioral economists interested in incorporating psychological insights into the realm of economic analysis. For example, economists' focus on incentives and cost-benefit decision making dictates an experimental method that uses salient rewards to motivate decision making. However, behavioral phenomena such as altruism and heuristic-based decision making may be strongly influenced not only by the mere presence of incentives but also, more profoundly, by the context and inferred intentions these incentives create. Thus, designing experiments with strong (internal and external) validity and clear testable hypotheses becomes of paramount importance to experimenting economists.

For behavioral economists, there is ample evidence that the psychological phenomena at work in decision making are heavily influenced by the context and implicit incentives people face. As a result, behavioral economists face an additional challenge in the design of experiments: attention must be paid not only to internal and external validity but also to the construct validity and psychological realism of experiments and theories. It is with these guides that behavioral economics draws its power in informing neoclassical economics of the important details inherent in individual decision making.

As behavioral economics "goes mainstream," more attention will be paid to the policy implications and normative import of behavioral research. This implies that we must pay close attention to the methods employed in empirically testing these new and emerging theories.

NOTES

1. For a lively debate on the role, and lack thereof, of deception in experimental economics, see Bonetti 1998; Hey 1998; McDaniel and Starmer 1998.

2. Other researchers have defined other types of validity that should be accounted for in experiments and, more generally, behavioral research. For example, Sommer and Sommer (2002) define, in addition to those above, content validity, criterion validity, concurrent validity, and predictive validity.

3. See Thaler 1980; Arkes et al. 1994. Recent experiments in this area include Cherry 2001; Cherry, Frykblom, and Shogren 2002; Oxoby and Spraggon forthcoming; Ruffle 1996.

4. Previous experiments in which participants have had to earn the endowments include taking exams (Ruffle 1996) and cracking walnuts (Fahr and Irlenbusch 2000).

5. Similarly, Spraggon and Oxoby (2003) find that "sophisticated" participants (defined as those having taken an undergraduate course in game theory) are more likely to choose Nash-type behaviors in public goods games.

6. A friend recounted a story regarding a series of bargaining experiments (ultimatum and trust games). One treatment was conducted a week prior to the terrorist attacks of September 11, 2001; the control sessions were conducted several weeks after the attacks. Although the results from the first and second sessions differed, he was unsure as to how much of the difference may be attributable to the events of September 11 and the emotional impact they had on people.

7. As an example of the way in which the hunger experienced before lunch can influence individuals' projection of future preferences, see Read and van Leeuwen 1998.

8. Relatedly, Charness, Frechette, and Kagel (2004) find that the presence of a payoff table significantly affects the way in which individuals behave in a gift-giving game. The presence of such a table may not only facilitate participants' calculations of payoffs but also change the way they approach the interactions occurring during the experiment.

9. Similar results are obtained in the two-period ultimatum game of Binmore, Shaked, and Sutton (1985).

10. In the Quentin Tarantino film *Reservoir Dogs,* the opening scene depicts the difficulty one may have violating a simple tipping norm. Oxoby 2003 documents the evolution and development of social and cultural norms over the 1990s.

11. This point is eloquently argued by Mook (1983).

12. In the context of identifying the relationship between outcome- and intention-based reasons for other-regarding behavior, good examples of experiments with strong internal and construct validity include Cox 2004 and McCabe, Rigdon, and Smith 2003.

13. These results indicate that team-based incentives resulted in *less* observed trust and trustworthiness when those contributing less to the team's output were assigned the role of proposer. Under tournaments, losers assigned the role of proposer displayed significantly more trust than did winners assigned the role of proposer.

14. Hey (1998) argues against the use of deception in experiments. He eloquently discusses the difference between the use of deception and "partial information" in experiments.

REFERENCES

Andersen, Steffen, Glenn W. Harrison, Morten I. Lau, and E. Elisabeth Rustrom. 2004. "Preference Heterogeneity in Experiments: Comparing the Field and Lab." Working paper, Department of Economics, University of Central Florida.

Arkes, Hal R., Cynthia A. Joiner, Mark V. Pezzo, Karen Siegel-Jacobs, and Eric Stone. 1994. "The Psychology of Windfall Gains." *Organizational Behavior and Human Decision Processes* 59, 3: 311–47.

Aronson, Elliot, Timothy D. Wilson, and R.M. Akert. 1994. *Social Psychology: The Heart and the Mind.* New York: HarperCollins.

Aronson, Elliot, Timothy D. Wilson, and Marilynn B. Brewer. 1998. "Experimentation in Social Psychology." In Daniel T. Gilbert, Susan T. Fiske, and Gardner Lindzey, eds., *The Handbook of Social Psychology,* 99–142. New York: McGraw-Hill.

Binmore, Ken. 1999. "Why Experiment in Economics." *Economic Journal* 109: F16–F24.

Binmore, Ken, A. Shaked, and J. Sutton. 1985. "Testing Non-Cooperative Game Theory: A Preliminary Study." *American Economic Review* 75, 5: 1178–80.

Binmore, Ken, J. Swierzsbinski, S. Hsu, and C. Proulx. 1993. "Focal Points and Bargaining." *International Journal of Game Theory* 22, 4: 381–409.

Blumenschein, Karen, Magnus Johannesson, Glenn C. Blomquist, Bengh Liljas, and Richard M. O'Conor. 1997. "Hypothetical Versus Real Payments in Vickery Auctions." *Economics Letters* 56, 2: 177–80.

Bolton, Gary E., and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review* 90, 1: 166–93.

Bonetti, Shane. 1998. "Deception and Experimental Economics." *Journal of Economic Psychology* 19, 3: 377–95.

Camerer, Colin. 1995. "Individual Decision Making." In John H. Kagel and Alvin E. Roth, eds., *Handbook of Experimental Economics*. Princeton, NJ: Princeton University Press.

———. 2003. *Behavioral Game Theory.* Princeton, NJ: Princeton University Press.

Camerer, Colin, Samuel Issacharoff, George Loewenstein, Ted O'Donoghue, and Matthew Rabin. 2003. "Regulation for Conservatives: Behavioral Economics and the Case for Asymmetric Paternalism." *University of Pennsylvania Law Review* 151: 1211–54.

Carter, John, and Michael Irons. 1991. "Are Economists Different, and if So, Why?" *Journal of Economic Perspectives* 5, 2: 171–77.

Charness, Gary, Guillaume Frechette, and John Kagel. 2004. "How Robust Is Laboratory Gift Exchange?" *Experimental Economics* 7, 2: 189–203.

Charness, Gary, and Matthew Rabin. 2002. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics* 117: 817–69.

Cherry, Todd L. 2001. "Mental Accounting and Other-Regarding Behavior: Evidence from the Lab." *Journal of Economic Psychology* 22, 5: 605–15.

Cherry, Todd L., Peter Frykblom, and Jason Shogren. 2002. "Hardnose the Dictator." *American Economic Review* 92, 4: 1218–21.

Clark, Jeremy. 2002. "House Money Effects in Public Good Experiments." *Experimental Economics* 5, 3: 223–31.

Cook, Thomas D., and Donald T. Campbell. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings.* Chicago: Rand-McNally.

Cox, James. 2004. "How to Identify Trust and Reciprocity." *Games and Economic Behavior* 46, 2: 260–81.

Davis, Douglas D., and Charles A. Holt. 1993. *Experimental Economics.* Princeton, NJ: Princeton University Press.

Dufwenberg, Martin, and Georg Kirchsteiger. 2004. "A Theory of Sequential Reciprocity." *Games and Economic Behavior* 47, 2: 268–98.

Eckel, Catherine C., and Philip J. Grossman. 1998. "Are Women Less Selfish Than Men? Evidence from Dictator Experiments." *Economic Journal* 108: 726–35.

Fahr, René, and Bernd Irlenbusch. 2000. "Fairness as a Constraint on Trust in Reciprocity: Earned Property Rights in a Reciprocal Exchange Experiment." *Economics Letters,* 66, 3: 275–82.

Fehr, Ernst, and Klaus Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics* 114: 817–68.

Forsythe, Robert, Joel L. Horowitz, N.E. Savin, and Martin Sefton. 1994. "Fairness in Simple Bargaining Games." *Games and Economic Behavior* 6, 3: 347–69.

Frey, Bruno S. 1997. *Not Just for the Money: An Economic Theory of Personal Motivation.* Brookfield, VT: Edward Elgar.

Friedman, Daniel, and Shyam Sunder. 1994. *Experimental Methods: A Primer for Economists.* New York: Cambridge University Press.

Gneezy, Uri. 2003. "The W Effect of Incentives." Working paper, Graduate School of Business, University of Chicago.

Gneezy, Uri, and Aldo Rustichini. 2000. "Pay Enough or Don't Pay at All." *Quarterly Journal of Economics* 115: 791–810.

Güth, Werner, K. Schmittberger, and B. Schwarze. 1982. "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization* 3: 367–88.

Harrison, Glen, and John A. List. 2004. "Field Experiments." *Journal of Economic Literature* 42: 1009–55.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath. 2001. "In Search of Homo-Economicus: Behavioral Experiments in 15 Small-Scale Societies." *American Economic Review* 91, 2: 73–78.

Hey, John D. 1998. "Experimental Economics and Deception: A Comment." *Journal of Economic Psychology* 19, 3: 397–401.

Holt, Charles A. 1995. "Psychology and Economics." Discussion paper presented at the annual meeting of the Allied Social Science Association, San Francisco.

Kahneman, Daniel, Jack Knetsch, and Richard Thaler. 1986. "Fairness and the Assumptions of Economics." *Journal of Business* 59, 4: S286–S300.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica* 47, 2: 263–91.

———. 1988. "Rational Choice and the Framing of Decisions." *Journal of Business* 59, 4: S251–S278.

Kruse Brown, Jamie, and Mark A. Thompson. 2001. "A Comparison of Salient Rewards in Experiments: Money and Class Points." *Economics Letters* 74, 1: 113–17.

Laibson, David. 1996. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics* 112, 2: 443–77.

Laver, M., and K. Shepsle, eds. 1996. *Making and Breaking Governments: Cabinets and Legislatures in Parliamentary Democracie*s. New York: Cambridge University Press.

Lazear, Edward. 2000. "Economic Imperialism." *Quarterly Journal of Economics* 115, 1: 99–146.

Loewenstein, George. 1999. "Experimental Economics from the Vantage-Point of Behavioral Economics." *Economic Journal* 109: F25–F34.

Loomes, Graham. 1998. "Probabilities vs Money: A Test of Some Fundamental Assumptions About Rational Decision Making." *Economic Journal* 108: 477–89.

Loomes, Graham, and Jane Beattie. 1997. "The Impact of Incentives upon Risky Choice Experiments." *Journal of Risk and Uncertainty* 14: 149–62.

McCabe, Kevin, Mary Rigdon, and Vernon Smith. 2003. "Positive Reciprocity and Intentions in Trust Games." *Journal of Economic Behavior and Organization* 52, 2: 267–75.

McClintock, C.G., and S.P. McNeel. 1967. "Reward and Score Feedback as Determinants of Cooperative and Competitive Game Behavior." *Journal of Personality and Social Psychology* 4: 606–13.

McDaniel, Tanga, and Chris Starmer. 1998. "Experimental Economics and Deception: A Comment." *Journal of Economic Psychology* 19, 3: 403–9.

Mook, Douglas G. 1983. "In Defense of External Invalidity." *American Psychologist* 38, 4: 379–87.

Oxoby, Marc. 2003. *The 1990's.* Westport, CT: Greenwood Publishing Group.

Oxoby, Robert J. 2005. "How Much Does Size Matter? The Proportion Heuristic and the Structure of Incentives." Working paper, Department of Economics, University of Calgary.

Oxoby, Robert J., and Colette Friedrich. 2002. "Trust and the Structure of Incentives." Working paper, Department of Economics, University of Calgary.

Oxoby, Robert J., and Kendra N. McLeish. 2003. "Specific Decision and Strategy Vector Methods in Ultimatum Bargaining: Evidence on the Strength of Other-Regarding Behavior." *Economics Letters* 84, 3: 399–405.

Oxoby, Robert J., and John Spraggon. Forthcoming. "Yours, Mine, and Ours: The Effect of Ersatz Property Rights on Outcome Based Fairness and Reciprocity." Technical Paper 041012, Institute for Advanced Policy Research, University of Calgary.

Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review* 83, 5: 1281–302.

———. 1998. "Psychology and Economics." *Journal of Economic Literature* 36, 1: 11–46.

———. 2002. "A Perspective on Psychology and Economics." *European Economic Review* 46, 4–5: 657–85.

Read, Daniel, and Barbara van Leeuwen. 1998. "Predicting Hunger: The Effects of Appetite and Delay on Choice." *Organizational Behavior and Human Decision Processes* 76, 2: 189–205.

Roth, Alvin E. 1988. "Laboratory Experimentation in Economics: A Methodological Overview." *Economic Journal* 98: 974–1031.

Roth, Alvin E., Vensa Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir. 1991. "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburg, and Tokyo: An Experimental Investigation." *American Economic Review* 81, 5: 1068–95.

Rubinstein, Ariel. 2001. "A Theorist's View of Experiments." *European Economic Review* 45, 4–5: 615–28.

Ruffle, Bradley J. 1996. "More Is Better, but Fair Is Fair: Tipping in Dictator and Ultimatum Games." *Games and Economic Behavior* 23, 2: 247–65.

Sears, Donald O. 1986. "College Sophomores in the Laboratory: Influences of a Narrow Data Base on Social Psychology's View of Human Nature." *Journal of Personality and Social Psychology* 51, 3: 515–30.

Selten, Reinhardt. 1975. "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games." *International Journal of Game Theory* 4, 1: 25–55.

Silvera, David H., Robert A. Josephs, and R. Brian Giesler. 2001. "The Proportion Heuristic: Problem Set Size as a Basis for Performance Judgments." *Journal of Behavioral Decision Making* 14, 3: 207–21.

Slonim, Robert, and Alvin Roth. 1998. "Learning in High Stakes Ultimatum Games: An Experiment in the Slovak Republic." *Econometrica* 66: 569–96.

Smith, Vernon L. 1987. "Experimental Methods in Economics." In J. Eatwell et al., eds., *The New Palgrave Dictionary of Economic Theory and Doctrine.* London: Macmillan.

Smith, Vernon L., and James M. Walker. 1993. "Monetary Rewards and Decision Costs in Experimental Economics." *Economic Inquiry* 31, 2: 245–61.

Sommer, Robert, and Barbara Sommer. 2002. *A Practical Guide to Behavioral Research,* 5th ed. New York: Oxford University Press.

Spraggon, John, and Robert J. Oxoby. 2003. "Can We Train Students to Be Nash Payoff Maximizers?" Working paper, Department of Economics, Lakehead University. Mimeo.

Thaler, Richard H. 1980. "Towards a Positive Theory of Consumer Choice." *Journal of Economic Behavior and Organization* 1, 1: 39–60.

———. 1992. *The Winner's Curse: Paradoxes and Anomalies of Economic Life.* Princeton, NJ: Princeton University Press.

———. 1999. "Mental Accounting Matters." *Journal of Behavioral Decision Making* 12, 3: 183–206.

Thaler, Richard H., and Cass R. Sunstein. 2003. "Libertarian Paternalism." *American Economic Review* 93, 5: 175–79.

Tversky, Amos. 1977. "Features of Similarity." *Psychological Review* 84, 4: 327–52.