Journal of Economic Behavior & Organization 73 (2010) 122–131

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

The science of experimental economics☆

Rachel Croson[a,*], Simon Gächter[b]

[a] University of Texas at Dallas, School of Economic, Political and Policy Sciences, 800 West Campbell Road, Richardson, TX 75080-3021, USA
[b] University of Nottingham, Centre for Decision Research and Experimental Economics, School of Economics, Sir Clive Granger Building, University Park, Nottingham NG7 2RD, United Kingdom

ARTICLE INFO

Article history:
Received 20 January 2009
Received in revised form 6 September 2009
Accepted 10 September 2009
Available online 3 October 2009

JEL classification:
B4
C9

Keywords:
Methodology
Experiments
Philosophy of science

ABSTRACT

In this paper we present the views of two practicing experimental economists on the role of economic experiments in the science of economics, and in particular on the interaction between economic theory and experimental design and data.

Games people play, you take it or you leave it
Things that they say, just don't make it right
If I'm telling you the truth right now, do you believe it
Games people play in the middle of the night

Where do we go from here?.. . .
Alan Parson's Project: Games People Play

1. Introduction

Some years ago, it was still common to argue that economics could never be an experimental science but was confined to be purely observational or theoretical. Today the integration of experimental economics into mainstream economics is an established fact. Economics *is* an experimental science, as well as a theoretical and observational one.

In this essay we discuss our views on what experiments can contribute to economics. Our focus is mainly on the relationship between theory and experiments. Experiments are historically closely linked to economic theory and recently empirical economists who typically use non-experimental observational data have begun to use experimental methods. Our eclectic views in this essay are those of "practitioners" of experimental economics, not those of professional methodologists. Like most economists we are normally busy doing research, not philosophy of science.

☆ The authors thank Barkley Rosser, Catherine Eckel, and participants at the 2008 Economic Science Association conference for helpful guidance and feedback. All errors (and opinions) remain our own.

* Corresponding author.
E-mail addresses: crosonr@utdallas.edu (R. Croson), simon.gaechter@nottingham.ac.uk (S. Gächter).

0167-2681/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jebo.2009.09.008

A large body of experiments has established "anomalies" in individual decision-making (summarized in a wonderful popular book by Thaler, 1991) and the body of evidence from experiments on bargaining, public goods, coordination, markets, auctions and individual decision-making has grown (Kagel and Roth, 1995, 2009; Plott and Smith, 2008). Standards of how to conduct economic experiments have emerged, textbooks teach methods of experimental economics (Davis and Holt, 1993; Friedman and Sunder, 1994; Friedman and Cassar, 2004), and a dedicated field journal, *Experimental Economics*, has an increasing impact in economics. Experimental economics has come of age.

This essay reflects our views on how we see experimental economics now. We want to acknowledge the pioneers who have shaped much of the methodological convictions of most experimental economists (e.g., Smith, 1982; Plott, 1982; Roth, 1995). A second wave of methodological thinking is taking shape in the form of specialized books tackling methodological issues (Guala, 2005; Caplin and Schotter, 2008; Bardsley et al., 2009), special issues (Methodology, e.g., Sugden, 2005a), and methodological articles in major journals (e.g., Samuelson, 2005; Levitt and List, 2007; Schmidt, 2008). Our goal is to contribute to this thinking about the field.

We begin with our perspectives on economic theory and experiments (Sections 2 and 3). We then move to a discussion of what theories can do for experiments, and what experiments can do for theories (Sections 4 and 5). Section 6 discusses how we should evaluate theories in light of experimental data. We summarize with our "Top Ten" list of things to do (and avoid) for experimentalists and theorists alike in Section 7. Section 8 concludes.

2. What is (economic) theory?

Economic theory provides a framework and tools for describing and analyzing economic situations. It makes behavioral assumptions to derive predictions and to provide explanations of economic and social phenomena. More specifically, an economic theory (model) is a description of a social situation, which involves specifying the actors, the choices they face, their information, and how they evaluate each possible outcome.

Behavioral assumptions enter in three ways: a first set of assumptions specifies how the individuals evaluate each possible outcome. The evaluations of outcomes are described by the preferences of the individual, including attitudes towards risk and uncertainty ("risk preferences"), the future ("time preferences"), and the extent to which outcomes and the behavior of others are relevant ("social preferences"). A second set of assumptions explicates the individuals' cognitive abilities (their degree of logical sophistication) and how they form beliefs about uncertain states of the world. A third set of assumptions specifies how the individuals will behave. These behavioral predictions are derived by applying "solution concepts", which describe how assumptions on preferences and beliefs translate into outcomes.

This description of economic theory is a very abstract framework, which is instantiated with more specific assumptions to derive predictions. We describe "standard economic theory" as theory which assumes:

(i) Economic decision makers are cognitively sophisticated—they are assumed to be rational; their preferences are complete, transitive and obey the other axioms of expected utility theory. Individuals also have rational expectations about relevant states of the world and the behavior of others. In strategic situations rationality is common knowledge, that is, everybody is rational, knows that everybody is rational, knows that others know that everybody is rational and so on.
(ii) Risk preferences obey the expected utility principle. Time preferences exhibit exponential discounting. Social preferences are assumed to be zero (or already captured in the utility function). In applications, functional forms for the utility function are specified, e.g., it is assumed that people have Cobb-Douglas utility functions or that they have constant relative or absolute risk aversion.
(iii) Solution concepts are typically equilibrium concepts, like competitive equilibrium, Nash equilibrium, or subgame perfect Nash equilibrium.

It is important to note that these assumptions are psychological (behavioral) because they specify how individuals evaluate outcomes, reason, and reach decisions. This standard model is an extremely powerful framework which provides intellectual clarity and tractable formalizations. This framework has generated considerable theoretical advances in all areas of economics (see, e.g., the textbooks by Mas-Colell et al., 1995 and Romer, 2005, and the essays by Becker, 1993 and Lazear, 2000). This framework has made economics the most influential of the social sciences.

More generally, a theory (model) is an abstraction from the world. Models are, by definition, wrong (descriptively inaccurate). However, this does not mean they are useless. In particular, theories make the logical implications of behavioral assumptions explicit and concrete. The standard model can be seen as a boundary case, but understanding boundary cases is critical for scientific development. This does not relieve theorists from developing models based on more descriptively accurate assumptions, but it provides an important benchmark against which these theorists can compare the implications of their new models.

In summary, economic theory provides a framework for analysis. The conceptual distinction between preferences, beliefs, and constraints is an extremely useful framework that has served economics well. The behavioral assumptions of the standard model are boundary cases, but the standard model has been invaluable in providing testable predictions and developing theoretical alternatives.

3. What are (economic) experiments?

Experiments are a controlled data generating process. "Control" means that most factors which influence behavior are held constant and only one factor of interest (the "treatment") is varied at a time. This controlled variation of factors is crucial for being able to draw causal inferences. In rare cases this controlled data generating process occurs naturally (called a "natural experiment"). In the majority of cases, however, it is the researcher who controls the data generating process. Notice that our definition of an experiment is very general—it applies to all experimental sciences, and hence also to economics. Also, neither in our definition, nor in our discussion of experiments do we make a distinction between lab or field experiment or between a classroom experiment, one conducted in an fMRI scanner, or an experiment conducted with students, professionals, or Joe The Plumber. Finally, what "control" means depends inextricably on the exact research question. The question defines which arguments need to be held constant (and which varied systematically).

Economic experiments are experiments motivated by (and designed to answer) economic questions. Economic experiments usually involve controlling the choice sets (what decision makers can do), the information conditions (what decision makers know), and the monetary incentive structure (how decisions translate into payoffs) (Smith, 1976, 1982). Thus, an experiment implements the most important aspects of an economic model as introduced above.

Like theoretical models, experiments are simplifications of the world. They are equally (or sometimes more) descriptively inaccurate than economic models. They often (but not always) involve very little context, artificial settings, and abstract instructions. That said, just as descriptively inaccurate theories are useful, descriptively inaccurate experiments are useful as well. They provide tests of theories, examples or illustrations of phenomena, and techniques to elicit and measure preferences in various populations. They can provide "existence proofs" or "nonexistence proofs" of theoretical constructs.

In summary, experiments are controlled data collection processes. Like theories, experiments can be descriptively inaccurate, and yet still critical for scientific progress.

4. What do theories do (re: experiments)?

The role of economic theory for experiments is most straightforward when experiments are used to test a theory. Theory tells the researcher the relevant variables which need to be controlled or manipulated, and makes predictions of what will happen in the experiment. A famous example is tests of competitive market theory, initiated in their modern form by Smith (1962). Market theories are designed to apply to *any* market and *any* supply and demand function. Smith "induced" particular supply and demand functions and used competitive price theory to derive the equilibrium price and quantity. Together with assumptions on trading rules (which typically are not part of the theory) and some further auxiliary assumptions (like privacy of payoffs) the experiments test whether actual prices and quantities correspond to those predicted by theory. Economic theory plays a particularly important role when the task is to discriminate between competing theories. The theories allow the researcher to identify the parameters (or experimental treatments) in which competing theories make different predictions.

Even when experiments are not explicitly designed to test theory, theory typically provides a framework to derive benchmark predictions (for instance, what does expected utility or prospect theory predict? What will the outcome be if everyone is rational and self-interested, or boundedly rational and inequity averse? How would a time-(in-)consistent person behave?). In this role, economic theory is useful even if we know from other data that the theory is often descriptively wrong, as it provides predictions which can be tested.

But perhaps the most critical role of economic theory is to formulate new models consistent with robustly observed behavioral regularities. Experiments can only provide facts; explanations have to come from theory. Prominent examples are prospect theory to explain violations of expected utility theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992; Schmidt et al., 2008), the theory of inequity aversion to explain other-regarding preferences (Bolton, 1991; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) and hyperbolic discounting to explain present-bias and excessive exponential discount rates (e.g., Loewenstein and Prelec, 1992; Laibson, 1997). These theoretical advances are useful for (at least) three reasons. First, as with any theoretical advance, the new theory typically organizes a body of empirical evidence better than the previous theory, often including the previous theory as a special case. Second, these new theories provide new testable predictions, which moves scientists to the next step of the dialectic between theory and data. Finally, as with the standard theory, new theories illuminate particular behavioral phenomena. Thus if and when the new theories are disproved, they provide a language to talk about the results which they helped to organize.

Consider the theory of inequity aversion (Bolton, 1991; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). Prior to these theories, experiments and observational data had produced a body of contradictory-looking evidence with ad hoc explanations: people behaved non-selfishly sometimes (e.g., in ultimatum games and public goods games with punishment), but (with experience) tended to conform to the self-interested prediction (e.g., in public goods games without punishment and in market experiments even with unfair outcomes). Theories of inequity aversion rationalized both sets of findings.

These new theories generated new predictions, which were themselves tested (e.g., Blount, 1995; Engelmann and Strobel, 2004; Falk et al., 2008) and inconsistencies discovered. Subsequently new theories of reciprocity were developed to account for these new observations (e.g., Charness and Rabin, 2002; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006; Cox et al., 2007). Just as the existence of inequity aversion theories did not violate the usefulness of standard theories in some settings, the development of reciprocity theories did not make inequity theories useless. Theories are useful not only for their predictive power, but also for the insights they produce.

In summary, what do theories do with regard to experiments? Economic theory provides an important input for the design of most economic experiments, not only theory-testing experiments. They provide a guide for choosing parameters, for identifying potential confounds and competing explanations. Economic theory also plays an important role for explaining established behavioral facts in novel and testable frameworks.

5. What do experiments do (re: theory)?

Experiments make contributions to scientific research in multiple fields. Experiments are a critical part of scientific discourse in the natural sciences (e.g., physics). Experiments have historically been a pivotal methodology in other social sciences (e.g., psychology). However, experiments have found their way to economics only recently. It is thus not surprising that we are still 'feeling our way' around the intersections of experiment, theory and observational data.

One mental model of how experiments might interact with these other types of economic research is as follows:

Theory ------------------------- Experiment (Lab/Field) ------------------------- Observational data

In this model, experiments (both lab and field) have the potential to serve multiple purposes. First, they enable us to address theories. Second, they allow us to examine regularities from the field (observational data) in a controlled, abstracted setting. This exercise can highlight which observed regularities (or anomalies) are particular to a specific field context, and which are generalizable and should be observed in other settings. While this second purpose of experiments is not the focus of this paper, it nonetheless represents an important contribution.

How do experiments help us to address theories? Below we discuss different ways experiments can test model predictions, provide behavioral models, refine theories, suggest new theories, and serve as measurement tools. This list is by no means complete, but we hope it will provide a starting-ground for thinking of the relationship between theory and experiments.

5.1. Experiments test predictions

Theories (models) are, by definition, simplifications of the world. The goal of a theory is to identify and isolate a phenomenon in order to understand its impacts. Ideally, theories yield unique and testable predictions. Economic theories are logical systems whose truth derives logically from the assumptions. Experiments test whether observed behavior corresponds to the predictions of a particular model. However, testing the predictions of theories is often difficult (and sometimes impossible) to do with observational data.

This difficulty stems from a number of sources. First, tests of model predictions using observational data are typically *joint hypotheses tests*. We need to test jointly whether the assumptions of the theory hold in the field, and whether the predictions of the theory hold in the field. Moreover, any test of a theory is also a joint hypothesis test of auxiliary assumptions that need to be made to implement a test (the "Duhem-Quine problem"). Experimental procedures in the lab can reduce the jointness of the test.

For example, imagine a test of auction theory. The theory's assumptions involve an independent draw of individual values from a known distribution. The theory then predicts bids which will be placed. Observations from naturally occurring auctions may or may not conform to the assumptions of the theories (in the field, values may [or may not] be correlated). Thus if the theory's predictions are not observed, we do not know whether this is due to a violation of the theory's assumptions (e.g., correlated values) or to the fact that individuals are not reasoning in the way the theory predicts.

Of course, experiments cannot eliminate the joint hypothesis testing problem entirely. For example, economists typically use induced valuation in our experiments, but one can never be sure that the individual's utility function is defined only over one's own payoff, as assumed by many economic theories. Some experiments use a binary lottery procedure, but we cannot ensure that participants are risk-neutral. However, the controlled laboratory situation implements the assumptions of the theory as closely as is possible (e.g., the number of players, the possible moves, the order of moves, the information sets, etc.).

A second and related way which controlled experiments can test theories is also related to this additional control. Observational data will, of necessity, involve omitted variables which may be important to the theory. As a result, tests of theory from observational data typically rely on testing the *comparative statics* of a theory. As price goes up, quantity demanded goes down. Observational data can rarely test point-predictions of a theory. Experimental data, with its additional control, can.

Experiments are useful for testing theories because they enable us to compare the predictions of economic models against outcomes of the experiment. If the experimental results confirm the theory we can alter the parameters of the experiment to identify the robustness of the model ("stress-testing"), or use the experimental results to estimate parameters of the model. If the predictions of the model are not observed we can search for reasons why. Hundreds of experiments have been conducted under this methodology.

A third way that experiments can test theories involves running *horse races* between competing theories. Often multiple theories make the same prediction about observational data (typically the theories were written with the observational data in mind). Conditions under which these theories would make different predictions may not naturally occur. In the context of the laboratory, however, we can design situations not found in nature in order to distinguish between competing theories.

For example, two types of theories have been developed to explain observations in the world (and in experiments) of other-regarding preferences. Outcome-based theories (e.g., Bolton, 1991; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) assume that individuals care about the inequity or inequality of outcomes between themselves and others. Intention-based theories (e.g., Rabin, 1993; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006; Segal and Sobel, 2007) assume that individuals care about the intentions of others, but not (only) about the outcomes *per se*. All these models are consistent with the observations we have from the field.

Experiments can distinguish among them. For example, in Blount (1995), individuals responded to ultimatum offers generated from another person, or from a randomization device. If individuals were motivated by outcome preferences, their attitudes towards a given offer (fair or unfair) should be independent of how that offer was generated. Instead, individuals were significantly more willing to accept low offers when they were generated from the device than from the person. This experiment nicely distinguishes between these two classes of theories.

A final way experiments can address theories is to limit the domains where the theories might or might not apply. Experiments can thus define the scope of a given theory. For example, expected utility theory suggests that individuals should not demonstrate preference reversals. Experiments have observed preference reversals in hypothetical and incentivized choices (e.g., Grether and Plott, 1979), but less often in real world settings with high incentives (e.g., Bohm and Lind, 1993; Bohm, 1994). Thus expected utility theory might not hold in hypothetical or low-stakes choices, but might be a good predictor of behavior in high-stakes situations. The world, of course, involves both types of situations (e.g., buying milk versus buying a house). Experiments can tell us when expected utility theory is likely to make accurate predictions and when it is not.

Experiments are also useful for testing institutions, like trading rules, matching mechanisms, or auction designs. The laboratory serves as a test bed, or "wind tunnel", to check the robustness of different institutions before they are implemented (for some examples see, e.g., Plott and Porter, 1996; Abbink et al., 2005; Ockenfels and Roth, 2006). In this line of research, experiments are used as tools for "economic engineering" (Roth, 2002).

We close this sub-section by reiterating a point made by Plott (1982) in discussing the role of the experimental lab as a midpoint between the theory and the field. The (well-designed) laboratory experiment gives a theory its 'best-shot' at making accurate predictions. The assumptions of the theory are designed into the lab experiment. For example, if an auction theory assumes that signals are independently drawn from a known and stationary distribution, the lab experiment addressing that theory will involve signals independently drawn from a known and stationary distribution.

This 'best-shot' logic also applies to the choice of experimental subject pool. For example, we might believe that expected utility theory is a good descriptor of the preferences of individuals. Then using smart, mathematically sophisticated participants like college students gives expected utility theory its best chance of making accurate predictions. If one conjectures that practitioners will behave differently, running experiments with them to can test these conjectures (e.g., Cooper et al., 1999).

Thus in many ways, experiments are an *existence proof*; for some set of individuals, for some sets of institutions, with some set of parameters, the theory's predictions are observed. This, of course, does not guarantee that the theory's predictions will *always* be observed (a uniqueness proof). But if under these best-shot conditions the theory's predictions are not observed, this is a strong statement indeed.

5.2. Experiments provide behavioral models

Economic models derive the logical implications of its assumptions for behavior. In contrast, behavioral models describe observed behavior in a particular framework. The usefulness of all models, theoretical and behavioral, rests in part on the predictions they make for related or appropriate situations, as well as the insights they illuminate.

Many of the most influential experiments provide abstract behavioral models about an underlying phenomenon. For example, consider the "Stroop task", a famous psychological experiment in which participants are asked to name the color of the font in which a word is printed (e.g., the word green printed in red; Stroop, 1935). Under time pressure many people say the word and not the color. The automated response (reading) fails to be overridden by the cognitive effort (identifying fonts color). While the Stroop task does not capture tasks one typically observes in daily life, many situations in daily life involve of the interaction of automated responses and cognitive effort. The Stroop task, though highly abstract and unrealistic if taken literally, is a behavioral model (paradigm) for this tension.

From economics, consider a public goods experiment with punishment. In the typical setup (following Fehr and Gächter, 2000) individuals make contributions to a linear public good and are then informed about each others' contribution. In the second stage individuals can punish others at a cost to themselves and to the punished group members. This experimental design is clearly artificial. But it provides a model of how informal sanctions (peer punishment) work (see Gächter and Herrmann (2009), for an overview). Thus, the public goods experiment with punishment is a behavioral model of how social sanctions can sustain cooperation.

Many of the most influential experiments of recent years can be seen as behavioral models of the situations they represent. These experiments serve as "exhibits" (Sugden, 2005b), or "material models" (Schmidt working paper). They are "existence proofs" that illustrate a behavioral phenomenon. These phenomena can then inspire new theoretical arguments. This dialectic is exactly what drives the scientific method and ultimately our knowledge of the world.

5.3. Experiments refine theories

A second role of experiments is to refine theories. For example, assume (for purposes of argument) that individuals have an exponential discount rate. One might want to estimate the population average discount rate. Observational data has the potential to do this. For example, Hausman (1979) used data from refrigerator purchases to estimate discount rates. However, this data provides very little information on the distribution of discount rates across the population.

Furthermore, these estimates seemed inconsistent with the other choices we presume individuals were making (e.g., saving money, going to college, etc.). Experiments can help to reconcile these inconsistencies. Participants in the lab can make multiple intertemporal choices, and these choices can be examined for consistency.

This leads to the natural question: what do we do when choices are inconsistent? As scientists we need to evaluate a number of alternatives, from adding errors to our models (if the errors are "not too large"), limiting the scope of the theory to those domains where parameters are consistent, adding an additional parameter to capture the inconsistency (e.g., perhaps individuals have different discount rates for financial and health domains), or revising the theory even more drastically (e.g., perhaps inconsistent time preferences can be reconciled with a hyperbolic discount function rather than an exponential one). Analysis of observational data typically examines one decision per individual; thus potential inconsistencies may not present themselves.

Furthermore, we have discussed how experiments with different subject pools can provide information on how parameters might vary with demographics. The identification of individual heterogeneity and the ability to predict an individual's choice without having observed a previous decision (simply from their demographics) is something that experiments can add in this setting.

In psychology experiments are routinely used to refine theory. For example, the original version of prospect theory included a probability-weighting function which was undefined at 0 and 1. A series of experiments investigated what happened very close to and at 0 and 1. The refined theory (cumulative prospect theory, Tversky and Kahneman, 1992) incorporated these results into its probability-weighting function. Similarly, further experiments have led to further developments of prospect theory (Schmidt et al., 2008).

5.4. Experiments suggest/construct new theories

A third, and major, way in which experiments address theories is to suggest and construct new theories. This might involve identifying parameters which were not previously included in a theory but can be shown to matter, like scope limitations discussed above.

Alternately, experiments can suggest entirely new theories. For example, a large body of experiments demonstrated that individuals do not maximize their own earnings in games, but instead react to 'fairness.' Theories pin down the specifics; e.g., what does 'fairness' mean? Experiments can then identify whether (and when) the new theory's predictions are observed, how robust it is to context, whether we observe individual heterogeneity (and on what basis).

Similarly, experiments can identify behaviors which are correlated within the individual, and suggest a theory which captures the underlying similarity between the behaviors. For example, imagine that the same people who contribute in social dilemmas also accept more risk. This suggests that some contribution behavior may be driven by a willingness to take risks, especially in the behavior of others.

5.5. Experiments serve as measurement tools

A final role of experiments is their usefulness in measuring individual's preferences (risk, time social, etc.), their degree of strategic sophistication and other behavioral assumptions. For example, consider measuring risk preferences. The typical procedure is to confront participants with induced risky prospects and to ask them to choose either between different risky prospects (or sometimes sure payments). The decisions are then used to infer risk preferences (e.g., Holt and Laury, 2002; Dohmen et al., 2005; Andersen et al., 2008; Eckel and Grossman, 2008). Of course, risk preferences as measured in the lab may or may not correspond with risk preferences as exhibited in the field in a particular setting. Indeed, research has demonstrated that risk preferences in the financial domain are significantly different than risk preferences over the health domain (Weber et al., 2002); thus lab-generated preferences between gambles may or may not predict smoking behavior. The extent to which the measurements from experiments can be used to predict behaviors in naturally occurring decision situations is an empirical question. Research on this issue is underway and has produced some encouraging results (e.g., Dohmen et al., 2005; Benz and Meier, 2008; Chabris et al., 2008; de Oliveira et al., 2008).

When using experiments as measurement tools the context and the subject pool are of generic interest. For example, experiments as measurement tools have been used to ask how measured preferences vary across genders (Croson and Gneezy, 2009), across interesting social groups (e.g., Carpenter et al., 2004; Tanaka et al., forthcoming; Burks et al., 2007; Chesney et al., 2007; Gächter and Herrmann, 2007; de Oliveira et al., 2008), and across cultures (e.g., Henrich et al., 2001; Oosterbeek et al., 2004; Buchan et al., 2006; Bohnet et al., 2008; Herrmann et al., 2008). A particularly ambitious use of experiments as measurement tools is investigating how behavioral regularities vary with demographics in representatively selected participant pools (e.g., Fehr et al., 2002; Dohmen et al., 2005; Huck and Müller, 2007; Bellemare and Kröger, 2007; Bellemare et al., 2008; Andersen et al., 2008).

In summary, experiments provide a unique opportunity to address theories, above and beyond what observational research can achieve. Experiments can be used to test theories, to refine theories, or to suggest new theories. Like any methodology, experiments have limitations as well. Nonetheless, they can be a critical component of our economic toolbox.

6. How should we evaluate theories?

How do we judge a theory is of value? Sometimes we value theories simply because of their elegance, or because they illuminate and capture a useful insight. However, for an economic theory to have practical value, it needs to make useful predictions about economic behavior.

Under what conditions do theories achieve this lofty goal? First, theories may offer an unanticipated prediction, different from the predictions of other, competing theories. If they do not, they fail to provide any marginal predictive value. Alternately, theories may offer the same prediction as previous theories but for a different reason. In this case, empirical researchers (who use experiments and observational data) need to think about how these theories might be differentiated.

Second, theories need to be falsifiable (Popper, 2002). That is, they must make a concrete prediction in a given setting which will enable us to test them. One criticism of expected utility theory is that it can predict (almost) anything. Any single observed behavior can be consistent with the theory; after all, we simply define a utility function (and constraints) for which the observed decision is the outcome of its maximization. Even if we observe repeated decisions from one individual, very few sets of observations will falsify expected utility theory (violations of transitivity, preference reversals, or violations of the independence axiom are in this small set). Experimental designs have been developed that explicitly test for these violations (Starmer, 2000).

Third, and related, theories need to make predictions in a large set of circumstances. Here we refer to this attribute of theories as the *scope* of a theory. More parsimonious theories with fewer parameters have larger scope. The principle of Occam's Razor is useful here; if two theories give the same prediction, we prefer the simpler one, with the fewest parameters or 'moving parts.' The reason is that when theories involve many parameters, an observer needs to know the values of each of those parameters in order to use the theory to make a prediction in a given setting. As the number of parameters increases, the set of situations where those parameters are known decreases.

For example, if we assume individuals are risk-neutral, we can predict their decisions in any risky decision simply by knowing the probabilities and the payoffs. If we assume, instead, that individuals have CARA utility functions, we need to know the CARA parameter in order to make a prediction about their behavior. If we assume, instead that individuals have prospect theory preferences, we need to know their probability-weighting function, the curvature of their utility function for gains and losses, and their loss-aversion parameter. Thus theories with large scopes are parsimonious; they have relatively few parameters.

However, there is a tradeoff between scope and accuracy. The more parameters (degrees of freedom) a theory has, the more likely that its predictions will be accurate in the situations where it is able to make a prediction (when the values of the parameters are known). This tradeoff is depicted in Fig. 1 below.

The left side of the graph depicts theories which are high in scope (very parsimonious, few parameters) but not very accurate. One might imagine a standard (self-interested, risk-neutral, expected utility) theory here. The theory has no parameters; the situation (possible outcomes, probabilities and earnings) exactly predict the decision individuals will make.

As we move away from this simplified version of *homo economicus*, theories become more complicated, but also more accurate. For example, adding a risk parameter to the utility function decreases the scope; now one can only predict economic decisions when we know the risk parameter, but increases the accuracy of the theory's predictions. Similarly, moving to outcome-based other-regarding utility functions (altruism as in Becker, 1974, warm-glow as in Andreoni, 1990, inequality or inequity aversion as in Bolton, 1991, Fehr and Schmidt, 1999, and Bolton and Ockenfels, 2000) adds more parameters but further increases the accuracy of the predictions of the theory, at least in some circumstances.

Moving to intention-based utility functions (e.g., Rabin, 1993; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006; Segal and Sobel, 2007) increases the accuracy even further, gut further decreases the slope. Prospect theory (Kahneman and Tversky, 1979) adds further parameters and again yields an increase in predictive power.

At the other end of the spectrum we have placed psychology. Rather than a single unifying theory, psychologists create theories designed to explain outcomes in a relatively small set of situations. These theories are highly accurate in predicting behavior in a somewhat narrow setting.

Another way theories can be evaluated is on the basis of their self-defined question. What is the theory trying to accomplish, and how well does it accomplish this? For example, theories of inequality-aversion were designed to explain results from experimental games. How well does it accomplish this goal? Is there a different theory, with fewer parameters, which could do as well on existing data? Good theories can be nested; they can include previous theories as special cases, which can allow the researcher to test the significance of given parameters. Theories can be evaluated on the basis of their predictive ability as well. Does inequality-aversion theory allow you to make (accurate) predictions about experiments not-yet run?

Finally, we reiterate a point we made earlier: theories may be valuable for their insight and illumination, even if they explain or predict nothing about actual behavior (e.g., Rubinstein, 1991). For example, theories about purely self-interested

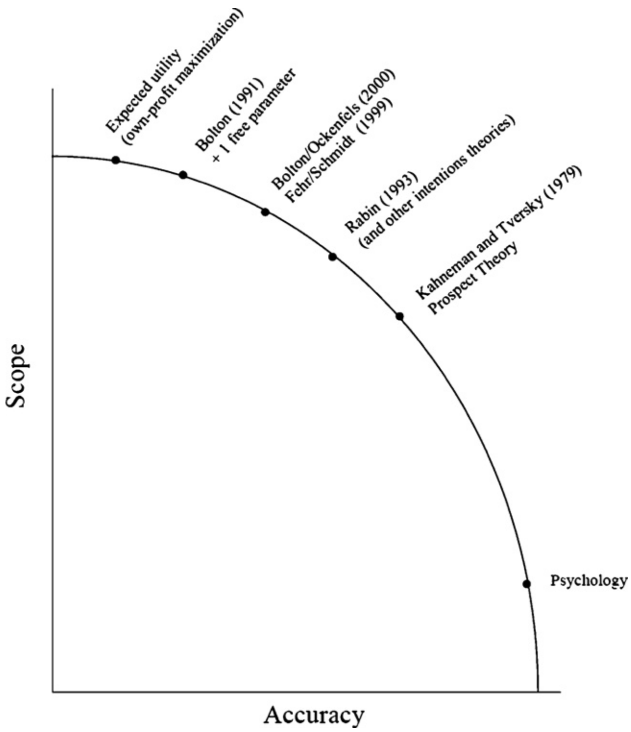

Fig. 1. A PPF of economic and psychological science: social preferences.

behavior cannot explain the data we observe from ultimatum games, but it can tell the reader what would happen if self-interested players played an ultimatum game. It is similarly useful to understand what would happen in a given context if people had (altruistic, envious, reciprocal, etc.) preferences. Theories can be evaluated on the new research they suggest, the concepts they illuminate, or the data collection (experimental or observational) and analysis they inspire.

In summary, there are many ways to evaluate a theory. Theories can be more or less accurate, have larger or greater scope, can explain/organize existing data or make (correct) predictions about new data. Theories can be right in their comparative static predictions, but wrong in their predictions about levels. But incorrect theories are also useful; they make predictions in domains where we might not observe data and they inspire data collection and analysis efforts which might not be otherwise undertaken.

7. Ten "commandments" of economic science (experimentalists and theorists)

Our article so far has outlined the theory of the interaction between economic theory and experiments (or, more generally, between economic theory and empirical work). However, as experiments and other empirical work tell us, there is often a long distance between theory and implementation. What should theorists do to make their work useful for experimentalists? What should experimentalists do to make their work useful for theorists? With apologies to David Letterman and to Moses, here are our top ten 'commandments' for experimentalists and theorists alike. A similar list could be created for empirical economists, but we leave that task to someone with more insight than we.

10. Experimentalists shall not Hypothesize After the Results are Known (HARK; see Kerr, 1998).
9. Experimentalists shall not criticize theory without suggesting (informally or formally) an improved alternative.
8. Experimentalists shall choose experimental parameters and designs that provide true tests of theories, ideally differentiating them from competing theories.
7. Experimentalists shall replicate and encourage replications, including making your data, instructions and software publically available.
6. Theorists shall not develop models in vain—no one needs a new model for every experimental or observational result.
5. Theorists shall not reject or criticize an experiment simply because its results do not support their theory.
4. Theorists shall respect optimization and equilibria—not because they are true (as quite a lot of research on bounded rationality has shown), but because they are unique to economics and should not be lightly put aside.
3. Theorists shall consider tradeoffs between parsimony and accuracy—we can write extremely accurate theories by adding many degrees of freedom, including saying that 'context matters' or 'social norms matter,' but these theories need loads of information in order to make a prediction (e.g., what is the social norm) and are thus difficult or impossible to falsify.

2. All economic scientists (theory, experimental or empirical) shall not oversell their results,and #1 on the top-10 list:
1. All areas of science (economics included) need to be tolerant of multiple methodologies.

Theory, lab experiments, field experiments, social experiments, neuroeconomics, observational research, simulations, surveys, case studies, etc., all these (and more) contribute to our understanding of the world. Scientists naturally develop comparative advantages in one or more of these methodologies, but these comparative advantages are only valuable in as much as they generate gains from trade. Intolerance of others' methods eliminates these gains.

8. Conclusion

This article has described our thoughts on the relationship between theoretical and experimental work in economics. We believe that practitioners of these two methodologies (as well as those who do other types of economics research) contribute significantly to our understanding of economic phenomena, decision-making in economic contexts, and policy analysis. However, the tolerant and constructive dialogue between them yields significant gains from trade and generates more insight and scientific progress than either in isolation.

References

Abbink, K., Irlenbusch, B., Pezanis-Christou, P., Rockenbach, B., Sadrieh, A., Selten, R., 2005. An experimental test of design alternatives for the British 3G/UMTS auction. European Economic Review 49, 505–530.
Andersen, S., Harrison, G., Lau, M.I., Rutström, E.E., 2008. Eliciting risk and time preferences. Econometrica 76, 583–618.
Andreoni, J., 1990. Impure altruism and donations to public-goods—a theory of warm-glow giving. Economic Journal 100, 464–477.
Bardsley, N., Cubitt, R., Loomes, G., Moffatt, P., Starmer, C., Sugden, R., 2009. Experimental Economics: Rethinking the Rules. Princeton University Press, Princeton.
Becker, G.S., 1974. Theory of social interactions. Journal of Political Economy 82, 1063–1093.
Becker, G.S., 1993. Nobel lecture—the economic way of looking at behavior. Journal of Political Economy 101, 385–409.
Bellemare, C., Kröger, S., 2007. On representative social capital. European Economic Review 51, 183–202.
Bellemare, C., Kröger, S., Van Soest, A., 2008. Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities. Econometrica 76, 815–839.
Benz, M., Meier, S., 2008. Do people behave in experiments as in the field?—evidence from donations. Experimental Economics 11, 268–281.
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.
Bohm, P., 1994. Behavior under uncertainty without preference reversal: a field experiment. in: Hey, J.D., (Ed.), Experimental Economics 19(2), 185–200.
Bohm, P., Lind, H., 1993. Preference reversal, real-world lotteries, and lottery-interested subjects. Journal of Economic Behavior & Organization 22, 327–348.
Bohnet, I., Greig, F., Herrmann, B., Zeckhauser, R., 2008. Betrayal aversion. Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. American Economic Review 98, 294–310.
Bolton, G.E., 1991. A comparative model of bargaining—theory and evidence. American Economic Review 81, 1096–1136.
Bolton, G.E., Ockenfels, A., 2000. ERC: a theory of equity, reciprocity, and competition. American Economic Review 90 (1), 166–193.
Buchan, N., Johnson, E.J., Croson, R., 2006. Let's get personal: an international examination of the influence of communication, culture and social distance on other regarding preferences. Journal of Economic Behavior & Organization 60, 373–398.
Burks, S.V., Carpenter, J., Götte, L., Monaco, K., Porter, K., Rustichini, A., 2007. Using behavioral economic experiments at a large motor carrier: the context and design of the truckers and turnover project. IZA Discussion paper no. 2789.
Caplin, A., Schotter, A. (Eds.), 2008. The Foundations of Positive and Normative Economics: A Handbook. Oxford University Press, Oxford.
Carpenter, J.P., Daniere, A.G., Takahashi, L.M., 2004. Cooperation, trust, and social capital in Southeast Asian urban slums. Journal of Economic Behavior & Organization 55, 533–551.
Chabris, C., Laibson, D., Morris, C., Schuldt, J., Taubinsky, D., 2008. Individual laboratory-measured discount rates predict field behavior. Journal of Risk and Uncertainty 37, 237–269.
Charness, G., Rabin, M., 2002. Understanding social preferences with simple tests. Quarterly Journal of Economics 117 (3), 817–869.
Chesney, T., Chuah, S.-H., Hoffmann, R., 2007. Virtual world experimentation: an exploratory study. CeDEx Discussion Paper No. 2007-14.
Cooper, D.J., Kagel, J.H., Lo, W., Gu, Q.L., 1999. Gaming against managers in incentive systems: experimental results with Chinese students and Chinese managers. The American Economic Review 89, 781–804.
Cox, J.C., Friedman, D., Gjerstad, S., 2007. A tractable model of reciprocity and fairness. Games and Economic Behavior 59, 17–45.
Croson, R., Gneezy, U., 2009. Gender differences in preferences. Journal of Economic Literature 47 (2), 448–474.
Davis, D.D., Holt, C.A., 1993. Experimental Economics. Princeton University Press, Princeton, NJ.
de Oliveira, A.C.M., Croson, R.T.A., Eckel, C., 2008. Are preferences stable across domains? An experimental investigation of social preferences in the field. CBEES Working Paper #2008-3, University of Texas at Dallas.
Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G., 2005. Individual risk attitudes: new evidence from a large, representative, experimentally-validated survey. IZA Discussion Paper Series No. 1730.
Dufwenberg, M., Kirchsteiger, G., 2004. A theory of sequential reciprocity. Games and Economic Behavior 47, 268–298.
Eckel, C.C., Grossman, P.J., 2008. Forecasting risk attitudes: an experimental study using actual and forecast gamble choices. Journal of Economic Behavior & Organization 68, 1–17.
Engelmann, D., Strobel, M., 2004. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. American Economic Review 94 (4), 857–869.
Falk, A., Fehr, E., Fischbacher, U., 2008. Testing theories of fairness—intentions matter. Games and Economic Behavior 62, 287–303.
Falk, A., Fischbacher, U., 2006. A theory of reciprocity. Games and Economic Behavior 54, 293–315.
Fehr, E., Fischbacher, U., von Rosenbladt, B., Schupp, J., Wagner, G.G., 2002. A nationwide laboratory examining trust and trustworthiness by integrating behavioral experiments into representative surveys. Schmoller's Jahrbuch 122, 519–542.
Fehr, E., Gächter, S., 2000. Cooperation and punishment in public goods experiments. American Economic Review 90, 980–994.
Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. Quarterly Journal of Economics 114 (3), 817–868.
Friedman, D., Cassar, A., 2004. Economics Lab. An Intensive Course in Experimental Economics. Routledge, London.
Friedman, D., Sunder, S., 1994. Experimental methods. A primer for economists. Cambridge University Press, Cambridge.
Gächter, S., Herrmann, B., 2007. The limits of self-governance when cooperators get punished—experimental evidence from urban and rural Russia. CeDEx Discussion Paper No. 2007-11, University of Nottingham.

Gächter, S., Herrmann, B., 2009. Reciprocity, culture, and human cooperation: previous insights and a new cross-cultural experiment. Philosophical Transactions of the Royal Society B-Biological Sciences 364 (1518), 791–806.
Grether, D.M., Plott, C.R., 1979. Economic theory of choice and the preference reversal phenomenon. American Economic Review 69, 623–638.
Guala, F., 2005. The Methodology of Experimental Economics. Cambridge University Press, Cambridge.
Hausman, J., 1979. Individual discount rates and the purchase and utilization of energy-using durables. Bell Journal of Economics 10, 33–54.
Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., 2001. In search of homo economicus: behavioral experiments in 15 small-scale societies. American Economic Review 91, 73–78.
Herrmann, B., Thöni, C., Gächter, S., 2008. Antisocial punishment across societies. Science 319, 1362–1367.
Holt, C.A., Laury, S.K., 2002. Risk aversion and incentive effects. American Economic Review 92, 1644–1655.
Huck, S., Müller, W., 2007. Allais for all: revisiting the paradox. CentER Discussion Paper No. 2007-99, University of Tilburg.
Kagel, J.H., Roth, A.E., 1995. The Handbook of Experimental Economics. Princeton University Press, Princeton.
Kagel, J.H., Roth, A.E., 2009. The Handbook of Experimental Economics, vol. II.
Kahneman, D., Tversky, A., 1979. Prospect theory—analysis of decision under risk. Econometrica 47, 263–291.
Kerr, N.L, 1998. HARKing: hypthesizing after the results are known. Personality and Social Psychology Review 2, 196–217.
Laibson, D., 1997. Golden eggs and hyperbolic discounting. Quarterly Journal of Economics 112, 443–477.
Lazear, E., 2000. Economic imperialism. Quarterly Journal of Economics 115, 99–146.
Levitt, S.D., List, J.A., 2007. What do laboratory experiments measuring social preferences reveal about the real world? Journal of Economic Perspectives 21, 153–174.
Loewenstein, G., Prelec, D., 1992. Anomalies in intertemporal choice—evidence and an interpretation. Quarterly Journal of Economics 107, 573–597.
Mas-Colell, A., Whinston, M., Green, J., 1995. Microeconomic Theory. Oxford University Press, Oxford.
Ockenfels, A., Roth, A.E., 2006. Late and multiple bidding in second price internet auctions: theory and evidence concerning different rules for ending an auction. Games and Economic Behavior 55, 297–320.
Oosterbeek, H., Sloof, R., van de Kuilen, G., 2004. Cultural differences in ultimatum game experiments: evidence from a meta-analysis. Experimental Economics 7, 171–188.
Plott, C.R., 1982. Industrial-organization theory and experimental economics. Journal of Economic Literature 20, 1485–1527.
Plott, C.R., Porter, D.P., 1996. Market architectures and institutional testbedding: an experiment with space station pricing policies. Journal of Economic Behavior & Organization 31, 237–272.
Plott, C.R., Smith, V.L. (Eds.), 2008. Handbook of Experimental Economic Results, vol. 1. North-Holland, Amsterdam.
Popper, K., 2002. The Logic of Scientific Discovery. Routledge Classics, London.
Rabin, M., 1993. Incorporating fairness into game-theory and economics. American Economic Review 83, 1281–1302.
Romer, D., 2005. Advanced Macroeconomics, 3rd ed. McGraw-Hill/Irwin, New York.
Roth, A.E. (Ed.), 1995. Introduction to Experimental Economics. Princeton University Press, Princeton.
Roth, A.E., 2002. The economist as engineer: game theory, experimentation, and computation as tools for design economics. Econometrica 70, 1341–1378.
Rubinstein, A., 1991. Comments on the interpretation of game theory. Econometrica 59, 909–924.
Samuelson, L., 2005. Economic theory and experimental economics. Journal of Economic Literature 43, 65–107.
Schmidt, K.M., 2008. The role of experiments for the development of economic theories. Working paper Ludwig Maximilans University of Munich.
Schmidt, U., Starmer, C., Sugden, R., 2008. Third-generation prospect theory. Journal of Risk and Uncertainty 36, 203–223.
Segal, U., Sobel, J., 2007. Tit for tat: foundations of preferences for reciprocity in strategic settings. Journal of Economic Theory 136, 197–216.
Smith, V.L., 1962. An experimental-study of competitive market behavior. Journal of Political Economy 70, 111–137.
Smith, V.L., 1976. Experimental economics—induced value theory. American Economic Review 66, 274–279.
Smith, V.L., 1982. Microeconomic systems as an experimental science. American Economic Review 72, 923–955.
Starmer, C., 2000. Developments in non-expected utility theory: the hunt for a descriptive theory of choice under risk. Journal of Economic Literature 38, 332–382.
Stroop, J.R., 1935. Studies of interference in serial verbal reactions. Journal of Experimental Psychology 18, 643–662.
Sugden, R., 2005a. Experiment, theory, world: a symposium on the role of experiments in economics. Journal of Economic Methodology 12, 177–184.
Sugden, R., 2005b. Experiments as exhibits and experiments as tests. Journal of Economic Methodology 12, 391-302.
Tanaka, T., Camerer, C.F., Nguyen, Q., forthcoming. Risk and time preferences: linking experimental and household survey data from Vietnam. American Economic Review.
Thaler, R., 1991. The Winner's Curse: Paradoxes and Anomalies of Economic Life. Free Press, New York.
Tversky, A., Kahneman, D., 1992. Advances in prospect-theory—cumulative representation of uncertainty. Journal of Risk and Uncertainty 5, 297–323.
Weber, E., Blais, A., Betz, N., 2002. A domain-specific risk-attitude scake: measuring risk perceptions and risk behaviors. Journal of Behavioral Decision Making 15 (4), 263–290.