IS PSYCHOLOGY THE FUTURE OF ECONOMICS?

Douglas Gale
Silver Professor of Economics

September 2005

NEW YORK UNIVERSITY

Is Psychology the Future of Economics?

I

For most of my career, being an economic theorist has involved building models of economic phenomena using two fundamental ideas, rational choice and equilibrium, as building blocks.

The assumption of rational choice in a model of human behavior is not as restrictive as it sounds. It simply requires that each decision maker have consistent preferences over all possible alternatives and that he choose the most preferred alternative from the feasible set. Consistency does not rule out preference for status or power; nor does it rule out feelings of envy and altruism. Consistency is an empty box and we can fill it as we wish.[1] Equilibrium, by contrast, is a more restrictive concept. One definition of equilibrium, due to Nash, plays a central role in the theory of games. It assumes that each player chooses a strategy that maximizes his payoff taking as given the strategies of his opponents. In other words, each player chooses a best response to the strategies of the other players.

The *neoclassical program* of explaining every economic phenomenon in terms of equilibrium and rational choice has been extremely successful over the

last fifty or sixty years. The entire discipline of economics has been reduced to the principles of rational behavior and economic equilibrium. Macroeconomics and microeconomics, which were once thought to rest on different principles, now share a common language and a common structure. Techniques that are developed in one field of economics are quickly applied to other fields or even to problems that were not previously thought to be amenable to economic analysis. As a result, we now have mathematical models of everything from marriage and divorce to ethnic conflict and from her behavior to suicide, as well as the more traditional topics of interest rates and prices. Theoretical models have also been crucial to empirical applications in every field of economics.

A recent interest of mine is the application of theoretical models to experimental data. It is a truism that empirical work in the social sciences is limited by the impossibility of conducting large-scale, controlled experiments. However much data we collect, the fact that everything is changing at once makes it very hard to trace causal relationships or disentangle the effects of different factors. There is no way to change in a single variable, holding everything else constant, which is what one would do in a classical experiment.

Another characteristic of economic data which makes empirical analysis difficult is the *endogeneity* of the variables we do observe. An example will make this clear. Suppose we are interested in determining the effect of education on earnings. In a classical experimental framework, we would attempt to measure this effect by taking a representative sample of the population, dividing it into

two or more groups, giving different amounts or qualities of education to each group and then, after the passage of several years, turning the subjects loose to see how they fare in the labor market. Because individuals are heterogeneous (e.g., have different abilities) and we cannot observe their innate characteristics, we are faced with a problem of statistical inference: how to distinguish the impact of innate characteristics, such as ability, from the impact of education on labor earnings. But since the treatment (education) is independent of ability, a large sample will reveal the average effect of education on earnings independently of the distribution of abilities.

Unfortunately, this is not how the data we observe are generated. Each individual uses private information to make choices about the amount of schooling she receives. A student with high ability may choose more education than a student with low ability because she finds it easier or more enjoyable or because she thinks the return to education is higher for her than it is for someone else. If people who choose more education tend to be more able, we cannot conclude that additional schooling is solely responsible for the higher earnings.

Although we cannot conduct experiments on the US economy, we can conduct very small scale experiments in the laboratory. The great attraction of experiments, in the social sciences and elsewhere, is that they allow us to control all the relevant variables. When subjects play a game in the laboratory, we know precisely the rules of the game, the information available to the subjects, and the payoffs corresponding to each profile of strategies, because we

have designed the experiment and chosen the parameters ourselves. The hope is that the experimental design will allow us to observe the effect of a change in the variable we are interested in, holding everything else constant. Unfortunately, the idea that experiments in the social sciences are just like experiments in the physical sciences is a delusion. We may be able to control some variables, but there are many important variables that we cannot control. In particular, while we can specify monetary payoffs for different outcomes of a game, we cannot prevent subjects from having their own objectives, such as wanting to win the game for its own sake or to frustrate an opponent. Likewise, we cannot prevent the occasional subject from making absurd choices out of boredom or a sense of mischief. Every subject is hopelessly contaminated by the "frame" that real life puts around the laboratory experiment.[2] One might hope that framing effects will wash out if there is enough data and the framing effects are purely random; but if the effects are systematic, either because the subjects are influenced by common factors or because they interact in the course of the experiment, simply averaging the data will not work. Because we cannot control all of the factors that may influence the behavior of subjects in an experiment, we need to use economic theory to make sense of the data.

Economic theory helps us design and interpret experiments in several other ways. A few examples may help to make this clear.

1. Backward Induction and Cooperation

An example of how economic theory can motivate an experimental design is found in Choi, Gale, and Kariv (2005b). The main objective of this work was to test a theoretical prediction that rational behavior implies a higher degree of cooperation in a class of dynamic games. The essential idea is illustrated by the following story.

> "A common feature of international treaties on the environment, such as the Kyoto Protocol, is a *minimum participation clause*. This clause states that the treaty does not become binding until it has been ratified by a certain number of signatories. Suppose that k is the minimum participation requirement and suppose there exists a set of at least k countries that all prefer to ratify the treaty assuming that $k - 1$ other countries in the set also ratify it. The purpose of the minimum participation clause is to protect signatories from the consequence of uncertainty about the number of eventual participants. It may be very difficult to coordinate the simultaneous ratification of the treaty by k countries, especially when we take into account the different political processes and interest groups in each country. The minimum participation requirement ensures that no country is disadvantaged by ratifying the treaty before $k - 1$ others do. On the one hand, if $k - 1$ other countries ratify the treaty, the ratifying country is better off. On the other hand, if some of the others fail to ratify the treaty, the ratifying country is no worse off because the treaty does not oblige the country to do anything. Thus, the minimum participation clause protects countries against free riders and, at the same time, it gives countries that are tempted to take a free ride an incentive to participate."

There are two key ideas in this example. The first is "backward induction," the idea that rational players will anticipate the rational responses of other players to their (the first players') actions. The second is that some actions are irreversible so that once a player makes a move (e.g., ratifies the treaty) he is committed to it, at least for some period of time. The combination of these two properties leads rational players to cooperate in signing the treaty. This backward induction argument holds for a broad class of games. To see whether human subjects can use backward induction to achieve cooperation, we asked them to play the following game.

> There are three players, each endowed with one token. The game is divided into five periods. In each period, the players simultaneously decide either to contribute a token to a public project or to keep it for another period. Once a player's token has been contributed, he cannot get it back. At the end of the game, if the number of tokens contributed is two or more, the project is completed and each player receives three tokens *plus* his own token if he has not contributed it. If the number of tokens is less than two, the project is not completed and each player keeps his own token if he has not contributed it.

At the end of the experiment, tokens are exchanged for money, so the player's monetary payoff is increasing in tokens.

This game is characterized by a free-rider problem: other things being equal, a player wants the project completed, but he would rather keep his token than contribute it. In other words, he wants someone else to pay for the project. In some settings, the free-rider problem means that the project will

not be completed in equilibrium. For example, if the players only have one chance to contribute, there exists an equilibrium in which no one contributes and the project is never completed. However, if the number of periods is greater than the number of tokens needed to complete the project, backward induction guarantees that the project is always completed with positive probability.

There are several problems with trying to "test" for backward induction."[3] Economic theory does not predict a single outcome for the underlying game, but rather a whole host of outcomes, each one corresponding to a different equilibrium. If a game has a unique equilibrium, it is possible to predict how a change in one or more parameters will affect the outcome. The multiplicity of equilibria makes such comparative static properties scarce. If subjects switch from one equilibrium to another as a parameter changes, the outcome will change in unexpected ways. In this case, we used the comparative static properties of the equilibrium set. For some parameter values, all equilibria are characterized by the property that the public good is provided with positive probability. For others parameter values, there also exist equilibria in which no provision occurs. The experimental design was set up so that different treatments corresponded to equilibrium sets with and without these no provision equilibria. The inclusion of a no-provision equilibrium was taken as a prediction that the probability of provision would be low. Some of the comparative static properties are surprising, for example, the finding that increasing initial endowments introduces an equilibrium with no provision.

A second problem in interpreting the data is that cooperation may arise for reasons other than backward induction. To identify a cooperative outcome as being the result of backward induction, one has to show that the subjects' behavior is more consistent with the theoretical model than with other explanations. Fortunately, economic theory provides many properties of equilibria that can be tested using experimental data. We used these properties to test the entire equilibrium theory and not just the prediction about the degree of cooperation. The more points at which the data fits the theory, the more likely it is that cooperation, if it arises, follows from backward induction, rather than from motives that are extrinsic to the game.

In these and other ways, a clear theoretical analysis allows one to understand the elements of a complex experiment and to make maximum use of the resulting data. Sometimes an experiment is designed to test a single "fact." Theoretical analysis may be less important because the data "speaks for itself," but the most interesting economic questions are not of this kind. Often we are interested in the subjects' interaction, for example, playing a game, and we may want to know something about their motivations or expectations or strategies, things which are not directly observable. The role of theoretical analysis in helping us understand what is going on in a complex game is well illustrated by the another piece of work, Choi, Kariv, and Gale (2000a), which studied social learning in networks.

2. Social Learning

Social learning refers to the process of information acquisition that occurs when one person observes a choice made by another person. Examples would be observing the make of car your neighbor drives, seeing the number of people going into a popular restaurant, or hearing that a neighboring farmer had planted a new type of grain. All of these observations provide information and may influence the decisions of the observer. In practice, each individual observes the choices made by a limited set of other individuals, which we call his "neighborhood." Information diffuses through the social network as each individual observes his neighbors, updates his beliefs and revises his behavior, and is observed in turn by other individuals.

This kind of learning raises some interesting questions about how individuals process information. If I observe someone adopting a new technology, I do not know whether he has private information about the value of that technology or is just adopting it because someone else adopted it (an example of "herd behavior"). In general, it may be very difficult to determine the informational value of an observation because it can depend on what has happened throughout the network. In principle, a rational individual can take into account all possible eventualities and their respective probabilities and then use Bayes' theorem to update his own beliefs about the value of a new technology, but whether individuals can actually do this, even in simple settings, is open to question. Even if individuals are perfectly rational, the result of social learning

may be inefficient. An important example of inefficiency is "herd behavior, which occurs when an individual rationally ignores his own information and imitates what he sees others doing. One purpose of the experiments reported in this paper was to see whether herd behavior is more likely in some networks than in others. Simply observing herd behavior is relatively easy. The hard part is determining whether the observed behavior is explained by the theoretical analysis of the model or is extraneous to the model. To "test" the validity of the theory, we used the following experiment.

> There are two urns, a "white" urn containing two white balls and one red ball and a "red" urn containing one white ball and two red balls. One of these urns is chosen by the computer and then a random selection of the subjects are given private information in the form of a ball drawn at random (with replacement) from the chosen urn. With this information, a subject is able to update his beliefs and make an informed guess as to which urn had been chosen. Subjects are asked to guess which urn has been chosen on six occasions, with different information each time. The first time, they have only their private information (the randomly drawn ball) to guide them. The second time, each subject is able to observe the previous prediction of one or more of the other subjects. How many others can be observed depends on the network structure used in the experiment. For example, in a circle network with three subjects, each subject observes exactly one other subject. After observing their neighbors' previous guesses, the subjects make a second prediction based on this information. At the third turn, the subjects observe their neighbor's choice of urn at the second turn, revise their beliefs again and make a new choice. This process continues until six decisions are made.

Although the setup is extremely simple, the decision problem can be very complex, requiring individuals to form beliefs about what their neighbor has seen his neighbor doing and what it implies about his (the neighbor's neighbor's) information. For example, consider a three-person game played by rational agents, in which each player receives a signal with probability ⅔. Assume that the network is star-shaped with player A in the center and B and C in the periphery, that is, player A can observe what B and C do but B and C can only observe what A does. The table below shows a particular profile of signals received and the corresponding equilibrium decisions made by the players.

	Player/Signal		
	A	B	C
Period	r	w	$\varnothing$
1	R	W	R
2	R	W	R
3	R	W	R
...	...	...	....

At the start of the game player A has seen a red ball, player B has seen a white ball, and player C has received no signal. A player who has seen a red ball drawn from the urn thinks that the red urn is more likely to have been chosen and conversely if he sees a white ball. At the first decision, players A and B will guess R and W respectively. Player C, being uninformed, thinks the two urns are equally likely, so he guesses randomly and ends up choosing R. At the second

turn, *A* observes the complete history of the predictions by *B* and *C* only observe *A*'s prediction. *B* will continue to choose *W* because he is informed whereas *A* might be uninformed; *C* will continue to choose *R*, because he is uninformed whereas *A* might be informed; *A* will continue to choose *R* because, from his point of view, *B* and *C* choices cancel each other.

At the third decision, *A* knows that *B* is informed (he can infer this from the fact that *B* does not switch at the second turn) and *C* may be informed or uninformed. The signals of *A* and *B* cancel each other and *C* tips the balance toward *R*, so *A* continues to choose *R*. Now *B* can infer that *A* is either informed or else is uninformed but observed *C* choose *R* at the first turn. Eventually, if *A* and *C* continue to choose *R*, their information overwhelms *B*'s information and *B* will switch to *R*. Notice that all three players have learned, in varying degrees, that the urn is more likely to be red, even though, conditional on the signals received, both colors are equally likely.

Deciding whether subjects are behaving rationally requires economic theory, not just because individual decision rules are complex, but because we have to solve for the equilibrium of the model in which everyone's behavior depends on everyone else, on the information available, and on the network architecture.

There are various measures that one can use to judge how well economic theory fits the data. The only way to take into account all of the data available is to use it to estimate the decision rules and compare them to 'equilibrium strategies. In this case, economic theory does a pretty good job of accounting for the

experimental data, but inevitably there are errors. To account for these errors we adopt a structural approach. First, we explicitly allow for the possibility of errors in our theoretical model. In this case, we used a model of Quantal Response Equilibrium (QRE), in which the probability of making an error is inversely related to the payoff difference between the optimal prediction and the alternative. In other words, subjects are assumed to be more likely to make a mistake when there is very little at stake. Secondly, we recalculate the equilibrium theory taking into account the fact that individuals make errors. The fact that players make mistakes changes the meaning of an optimal prediction. For example, if I know a particular subject makes mistakes, often guessing the white urn even though their information indicates the red urn was chosen, then I should put much less weight on the observation of that subject's prediction in updating my beliefs. Also, if some decisions are more difficult than others, the likelihood of mistakes may vary during the game and this too should be taken into account by rational players. Finally, we estimate the parameters of the QRE model so that it minimizes the difference between the predictions of the QRE model and the observed behavior.

In estimating the QRE, we are basically determining how much influence the optimal strategy has on the subjects' choices. At one extreme, it could be the only thing that matters; at the other, it could have no influence so that choice is essentially random. The larger the parameter measuring the influence of the optimal choice, the better the economic theory predicts the data. In this

case we developed a method to estimate the QRE recursively: first, we estimated a QRE of the first decision, then we used those estimates to determine the payoffs from the different predictions at the second turn and used those payoffs to estimate the QRE for the second turn, and we continued in this way until all the data had been used.

3. Estimating Cognitive Hierarchies

Although the QRE gives a fairly good account of the average data, it appears on closer inspection that there is heterogeneity in the subjects' behavior. Some subjects' behavior is much closer to the optimal behavior described by the economic theory than others'. This is hardly a surprise. When subjects behave in a way that is approximately optimal, it is usually because they have discovered a heuristic that does fairly well in that situation. Different individuals will use different heuristics (or sets of heuristics) resulting in a certain degree of heterogeneity. In a remarkable extension of this research, Choi (2005) identifies a number of these heuristics and treats them as behavioral "types" which characterize a subject's play in a particular game. Each type corresponds to a different level of cognitive ability and determines how much information he can usefully process. The lowest type cannot process any information and makes random predictions. The second lowest type can only process his own private signal and makes a prediction based on that at each turn. The third lowest can process his own signal plus the information he obtains from observing other choices at the

first turn, and so on. Choi constructs a model based on this hierarchy of types in which each player, faced with a distribution of types as opponents, responds optimally given the information he processes. Using the experimental data, it is actually possible to calculate the probability distribution of the different types. This tells us not only how many subjects behave like fully rational subjects and how many have bounded rationality, but it also tells us the behavior of a rational player who recognizes the presence of different types would differ from the behavior of players when there was common knowledge of rationality.

The cognitive hierarchy model provides a much better account of the data than the Bayesian model. This may not be surprising since the Bayesian model is a special case of the cognitive hierarchy model; however, the criterion used to estimate the model does not guarantee that the model will give a better fit according to other criteria such as predictions of herd behavior. More importantly, by distinguishing different cognitive types, this exercise gives us a precise sense of how rational individual subjects are. Choi finds that the proportion of rational types is very high, sometimes greater than 80%, and their behavior fits the predictions of the economic theory quite well. Thus, recognizing bounded rationality and incorporating an economic theory of mistakes actually provides a strong confirmation of the relevance of rational behavior. Of course, the meaning of rational behavior has changed somewhat: these rational players have to take into account the existence of boundedly rational players in the population.

II

Despite the success of the neoclassical program, economists have recently shown a growing interest in exploring aspects of human behavior that have traditionally been the province of other social sciences. The new fields of *behavioral economics* and *neuroeconomics* address new questions and use new methods. Behavioral economics has its origin in experimental studies which challenged the empirical validity of economic models of rational choice. For example, the famous Allais paradox provides an example of preference reversals when subjects make choices between pairs of lotteries that are theoretically equivalent. Thus, a subject may prefer lottery *A* to lottery *B* and lottery *B'* to lottery *A'*, even though the pair (*A, B*) is isomorphic to (*A", B'*). Paradoxes such as these have stimulated a lot of good research, some of which has led to the development of new models. Examples include prospect theory, which explains systematic empirical biases observed in decisions under uncertainty, and the B, Smodel of time preference, which explains present bias in intertemporal decisions. In other cases, empirical puzzles have encouraged economists to try to extend the classical paradigm to account for these anomalies. In the last few years, there have been attempts to study phenomena such as self-control or memory or anxiety using the tools of economic theory, that is, maximizing behavior and equilibrium. At its best, behavioral economics does not mean abandoning the classical paradigm; it means applying it more creatively.

Part of the excitement about the prospects of behavioral economics is

driven by the rapidly growing use of experimental methods to explore behavioral phenomena. While laboratory experiments involving human subjects have been a standard research tool in social psychology for many years, until recently they were the province of a small group of specialists in economics. Most departments did not even have an experimental laboratory ten years ago. Now experimental laboratories are common and may be considered essential in a major department. One exciting branch of experimental economics is the budding field of neuroeconomics, in which economists and neuroscientists collaborate to study the brain function of subjects who are performing economic tasks. The new questions about individual behavior, the new methods of empirical research, and the prospect of revolutionary advances in behavioral economics provide a very exciting combination which is certain to attract many of the brightest economic minds in the years to come.

Whereas neoclassical economics tries to explain every economic phenomenon in terms of rational choice and equilibrium, it sometimes seems that anything goes in behavioral and experimental economics. The new empiricism is often motivated by naive intuitions and ideas borrowed from psychology and sociology. One of the risks of the new empiricism is that, without the discipline of careful economic reasoning, the pursuit of "facts" will end up producing a large number of well established anecdotes, but no generalizable laws governing economic phenomena. The best way to understand why this trend is both worrying and seductive is to consider a couple of examples. The

two papers I am going to discuss are very recent (one as yet unpublished) and have already received a lot of attention. They appear to demonstrate remarkable and important results in areas where little or no previous research has been carried out by economists. Part of their charm is that they rely on experimental methods that are easily understood even by laymen and avoid making reference to difficult mathematical models.

1. Oxytocin and Trust

Kosfeld, Heinrichs, Zak, Fischbacher and Fehr (2005) observe that

"Trust pervades human societies. Trust is indispensable in friendship, love, families and organizations, and plays a key role in economic exchange and politics. In the absence of trust among trading partners, market transactions break down. In the absence of trust in a country's institutions and leaders, political legitimacy breaks down. Much recent evidence indicates that trust contributes to economic, political and social success. That trust is important ingredient of society who wish to investigate the biological basis of trust in humans."

They go on to note that "Little is known, however, about the biological basis of trust among humans." In the remainder of the paper they attempt to establish that the "intranasal administration of oxytocin, a neuropeptide that plays a key role in social attachment and affiliation in non-human mammals, causes a substantial increase in trust among humans."

In order to carry out this research, one needs a measure of the degree

of trust exhibited by subjects. Since there is no direct way to measure trust, they have the subjects play a simple but ingenious game of economic exchange, known in the experimental literature as the *trust game*. The game is defined as follows:

> There are two players, an *investor* and a *trustee*, and each is given an initial endowment of money. The investor moves first and makes an ex gratia payment (the *transfer*) to the trustee. The trustee receives an amount equal to three times the original transfer (the experimenter adds an amount equal to twice the transfer) and adds it to his endowment. Then it is the trustee's turn to move and he can make an ex gratia payment (the *back-transfer*) to the investor. After both players have moved the game ends and the players receive their payoffs. The investor's payoff equals his initial endowment *minus* the transfer *plus* the back-transfer. The trustee's payoff equals his initial endowment *plus* three times the transfer *minus* the back-transfer.

The use of this game in the experiment illustrates an important product of the collaboration between economists and neuroscientists. Economists have a large reservoir of games representing different phenomena. The cognitive tasks that subjects perform in these games allow neuroscientists to observe the performance of the brain in well defined situations that correspond to activities that are of interest in real life.

The trust game was introduced in a paper by Berg, Dickhaut, and McCabe (1995). According to classical game theory, the game has a unique equilibrium, in which the investor transfers nothing to the trustee and the trustee transfers

nothing to the investor. To see this, consider the trustee's decision once the transfer has been made by the investor. The trustee is assumed to care only about his own payoff, which is defined to be equal to the amount of money he holds at the end of the game. Obviously, the way to maximize this amount, taking the behavior of the investor as given, is to choose a back-transfer of zero. So, whatever transfer the investor chooses, the best response for the trustee is to transfer nothing. If the trustee adopts this strategy, then the best response by the investor is to keep his money for himself. Anything he gives away reduces his payoff and he gets nothing in return. Thus, the two strategies described constitute the unique equilibrium of the game.

Although the state of affairs we have described is an equilibrium, it is not a very happy one. Each player is maximizing his welfare, given the behavior of the other, but he fails to realize any gains from trade. Since any money transferred by the investor is tripled and both players can be made better off by an appropriate back-transfer, the players' welfare is maximized only if the investor transfers all of his initial endowment. For example, suppose that each player begins the game with $10. Then in the equilibrium, since there are no transfers, each player receives a payoff of exactly $10. But if the investor were to transfer $10 to the trustee, the trustee would have $10 + 3 \times $10 = 40 to divide between the two of them. Any back-transfer between $10 and $30 would give them each more than the $10 they receive in the equilibrium.

The theoretical prediction is clear. What happens in practice? In many

implementations of the trust game, contrary to economic theory, subjects in the role of the investor make positive transfers and subjects in the role of trustees make positive back-transfers. Moreover, in the study of Kosfeld et al., those who received the oxytocin made larger transfers and backtransfers than those who received the placebo. The authors conclude that the application of oxytocin increases trust.

When it comes to the economic interpretation of the results, the lack of theory raises some problems.

The definition of trust. Without a definition in terms of observables, the result is not generalizable. We have no idea how to apply the results of this experiment to other situations. Of course, we all have an intuitive notion of what trust means and how it influences our behavior, but these subjective understandings are not enough. We need something that can be objectively quantified in every situation. That is one of the major problems with a theory-free approach to experimentation.

Identifying the operation of trust. Berg et al. say that "To trust someone implies taking some risk. To be trustworthy implies that you will try to reduce risks to those who trust you even if it is costly to you to do so. For economists trust reduces the transaction cost of trading." Other experimenters have established a positive correlation between transfers in the trust game and measures of trust based on surveys, reports of past behavior, or self-evaluation by subjects. So there may be something to the argument that the game does

measure trust, but we still do not know what trust is in terms of observable characteristics and we cannot be sure that the reason why individuals are cooperating in this context has anything to do with trust.

The natural way for an economist to explain the outcome would be to hypothesize a different set of preferences. In other words, assume that the players care about something other than their own monetary payoffs. For example, suppose the trustee has preferences over the monetary payoffs received by both players, investor and trustee. This defines a new game and the standard analysis implies that the trustee will choose a back-transfer to maximize his preferences given the amounts held by both players at the beginning of the second stage. A result due to Gary Becker and known as *The Rotten Kid Theorem* says that as long as the back-transfer is positive, the equilibrium distribution of money between the two players is a function of the total amount of money at the second stage. So, if the investor's payoff is a normal good for the trustee, it will be an increasing function of the total amount of money at the second stage. In equilibrium, the investor will maximize his own payoff if he acts in such a way as to maximize the total amount of money in the game. Obviously, he does this by transferring all of his money to the trustee. Thus, under quite mild conditions, the assumption of altruistic preferences on the part of the trustee implies the investor transfers the largest possible amount.

There is nothing complicated or non-standard about this explanation. At the same time there is no reason to interpret any of this in terms of the

concept of trust. What we are describing is altruism on the part of the trustee. That was certainly Becker's interpretation of his model. The trustee will make a positive transfer to the investor whether the investor "trusts" him or not. The investor, for his part, is only exhibiting "trust" to the extent that he expects the trustee to choose a best response, as he would in any equilibrium.

Assuming game theory is valid. Trust is measured by the deviation of the observed transfer from the prediction of classical game theory. The authors appear to reason as follows: "Game theory predicts no transfers, so positive transfers must imply the existence of trust." This is one of the most curious aspects of the experiment, for it assumes something that is apparently falsified by the data.

Accounting for the data. There are several reasons for wanting to account for the experimental data using a structural model. The most important is that, until we can structurally account for the data, we cannot claim to have shown that the outcome is the result of trust or anything else. Another is that, having estimated a structural model of the data, we can use it for comparative static exercises that will tell us how the outcome will change if we change the parameters. Another is that there are many reasons why subjects cooperate in an experiment. It will reassure us that we are on the right track if the data fits the model at several points. In the trust-game literature, the approach seems to be that a deviation from the standard theory allows one to draw any conclusion one likes.

The theoretical explanation offered earlier does not explain why the

observed transfers are less than the maximum in the experimental data. This might be the result of risk aversion: if the investor is uncertain whether the trustee has altruistic preferences, a risk averse investor would protect himself against being left with nothing by withholding part of his initial endowment as insurance. Partial transfers might also be explained by the investor's sense of fairness. The Rotten Kid Theorem assures us that the investor will transfer all his money if his payoff is increasing in total income, but this is consistent with the trustee taking most of the gains and leaving the investor with only slightly more than his initial endowment. If this strikes the investor as unfair, he may refuse to cooperate fully. By reducing his transfer the investor makes himself slightly worse off, but the trustee is much worse off and this may be a preferred situation from the investor's point of view.

The bottom line is that, the more we try to understand the experiment in terms of economic theory, the more obscure the motivations of the subjects and the implications of the data become. Whereas even the superficial theorizing in the preceding paragraphs suggest interesting questions and hypotheses, the experiment as presented is pretty much a dead end in the sense that it does not suggest more complex explanations or directions for future research.

2. Women and Competition

Larry Summer's recent remarks have focused attention once again on the contentious question of why so few women occupy the highest positions

in academia and the business world. Among the explanations that have been offered are differences in career preferences, differences in ability, discrimination, and domestic responsibilities. Statistical studies have to cope with the possibility that all of these factors may be present, so it is hard to attribute the observed disparities to a single factor. Muriel Niederle and Lise Vesterlund (2005) have designed an experiment to address one aspect of this puzzle, whether women have a lower tolerance for competition. Unlike the real world, where a variety of factors are always in play, an experiment provides the chance to study women's choices in a setting where there are no differences in ability or workload and there is no discrimination.

> The experiment consists of a series of simple mathematical tasks. The subjects are formed into groups of four, each group consisting of two women and two men. The tasks consist of simple problems in addition. The subjects complete as many of them as they can in a limited period of time. There are four experimental treatments, corresponding to different reward schemes. In the first treatment, subjects are paid a piece rate of fifty cents for each correct answer. In the second treatment, the subjects play a tournament in which the one who gets the most correct answers is paid two dollars for each correct answer and the others receive nothing. In the third treatment, subjects are given the choice of being paid the piece rate or being paid according to the tournament, where the winner is determined by the subject's current performance compared to the performance of other three subjects in the second treatment. In the final treatment, subjects do not have to perform any calculations. They are simply asked to choose whether they would like to be rewarded for their performance under the first treatment according to a piece rate or

a tournament, where the winner of the tournament is determined by the performance of subjects under the first treatment.

Their striking observation is that, given the choice between a piece rate and a tournament, men are twice as likely to choose the tournament as women. The authors conclude that, relative to men, women tend to avoid competition. The experiment illustrates many of the features of a good experimental design. It begins with an interesting question, it presents the subjects with a transparent and familiar task, and it provides a rich data set. Although the authors do not present a formal theory, they are aware of other explanations that might be offered for the subjects' choices and use their design to isolate subjects' attitudes toward competition. For example, a subject might feel confident that she will win the tournament, but prefer the piece rate because it seems fairer than the tournament, in which the winner takes all. The third and fourth treatments avoid this confounding effect, because the subject's choice of reward scheme has no impact on the payoffs of the other subjects. Thus, attitudes towards fairness should have no effect on the subject's choice in those tasks. Although this is a clever design, the lack of a formal economic theory creates problems when it comes to interpreting the results. This is unfortunate because it leaves many questions unanswered.

As in the experiment of Berg et al., the classical economic theory is always there in the background. Here the authors use it as an implicit benchmark, assuming that any deviation from the outcome predicted by the classical economic theory must be attributable to a preference for, or aversion to, compe-

tition. The authors argue that if subjects have the same average ability, they would each expect to win the tournament with probability 0.25, so the expected earnings from the tournament would be fifty cents for each correct answer. Then risk neutral subjects should be indifferent between the tournament and the piece rate. If subjects are not indifferent, it must reflect some other preference. Because they present evidence that average abilities are the same, they conclude that the explanation must lie in different attitudes toward competition. Note that evidence of equal abilities is not the same thing as evidence of equal *perceptions* of ability. Given popular beliefs about the distribution of mathematical ability, perceptions may well be an independent factor in the outcome of the experiment.

Even if we assume that subjects are risk neutral and have the same average (perceived) ability, it does not follow that they should be indifferent according to the classical economic theory. If men have a higher variance than women, women may rationally prefer the piece rate. For example, suppose that women always get 12 answers correct, whereas men are equally likely to get 14 or 10 correct. Although the average number of correct answers is the same, the probability that a man will win the tournament is 0.75, so a woman would prefer the piece rate. Using an explicit model might have made the authors aware of this pitfall.

In addition, all the issues mentioned in connection with the experiment of Kosfeld et al. are present here. The idea of preference for competition is so obscure that we have no idea how to analyze its role in this particular game,

still less to apply it to other settings. How do we know which of two careers is more competitive? What observable characteristics determine competitiveness? We cannot use gender differences to measure competitiveness because we know there are other factors that are surely important. Similarly, we cannot be sure that what we are observing here is caused by what we perceive to be a difference in competitiveness.

III

The role of economic theory is to check the logic of arguments, to suggest interesting questions, to allow us to clarify concepts and organize thoughts and data. All of these functions are needed in the design and interpretation of experiments, but they are lacking when there is no theoretical framework. Experimental economics and neuroeconomics offer the prospect of new discoveries that may advance our understanding of economic phenomena. This potential will only be realized if experimental results can be given a coherent economic interpretation. The danger is that the lure of generating easy results using experimental methods and the excitement of playing amateur psychologist may lead economists to abandon the analytical method without which there can be no economics.

---

1 These days even consistency is too much for some economists. There is a very active literature on the implications of intertemporal inconsistency in preferences.

2 We might note in passing the difference between psychology, where framing effects are an essential object of study, and economics, where they are considered a nuisance.

3 The word "test" is used loosely in this context. All theoretical models are approximations. Whether a particular model is "close enough" is bound to be a subjective judgment. It would be more accurate to say that we are comparing theoretical predictions with the data, without assigning a pass or fail grade.

References

Berg, J., J. Dickhaut, and K. McCabe (1995). "Trust, Reciprocity, and Social History," *Games and Economic Behavior* 10, 122-132.

Choi, Syngjoo (2005). "A Cognitive Hierarchy Model of Learning in Networks," New York University, unpublished.

—, D. Gale, and K. Kariv (2005a). "Learning in Networks: An Experimental Study," New York University, unpublished.

—, —, and — (2005b). "Sequential Equilibrium in Monotone Games: A Theory-Based Analysis of Experimental Data," New York University, unpublished.

Kosfeld, M., M. Heinrichs, P. Zak, U. Fischbacher and E. Fehr (2005). "Oxytocin increases trust in humans," *Nature* 435, 673-676.

Niederle, M. and L. Vesterlund (2005). "Do Women Shy Away from Competition? Do Men Compete too Much?" Stanford University, unpublished.