

# Is Behavioral Economics Doomed?

*The ordinary versus the extraordinary*<sup>1</sup>

Max Weber Lecture

June 8, 2009

David K. Levine<sup>2</sup>

---

<sup>1</sup> I am grateful especially to my coauthors Drew Fudenberg and Tom Palfrey with whom I've worked and discussed these issues for many years, to Ramon Marimon, the Max Weber fellows of the EUI, to Karin Tilmans for careful proofreading and to NSF grant SES-03-14713 for financial support.

<sup>2</sup> Department of Economics, Washington University in St. Louis. Email: david@dklevine.com

## Introduction

As Max Weber was a professor of economics, it is perhaps appropriate to discuss modern “behavioral economics” in a lecture in his honor. Indeed – modern economics has returned to many of the issues that fascinated Weber, ranging from political economy to the theory of organizations. Behavioral economics purports to be instrumental in these extensions – my goal in this lecture is to address the question of what – if anything – behavioral economics brings to economics.

Certainly behavioral economics is all the rage these days. The casual reader might have the impression that the rational *homo economicus* has died a sad death and the economics profession has moved on to recognize the true irrationality of humankind. Nothing could be further from the truth. Criticism of *homo economicus* is not a new topic. In 1898 Thorstein Veblen wrote sarcastically rational economic man as “a lightning calculator of pleasures and pains, who oscillates like a homogenous globule of desire of happiness under the impulse of stimuli.” This description had little to do with economics as it was practiced then – and even less now. Indeed, for a long period of time during the 60s and 70s, irrational economic man dominated economics. The much-criticized theory of rational expectations was a reaction to the fact that irrational economic man is a no better description of us than that of a “lightning calculator of pleasures and pains.” In many ways the rational expectations model was a reaction to “[t]he implicit presumption in these ... models ... that people could be fooled over and over again,” as Robert Lucas wrote in 1995.

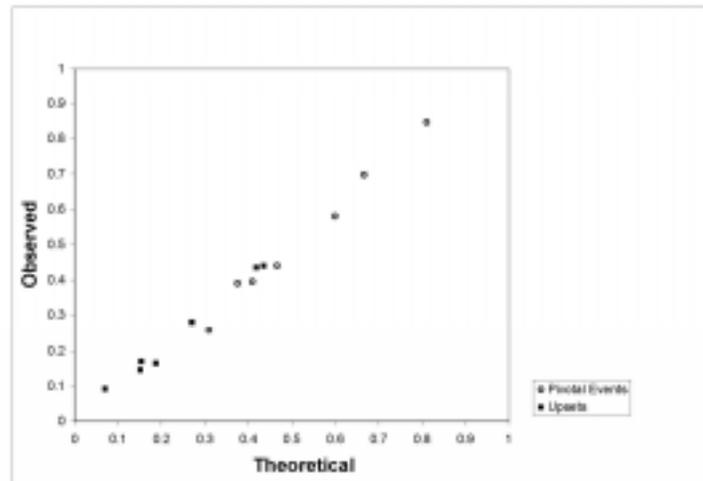
The modern paradigmatic man (or more often these days woman) in modern economics is that of a decision-maker beset on all sides by uncertainty. Our central interest is in how successful we are in coming to grips with that uncertainty. My goal in this lecture is to detail not the theory as it exists in the minds of critics who are unfamiliar with it, but as it exists in the minds of working economists. The theory is far more successful than is widely imagined – but is not without weaknesses that behavioral economics has the potential to remedy.

## Theory That Works: Voting

One of the most widespread empirical tools in modern behavioral economics is the laboratory experiment in which people – many times college undergraduates, but often other groups from diverse ethnic backgrounds – are brought together to interact in artificially created social situations to study how they reach decisions individually or in groups. Many anomalies with theory have been discovered in the laboratory – and rightfully these are given emphasis among practitioners, as we are most interested in strengthening the weaknesses in our theories. However, the basic fact should not be lost that the theory works remarkably well in the laboratory.

The heart of modern “rational” economic theory is the concept of the Nash (or non-cooperative) equilibrium of a game. A game is simply a careful description of a social situation specifying the options available to the “players,” how choices among those options result in outcomes, and how the participants feel about those outcomes. One of the most controversial applications of the theory is to voting. Modern voting theory, for example, the theory of Feddersen and Pesendorfer [1996], is based on the idea that your vote only matters when it decides an election – when it is pivotal. This has implications for voter participation – that elections must be close enough to give voters an incentive for costly participation. Whether this is how voters behave is quite controversial. Levine and Palfrey [2007] examined this question in the laboratory. We divided participants into unequal teams of voters, and each voter was randomly assigned a cost of participating in the election – known only to that voter. Each voter additionally received a prize if their team received the most votes. We then computed, using the theory of Nash equilibrium and the assumption that voters were completely selfish and cared only about their own money income, the unique Nash equilibrium of the game. This is a difficult computation, hinging critically on the fact that the participation rate must be such as to make the pivotal voter indifferent between participating and abstaining. Indeed, we were able to solve the problem only numerically.

We then re-created the theoretical environment in the laboratory. The key aspect is that we had no expectation that voters could guess, calculate, or otherwise intuitively figure out how best to behave. Rather, as is central to modern economic theory (see the quote of Lucas above) we imagined that given an opportunity to learn they would reach an equilibrium. So we gave them ample opportunity to learn – voters got to participate in



fifty elections each. The key measure of how well the theory worked is to ask how the empirical frequency of pivotal events and upset elections compared to the prediction of the theory. The figure above, from Levine and Palfrey [2007], plots the theoretical predictions on the horizontal axis and the empirical frequencies on the vertical axis. It should be emphasized that there are no free parameters – the theory is not fit to the data, rather a direct computation is made from the parameters of the experiments. If the theory worked perfectly the observations would align perfectly on the 45 degree line. As can be seen, they do.

This example of theory that works is but one of many. Other examples are double oral auctions [Plott and Smith, 1978], and more broadly competitive environments [Roth et al, 1991], as well as games such as best-shot [Prasnikar and Roth, 1992].

### Theory That Works? Ultimatum Bargaining

Despite the fact that the theory works extremely well for a variety of games, there are some famous “failures.” One of the most famous is the ultimatum bargaining game. Here one player proposes the division of an amount of money – often \$10, and usually in increments of 5 cents – and the second player may accept, in which case the money is divided as agreed on, or reject, in which case neither player gets anything. This game is frequently analyzed using a “refinement” of Nash equilibrium that requires that a Nash

equilibrium must occur whatever the history of past play. In particular, in ultimatum bargaining, if the second player is selfish, he must accept any offer that gives him more than zero. Given this, the first player should ask for – and get – at least \$9.95.

Not surprisingly this prediction – that the first player asks for and gets \$9.95 – is strongly rejected in the laboratory. The table below shows the experimental results of Roth, Prasnikar, Okuno-Fujiwara and Zamir [1991]. The amount  $x$  represents the part of the \$10 offered to the second player. (The data is rounded off to the nearest 25 cents.)

$x$	Offers	Rejection Probability
\$2.00	1	100%
\$3.25	2	50%
\$4.00	7	14%
\$4.25	1	0%
\$4.50	2	100%
\$4.75	1	0%
\$5.00	13	0%
	27	

US \$10.00 stake games, round 10

The number of offers of each type is recorded in the second column, and the fraction of second players who reject is in the third column. Notice that the results cannot easily be attributed to confusion or inexperience, as players have already engaged in 9 matches with other players. It is far from the case that the first player asks for and gets \$9.95. Most ask for and get \$5.00, and the few that ask for more than \$6.00 are likely to have their offer rejected.

The failure of the theory here is more apparent than real. First, the theory does not demand that players be selfish, although that may be a convenient approximation in certain circumstances, such as competitive markets. It is clear from the data that they are not: a selfish player would never reject a positive offer, yet ungenerous offers are clearly likely to be rejected. Technically this form of social preference is called spite: the willingness to accept a loss in order to deprive the opponent of a gain. Once we take

account of the spite of the second player, the unwillingness of the first player to make large demands becomes understandable.

Let us look more closely at what theory really tells us about this game. Any theory is an idealization. The preferences – in this case selfish preferences – we write down are at best an approximation to players’ “true” preferences. Theorists incorporate this idea through Radner’s [1990] concept of approximate or  $\varepsilon$ -equilibrium. Suppose that  $s_i$  is a strategy choice by player  $i$ , that  $\mu_i$  are his beliefs about the play of his opponents, and that  $u_i(s_i | \mu_i)$  is a numerical “utility” or “payoff” that player  $i$  expects to receive given his own strategy and beliefs. The condition for  $\varepsilon$  equilibrium is that each player should choose a strategy  $s_i$  that loses no more than  $\varepsilon$

$$u_i(s_i | \mu_i) + \varepsilon \geq u_i(s'_i | \mu_i).$$

and that his beliefs  $\mu_i$  should be correct. If  $\varepsilon = 0$  this is the definition of Nash equilibrium. Why allow for  $\varepsilon > 0$ ? Simply put,  $\varepsilon$  is our measure of how much the “true” preferences of the player differ from the preferences  $u_i$  that we have written down. So we allow the possibility that the true “payoff” to player  $i$  from playing  $s_i$  might be somewhat larger than we have written down, but by no more than  $\varepsilon$ . In effect  $\varepsilon$  is a measure of the approximation we think we made when we wrote down a formal mathematical model of player play, or of the uncertainty we have about the accuracy of that model.

A measure of the accuracy of our model then is not given by whether play “looks like an equilibrium” but rather by whether  $\varepsilon$  is small. Take the case of ultimatum bargaining. We can easily compute the losses to players playing less than a best-response to their opponent as averaging \$0.99 per game out of the \$10.00 at stake. What is especially striking is that most of the money is not lost by second players to whom we have falsely imputed selfish preferences, but rather by first movers who incorrectly calculate the chances of having their offers rejected. Notice, however, that a first player who offers a 50-50 split may not realize that he could ask for and get a little bit more without being rejected, nor if he continues to offer a 50-50 split, will he learn of his mistake.

In mainstream modern economic theory, a great deal of attention is paid to how players learn their way to “equilibrium” and what kind of equilibrium might result. It has long been recognized that players often have little incentive to experiment with

alternative courses of action, and may as a result, get stuck doing less well than they would if they had more information. The concept of *self-confirming equilibrium* captures this idea. It requires that beliefs be correct about the things that players actually see – the consequences of the offer they actually make – but not that they have correct beliefs about things they do not see – the consequences of offers that they do not make. Using this concept we can distinguish between knowing losses, representing losses a player might reasonably know about, and unknowing losses due to imperfect learning. In ultimatum bargaining, of the \$0.99 per game that players are losing, \$0.34 are knowing losses due to second players rejecting offers, and the remaining \$0.63 are due to incomplete learning by the first mover. The details of these calculations can be found in Fudenberg and Levine [1997].

One message here is that between social preferences – a major focus of behavioral economics – and learning – a major focus of mainstream economics – in this experiment the role of learning is relatively more important than social preferences. The second message is that the failure of the theory is much less than a superficial inspection suggests. Simply comparing the prediction of subgame perfection to the data indicates an abysmal failure of the theory. Yet a reasonable measure of the success of the theory is that players lose only \$0.34 out of the possible \$10.00 that they can earn.

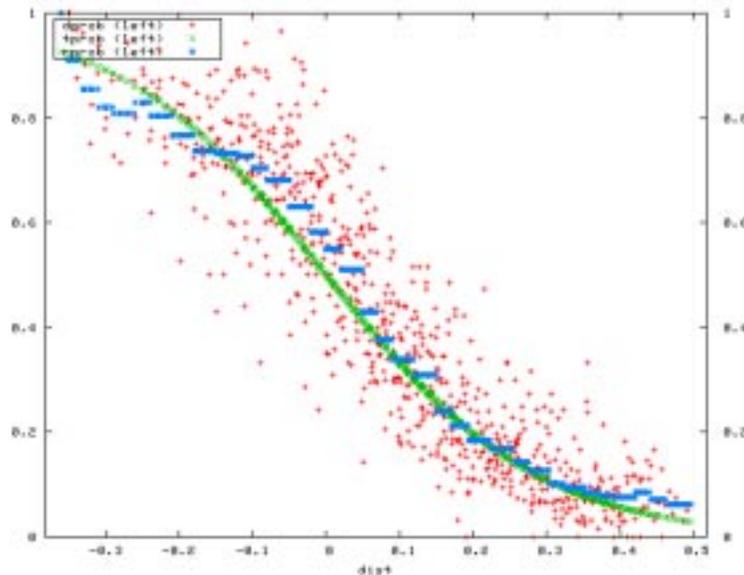
### Equilibrium: The Weak versus the Strong

The key problem with  $\varepsilon$ -equilibrium is not that it makes inaccurate predictions, but rather than it can be a weak theory, often making far too broad a range of predictions. The ultimatum bargaining game is a perfect example: with  $\varepsilon = \$0.99$  the observed behavior is as much an equilibrium as is all the first players demanding \$9.95 and getting it. While weakness is not a good thing in a theory, it is important to recognize that the theory itself tells us when it is weak and when it is strong. When there is a narrow range of predictions – as in the voting game, or in games such as Best Shot or competitive bidding – the theory is useful and correct. When there is a broad range of predictions such as in ultimatum bargaining the theory is correct, but not as useful.

The role for behavioral economics – if there is to be one – is not to overturn existing theory, but it must be instead to strengthen it. That is, psychological factors are

weak compared to economic factors, but in certain types of games that may make a great deal of difference.

To get a sense of the limitations of existing theory, it is useful to take a look under the hood of the voting game described above. At the aggregate level the model predicts with a high degree of accuracy. However, as anyone who has ever looked at raw experimental data can verify, individual play is very noisy and poorly described by the theory. The figure below from Palfrey and Levine [2007] summarizes the play of



individuals. The optimal play for an individual depends on the probability of being pivotal (deciding the election) and on the cost of participation. The horizontal axis measures the loss from participating depending on the cost that is drawn. If – in the given election – the cost drawn should make the player indifferent to participating, the loss is zero. Otherwise it can be negative or positive, depending on how much is lost from participating. The vertical axis is the empirical probability of participating. The red dots are the results of individual elections. The blue dots are averages of the red dots for each loss level, and the green curve is a theoretical construct described below. The theory says that this “best response” function should be flat with the probability of participating equal to one until gains (negative losses) reach zero on the horizontal axis, then a vertical line, then again flat with a value of zero for all losses that are bigger than zero. This is far from the case: some players make positive errors, some make negative errors. The key is that in this voting game, the errors tend to offset each other. Over voting by one voter causes other voters to want to under vote, so aggregate behavior is not much effected by the fact

that individuals are not behaving exactly as the theory predicts. A similar statement can be made about a competitive auction and other games in which equilibrium is strong and robust. By way of contrast, in ultimatum bargaining, a few players rejecting bad offers changes the incentives of those making offers, so that they will wish to make lower offers – moving away from the subgame perfect equilibrium, not towards it.

A key feature of the individual level data is that behavior is sensitive to the cost of “mistakes.” That is, voters are more likely to play “sub-optimally” if the cost of doing so is low. The same is true in ultimatum bargaining: bad offers are less costly to reject than good ones, and are of course rejected more frequently.

This fact: the weakness of incentives when players are near indifferent, can be captured without any “psychological” analysis quite effectively through what has become known as quantal response equilibrium, or QRE. This logistic choice model that has been used by economists since McFadden’s [1980] work has become popular due to the work of McKelvey and Palfrey [1995] in analyzing experimental data. It supposes that play is somewhat random. Suppose that  $\sigma_i(s_i)$  is the probability with which player  $i$  plays the strategy  $s_i$ . Let  $\lambda_i > 0$  be a parameter of the choice function. We first define propensities with which strategies are played,  $p_i(s_i) = \exp(\lambda_i u_i(s_i, \sigma_{-i}))$ . This says that strategies that yield higher utilities have higher propensities of being played. The QRE equilibrium probabilities are given by normalizing the propensities to add up to one.

$$\sigma_i(s_i) = p_i(s_i) / \sum_{s_i} p_i(s_i).$$

Notice that this formulation contains an unknown preference parameter  $\lambda_i$ . If  $\lambda_i = 0$  play is completely random. As  $\lambda_i$  becomes large, the probability of player the “best” response approaches one. So we can interpret  $\lambda_i$  as a kind of index of rationality. To give an idea how this theory works, in the voting experiment we can estimate a common value of  $\lambda_i$  for all players. The corresponding equilibrium probabilities of play are given by the green curve in the figure above, which does an excellent job of describing individual play – although it makes roughly the same predictions for aggregate play as Nash equilibrium.

While QRE is useful in explaining a great many experimental deviations from Nash equilibrium in games where Nash equilibrium is weak, it captures only the cost side of preferences. That is, it recognizes – correctly – that departures from standard “fully rational” selfish play are more likely if they are less costly in objective terms, but it does

not attempt to capture the benefits of playing non-selfishly. It cannot capture, for example, the fact that under some circumstances players are altruistic, and in other spiteful. The modern literature on social preferences and fairness including Rabin [1993], Levine [1998], Fehr and Schmidt [1999], Bolton and Ockenfels [2000], and Gul and Pesendorfer [2004] attempt to capture that idea. On the other hand we already observed that incomplete learning is a more important source of deviations from the “pure” theory than are social preferences. The QRE does a good job of capturing errors that arise from incomplete learning – indeed, it is implied by learning models such as the smooth fictitious play of Fudenberg and Levine [1995].

### Learning and Self-confirming Equilibrium

Learning and incomplete learning – whether or not we regard this as “behavioral” economics – are an important part of mainstream economics and have been for quite some time. An important aspect of learning is the distinction between active learning and passive learning. We learn passively by observing the consequences of what we do simply by being there. However we cannot learn the consequences of things we do not do, so unless we actively experiment by trying different things, we may remain in ignorance.

As I indicated, the notion of *self-confirming equilibrium* from Fudenberg and Levine [1993] captures this idea. A simple example adapted from Sargent, Williams and Zhao [2006a] by Fudenberg and Levine [2009] shows how this plays a role in mainstream economic thought. Consider a simple economic game between a government and a typical or representative consumer. First, the government chooses high or low inflation. Then in the next stage consumers choose high or low unemployment. Consumer always prefer low unemployment, while the government (say) gets 2 for low unemployment plus a bonus of 1 if inflation is low. If we apply “full” rationality (subgame perfection), we may reason that the consumer will always choose low unemployment. The government recognizing this will always choose low inflation. Suppose, however, that the government believes incorrectly that low inflation leads to high unemployment – a belief that was widespread at one time. Then they will keep inflation high – and by doing so never learn that their beliefs about low inflation are false.

This is what is called a *self-confirming* equilibrium. Beliefs are correct about those things that are observed – high inflation – but not those that are not observed – low inflation.

This simple example cannot possibly do justice to the long history of inflation – for example in the United States. Some information about the consequences of low inflation is generated if only because inflation is accidentally low at times. Sargent, Williams and Zhao use a sophisticated dynamic model of learning about inflation to understand how in the U.S. Federal Reserve policy evolved post World War II to ultimately result in the conquest of U.S. inflation.

While the current economic crisis is surprising and new to non-economists, it is much less so to economists who have observed and studied similar episodes throughout the world. Here too learning seems to play an important role. Sargent, Williams, Zha [2006b] examine a series of crises in Latin America from a learning theoretic point of view. They assume that consumers have short-run beliefs that are correct, but have difficulty correctly anticipating long run events (the collapse of a “bubble”). Periodic crises arise as growth that is unsustainable in the long-run takes place, but consumers cannot correctly foresee that far into the future.

In talking about the crisis, there is a widespread belief that bankers and economists “got it wrong.” Economists anticipate events of this sort, but by their nature the timing is unpredictable. Bankers by way of contrast can hardly be accused of acting less than rationally. Their objective is not to preserve their banks or take care of their customers – it is to line their own pockets. They seem to have taken advantage of the crisis to do that very effectively. If you can pay yourself bonuses during the upswing, and have the government cover your losses on the downswing, there is not much reason to worry about the business cycle.

## Behavioral Theories

While behavioral economics points to many paradoxes and problems with mainstream economics, their own models and claims are often not subject to a great deal of scrutiny. Here I examine some popular behavioral theories.

**The naif at the health club:** Consider the following facts from Della Vigna and Malemendier [2006] about health club memberships. First, people who chose long-term memberships rather than pay per visit paid on average \$17 per visit as against a \$10 per

visit fee. Leaving aside the hassle factor of availability of lockers and the need to pay each visit, we can agree that this is some evidence that people are trying to make a commitment to attending the health club.

In the idealized world usually studied by economists, there is no need for a single decision-maker ever to commit. In reality we often choose to make commitments to avoid future behavior we expect to find tempting but with bad long-term consequences: the drug addict who locks himself in a rehab center would be an obvious example. The long-term membership in a health club has a similar flavor. Skipping a workout can be tempting but has bad long-term consequences for health. Having to pay \$10 will make it easier to find excuses to avoid going.

So far so good for behavioral economics. They have identified a phenomenon that standard models cannot explain – the desire for commitment in single-person decision problems. Of course even with the commitment, some people eventually give up and stop going to the health club. However, Della Vigna and Malmendier's data shows that people typically procrastinate for an average of 2.3 months before canceling their self-renewing membership. The average amount lost is nearly \$70 against canceling at the first moment that attendance stops.

Leaving aside that fact that it may take a while to make the final decision to quit the club, we are all familiar with procrastination. Why cancel today when we could cancel tomorrow instead? Or given the monthly nature of the charge, why not wait until next month. One behavioral interpretation of procrastination is that people are *naïve* in the sense that they do not understand that they are procrastinators. That is, they put off until tomorrow, believing they will act tomorrow, and do not understand that tomorrow they will face the same problem and put off again. There may indeed be some people that behave this way. But if we grant that people who put off cancellation are making a mistake, there are several kinds of untrue beliefs they might hold that explains this. One is that they are procrastinators and do not know it. Another is that it is really simple and inexpensive to cancel their membership, but people incorrectly perceive that it will be an time consuming hassle involving endless telephone menus, employees who vanish in back-rooms for long periods of times, and all the other things we are familiar with whenever we try to cancel a credit card charge.

The question to raise about the “naïve” interpretation then is this. Which is more likely: that people are misinformed about something they have observed every day for their entire lives (whether or not they are procrastinators) or something that they have observed infrequently and for which the data indicates costs may be high (canceling)? Learning theory suggests the latter – people are more likely to make mistakes about things they know little about.

Another point worth mentioning is that so called “impulsive” behavior – that is, giving in to temptation – is often everything but. Take Eliot Spitzer who lost his job as governor of New York because of his “impulsive” behavior in visiting prostitutes. Yet the fact is that he paid months in advance (committing himself to seeing prostitutes rather than the other way around) and in one case flew a prostitute from Washington D.C. to New York – managing to violate Federal as well as State law in the process. Similarly, when Rush Limbaugh was discovered to be carrying large quantities of viagra from the Dominican Republic it was widely suspected that he had gone there on a “sex vacation” – hardly something done impulsively at the last minute. Or perhaps a case more familiar to most of us – how about the Las Vegas vacation? This also is planned well in advance, with the anticipation of the rush of engaging in impulsive behavior. Of course, the more sensible among us may plan to limit the amount of cash we bring along.

The point here is simple: our “rational” self is not intrinsically in conflict with our impulsive self. In fact the evidence is that our rational self often facilitates rather than overrides the activities of our impulsive self.

**Prospect theory to the rescue:** Psychologists widely regard the decision theoretic model used by economists – expected utility theory – as nuts. The reasons for this are subtle – although there is a real problem we discuss in the conclusion – but psychologists also have a serious alternative called prospect theory. This has two parts: one is that gains and losses are measured relative to a reference point. Unfortunately for economists the reference point seems to vary from setting to setting in a not entirely explained manner. Let us focus on the second part of prospect theory – the part that says that people overweight low probabilities and underweight high probabilities. Bruhin, Fehr-Duda, and Epper [2007] (economists, by the way) carry out a careful experimental study to find what the probability weighting function might be. Suppose that  $p_i$  is the

chance of winning one of two prizes  $x_i \geq 0$ . They find that for gains, many people choose gambles as if they maximized the utility function

$$U = \sum_i \frac{.846p_i^{.414}}{.846p_i^{.414} + (1 - p_i)^{.414}} x_i^{1.056} .$$

One issue with theories, however, is that they make a range of predictions – not only in the laboratory, but also outside the laboratory. Which would you rather have:

A. \$5,000 for sure (or)

B. a 50-50 coin-flip between \$9,700 dollars and nothing

Most people I imagine would prefer A. However an individual with the “typical” utility function above will choose B. So prospect theory is not without its own paradoxes.

To pursue this further, prospect theory is motivated in part by an important decision theoretic puzzle called the Allais paradox. Consider the following two scenarios: In Scenario 1 you choose between a certain \$1 million and a lottery offering a nothing with a 1% probability, \$1 million with an 89% probability, and \$5 million with a 10% probability. Most people choose the certain \$1 million. In Scenario 2 you are offered the choice between two lotteries. The first lottery offers nothing with 89% probability and \$1 million with 11% probability, while the second offers nothing with 90% probability and \$5 million with 10% probability. Here most people choose the 10% chance of \$5 million. However no expected utility maximizer can make these choices: if \$1 million for sure was chosen in Scenario 1 then any expected utility maximizer must choose the 11% chance of \$1 million in Scenario 2. Prospect theory offers a possible resolution of this paradox because smaller probabilities are exaggerated, making the first choice relatively unattractive in Scenario 1, but not so much so in Scenario 2. Unfortunately the Bruhin, Fehr-Duda, and Epper [2007] utility function above does not lead to the Allais reversal. While explaining the laboratory data, it fails to explain the Allais paradox.

**Becker, Marschak and DeGroot:** Returning to the theme of which types of mistakes are most likely, another paradox of behavioral economics is the so called willingness to pay versus willingness to accept. For example, if we ask people how much they are willing to pay for a coffee cup they will state a relatively low value; if we give them a coffee cup and ask how much they will sell it for they will state a relatively high value. On the surface this is not a paradox: we all know to buy low and sell high. However: the elicitation of values is done using a method called the Becker Marschak

DeGroot elicitation procedure. A willingness to pay or accept payment is stated, then a random draw is made. If the random draw is lower than the stated value (in the willingness to pay case) then the item is sold at the randomly drawn price. If the draw is higher than the stated value then no transaction takes place.

Is it obvious to you that when this procedure is used that the unambiguously best course of action is to bid your true value and not buy low and sell high? It is true, and subjects are often informed of this fact. So: is there a paradox here, as some behavioral economists and psychologists would argue, or is it simply the case that people have trouble understanding a complex and unfamiliar procedure? The answer is the latter: Zeiler and Plott [2004] show that if subjects are well trained in understanding the elicitation procedure – so they clearly understand that the best thing to do is to state their true value – then there is no difference between willingness to pay and willingness to accept payment. If the observation that people have trouble understanding complex decisions and sometimes make mistakes is “behavioral” then we scarcely need experimental evidence to prove the point – the fact that students get exam questions wrong should be proof enough that people fall short of complete and total rationality.

### Psychology versus Economics

Much of behavioral economics arises from the fact that people have an emotional irrational side that is not well-captured by mainstream economic models. By way of contrast, psychologists have long been fascinated with this side of humankind, and have many models and ideas on the subject. Not surprisingly much of behavioral economics attempts to import the ideas and models developed by psychologists. Those who know me know that I have a series of “why psychologists are dumber than pigeons” jokes. Since the psychologists I know are all smarter than me, you can see where that leaves me. More to the point – it is important to recognize that the goals of psychologists and economists are very different, and that this has implications for importing ideas from psychology into economics.

The key difference between psychologists and economists is that psychologists are interested in individual behavior while economists are interested in explaining the results of groups of people interacting. Psychologists also are focused on human dysfunction – much of the goal of psychology (the bulk of psychologists are in clinical

practices) is to help people become more functional. In fact, most people are quite functional most of the time. Hence the focus of economists on people who are “rational.” Certain kinds of events – panics, for example – that are of interest to economist no doubt will benefit from understanding human dysfunctionality. But the balancing of portfolios by mutual fund managers, for example, is not such an obvious candidate. Indeed one of the themes of this essay is that in the experimental lab the simplest model of human behavior – selfish rationality with imperfect learning – does an outstanding job of explaining the bulk of behavior.

The difference between group and individual behavior is quite crucial in other ways as well. There is a small segment of the psychology literature that effectively commits a fallacy of composition, reasoning that if we can explain individual behavior, this carries over immediately to the group. The most obvious example of this is the idea that if we could somehow make people better – more altruistic, say – then society at large would be better off. This is far from the case – a nice example of an interactive setting where better people result in an inferior society can be found in Hwang and Bowles [2008]. In a more intuitive way – from the point of view of the clinical practitioner, making a Mafia don more functional might be a good thing. Indeed medical ethics is entirely focused on the patient, with no allowance for the role of the patient in society. Of course making a Mafia don more functional can be extremely bad for society more broadly. The bottom line is that what is good for the individual is not always good for society, and we need and use game-theoretic and related models in order to understand the consequences of individual behavior for the group.

The need to study groups of potentially large numbers of individuals imposes constraints on economic models of individual decision making that are not present for psychologists. Economists need simple broad models of behavior. Narrow complex models of behavior cannot easily be used to study the behavior of many people interacting. Hence the focus by economists on axiomatic models that provide some assurance that the model explains not only particular data, but gives good results over a broad range of social settings. To take a particular example, research in psychology on hyperbolic discounting focuses on finding clever functional forms that will fit a broad range of data on human (and animal) behavior involving delayed rewards. From an

economist's perspective, such models can be useful in testing and calibrating our own models – but they cannot be usefully embedded in complex social situations.

Finally, there is the question of decision making and the physiology of decision making – for example as mapped out by peering into the brain with fMRI machines. The key thing to understand is that for the kinds of decisions that economists are interested in much of the action does not take place in the brain, nor is it subject to the memory and other limitations of the human brain. Even before we all had personal computers, we had pieces of paper that could be used not only for keeping track of information – but for making calculations as well. For most decisions of interest to economists these external helpers play a critical role – and no doubt lead to a higher level of rationality in decision making than if we had to make all decisions on the fly in our heads.

A useful summing up is by considering the main theme of this lecture: that behavioral economics can contribute to strengthening existing economic theory, but, at least in its current incarnation, offers no realistic prospect of replacing it. Certain types of “behavioral” models are already important in mainstream economics: these include models of learning; of habit formation; and of the related phenomenon of consumer lock-in. Behavioral criticisms that ignore the great increase in the scope and accuracy of mainstream theory brought about by these innovations miss the mark entirely. In the other direction are what I would describe as not part of mainstream economics, but rather works in progress that may one day become part of mainstream economics. The ideas of ambiguity aversion, and the related instrumental notion that some of the people we interact with may be dishonest is relatively new and still controversial. The use of models of level-k thinking to explain one-time play in situations where players have little experience works well in the laboratory, but is still unproven as a method of analyzing important economic problems. The theory of menu choice and self-control likewise has still not been proven widely useful. The theory of interpersonal (or social) preferences is no doubt needed to explain many things – but so far no persuasive and generally useful model has emerged.

### The Rabin Paradox

So far we have focused on those puzzles and paradoxes that have captured the imagination of economists and psychologists. The fact is that these are all of somewhat

secondary importance to economists. There is, however, one dramatic flaw in the existing theory of expected utility that might make a good ending point for this essay. This flaw is by no means of second order importance, yet has received relatively little attention from behavioral economists – or for that matter psychologists. This is the Rabin [2000] paradox.

I will give my own version of the Rabin paradox – drawn from years of watching experimental papers presented in which attitudes towards risk are measured – without once mentioning that the results are nonsense by three orders of magnitude. One measure economists use of risk aversion is the so-called “coefficient of relative risk aversion.” The bigger this number, the more you are willing to pay for a certain alternative to a risky prospect. Suppose that your lifetime wealth is \$860,000 which is about the median in the United States. Suppose also that you are indifferent between a 70% - 30% chance of A: \$40 and \$32 and B: \$77 and \$2 – which many people are in the laboratory [Holt and Laury 2002]. Then your coefficient of relative risk aversion is 27,950. If this sounds like a big number it is. One important puzzle much studied by economists is why the rate of return on stocks is so much higher than on bonds, given that stocks are not all that much more riskier. In fact we can also calculate how risk averse a person would be who is on the margin between buying stocks (S&P 500 index mutual fund) versus U.S. government bonds – a situation many of us are in. The answer is that the corresponding coefficient of relative risk aversion is 8.84. This is over three orders of magnitude different than the answer we find in the laboratory.

To give credit to psychologists – their main model of attitudes towards risk is prospect theory, which allows attitudes towards risk to depend on the context (or “reference point”). There is no way to explain the wildly different attitudes towards large and small risks without some model of context dependence. Here is surely an area where rapid and immediate progress is needed.

## References

- Bolton, GE and A Ockenfels [2000]: “ERC: A Theory of Equity, Reciprocity, and Competition,” *American Economic Review*.
- Bruhin, A., H. Fehr-Duda, and T. Epper [2007] “Risk and Rationality: Uncovering Heterogeneity in Probability Distortion,” mimeo.
- Della Vigna, S. and U Malmendier [2006]: “Paying not to go to the gym,” *American Economic Review*.
- Feddersen, TJ and W Pesendorfer [1996], “The swing voter's curse,” *American Economic Review*.
- Fehr, E and KM Schmidt [1999]: “A Theory Of Fairness, Competition, and Cooperation,” *Quarterly Journal of Economics*.
- Fudenberg, D and DK Levine [1993]: “Self-Confirming Equilibrium,” *Econometrica*.
- Fudenberg, D and DK Levine [1995]: “Consistency and cautious fictitious play,” *Journal of Economic Dynamics and Control*.
- Fudenberg, D and DK Levine [1997]: “Measuring Players' Losses in Experimental Games,” *Quarterly Journal of Economics*.
- Fudenberg, D and DK Levine [2009], “Self-confirming Equilibrium and the Lucas Critique,” forthcoming *Journal of Economic Theory*.
- Gul, F and W Pesendorfer [2004]: “The Canonical Type Space for Interdependent Preferences”.
- Holt, C and S Laury [2002]: “Risk Aversion and Incentive Effects,” *American Economic Review*.
- Hwang, SH and S Bowles [2008]: “Is Altruism Bad For Cooperation?” mimeo
- Levine, DK [1998]: “Modeling Altruism and Spitefulness in Experiments,” *Review of Economic Dynamics*.
- Levine, DK and TR Palfrey [2007]: “The Paradox of Voter Participation: A Laboratory Study,” *American Political Science Review*, 101: 143-158.
- Lucas, R [1995] quoted in “Economics dynasty continues: Robert Lucas wins Nobel Prize,” Chicago Journal, The University of Chicago Magazine, December.
- McFadden, D [1980]: “Econometric models for probabilistic choice among products,” *Journal of Business*.

- McKelvey, RD and TR Palfrey [1995]: “Quantal Response Equilibria for Normal Form Games,” *Games and Economic Behavior*.
- Plott, CR and VL Smith [1978]: “An Experimental Examination of Two Exchange Institutions,” *Review of Economic Studies*.
- Plott, CR and K Zeiler [2005]: “The Willingness to Pay-Willingness to Accept Gap, the Endowment Effect and Subject Misconceptions,” *American Economic Review*.
- Prasnikar, V and AE Roth [1992]: “Considerations of fairness and strategy: Experimental data from sequential games,” *The Quarterly Journal of Economics*.
- Rabin, M. [1993]: “Incorporating fairness into game theory and economics,” *The American Economic Review*.
- Rabin, M [2000]: “Risk Aversion and Expected-utility Theory: A Calibration Theorem,” *Econometrica*.
- Radner, R [1990]: “Collusive Behavior in Noncooperative Epsilon-Equilibria of Oligopolies with Long but Finite Lives,” *Journal of Economic Theory*.
- Roth, AE, V Prasnikar, M Okuno-Fujiwara, and S Zamir [1991]: “Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study,” *American Economic Review*
- Sargent, T., N. Williams, and T. Zhao [2006a], “Shocks and Government Beliefs: The Rise and Fall of American Inflation,” *American Economic Review*, **96**: 1193-1224.
- Sargent, T., N. Williams, and T. Zhao [2006b], “The Conquest of South American Inflation,” NYU.
- Veblen, T [1898]: “Why is Economics not an Evolutionary Science?” *Quarterly Journal of Economics*.