

Neuroeconomics?¹

March 3, 2011

David K. Levine²

¹ NSF grants SES-03-14713 and SES-08-51315 provided financial support. I am grateful to Colin Camerer, Guillaume Frechette, Drew Fudenberg, Glenn Harrison, Rosemarie Nagel, Tom Palfrey, and Jie Zheng for many discussions.

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

1. Introduction

The point of departure for most neuroeconomics is what has come to be called behavioral economics. A reader of this literature might rightfully conclude that the old rational man of economics has been rejected in favor of more realistic models of human behavior that incorporate insights from psychology. The old model supposes hyperrationality of individuals and as a consequence existing economic theory is useless. We can complete the behavioral revolution in economic thought by peering into the brain in order to build far more models of decision-making.

Unfortunately this assessment of the nature and state of modern economics is false. In this paper I will document the basic strength of existing theory. It is only by understanding its true weaknesses – which revolve around preferences far more than decision making – that can we understand what if any role there can be for neuroeconomics.

2. Theory That Works

The critical thing to understand is that modern economic theory works well for the problems that economists are the most interested in – both in the field and in the laboratory. Since shortcomings of the theory naturally get more emphasis among practitioners this basic fact often is inadequately understood.

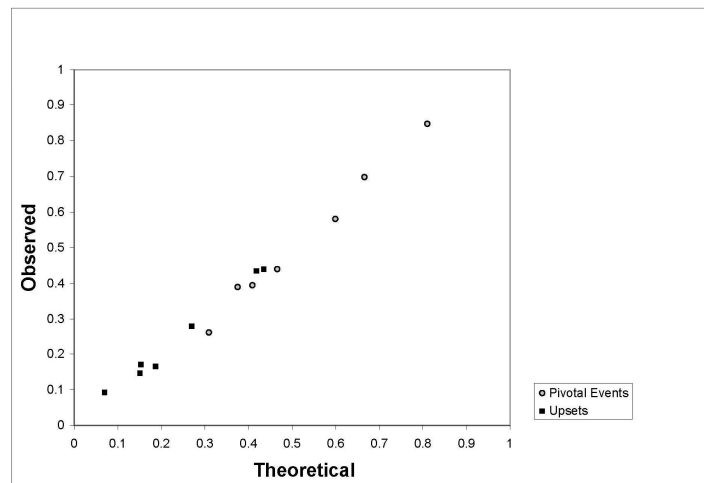
Let us start in the laboratory. Here people – many times college undergraduates, but often other groups from diverse ethnic backgrounds – are brought together to interact in artificially created social situations to study how they reach decisions individually or in groups. The heart of modern “rational” economic theory is the concept of the Nash 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. Nash equilibrium says that players do the best they can given the actual choices of their opponents.

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 voter participation 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.

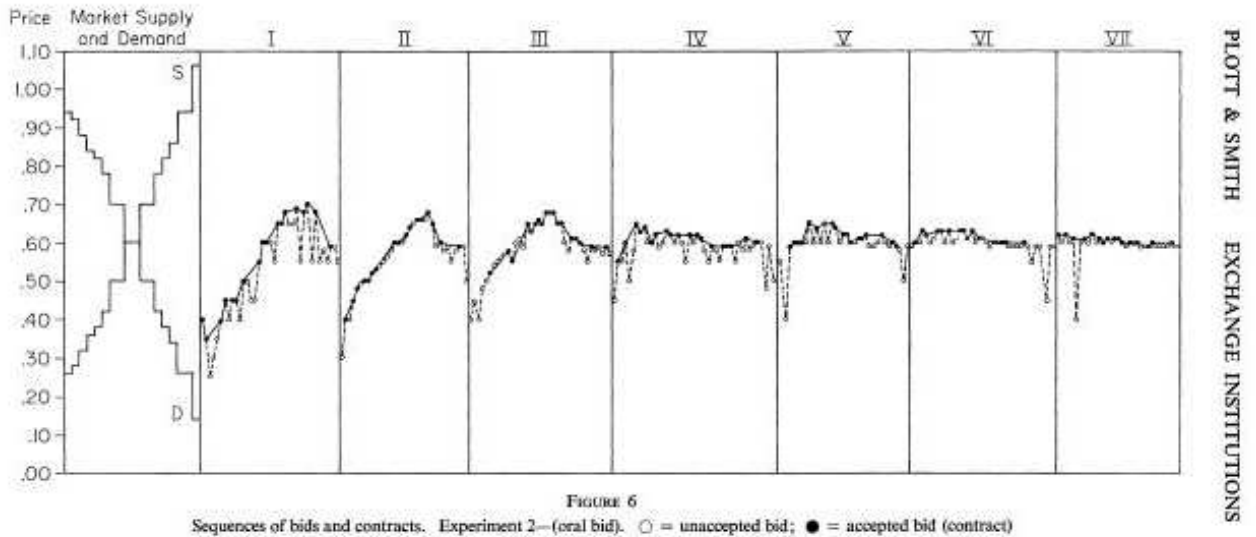
We then re-created the theoretical environment in the laboratory. The key aspect is that we had no expectation that voters could guess, calculate, or otherwise intuitively figure out how best to behave. Rather, as is central to modern economic theory we imagined that given an opportunity to learn they 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 voting experiment is but a recent example a long tradition of experimental economics – the most important finding of which is that market clearing – the heart of



most economics – happens quickly and easily in the laboratory. The graph above is taken from Plott and Smith [1978] and shows how a sequence of bids in an experimental double oral auction converges quickly to the intersection of the supply and demand curve at \$0.60.

3. Theory That Works? *Ultimatum Bargaining*

Not all economic theory works so well in the laboratory. One of the most difficult areas in economics is the theory of bargaining – about which there is no widely agreed upon theory. One of the most famous “failures” of economic theory in the laboratory takes place in 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 first column shows how much of the \$10 is offered to the second player. (The data is rounded off.) 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.

<i>Amount of Offer</i>	<i>Number of Offers</i>	<i>Rejection Probability</i>
<i>\$3.00 or less</i>	3	66%
<i>\$4.75 to \$4.00</i>	11	27%
<i>\$5.00</i>	13	0%
U.S. \$10.00 stake games, round 10		

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.

Looking at the data a simple hypothesis presents itself: players are not strategic at all they are “behavioral” and fair-minded and just like to split the \$10.00 equally. Aside from the fact that this “theory” ignores slightly more than half the observations in which the two players do not split 50-50, it might be wise to understand whether the “economic theory” of rational strategic play has really failed here – and if so how.

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 ε -equilibrium. Suppose that s_i is a strategy choice by player i , that μ_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 ε equilibrium is that each player should choose a strategy s_i that loses no more than ε

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

and that his beliefs μ_i should be correct. If $\varepsilon = 0$ this is the definition of Nash equilibrium. Why allow for $\varepsilon > 0$? Simply put, ε 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 ε . In effect ε 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 ε 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].

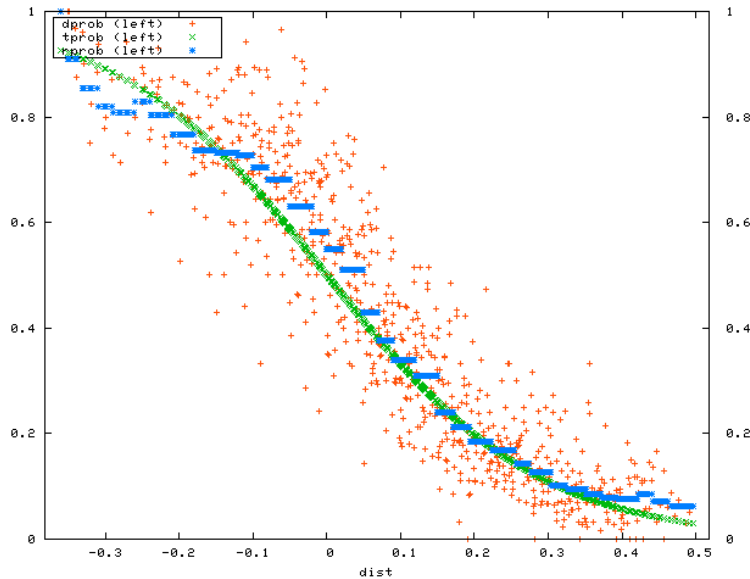
One message here is that between social preferences – a major focus of behavioral economics – and learning – a major focus of mainstream economics – in this experiment the role of learning is relatively more important than social preferences. The second message is that the failure of the theory is much 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.

4. Equilibrium: The Weak versus the Strong

The key problem with ε -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 competitive market games – 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.

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 affected by the fact that individuals are not behaving exactly as the theory predicts. A similar statement can be made about a competitive market games. 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 $\lambda > 0$ be a parameter of the choice function. We first define propensities with which strategies are played, $p_i(s_i) = \exp(\lambda u_i(s_i, \sigma_{-i}))$. 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 λ . If $\lambda = 0$ play is completely random. As λ becomes large, the probability of playing the “best” response approaches one. So we can interpret λ 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 λ 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 others 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].

5. Selling a Jar of Pennies

One (incorrect) interpretation of economic theory is that nobody can every make a profit. In fact there is a surefire way to do so. Put a bunch of pennies in a jar bring it to class. Then auction off the jar of pennies. You will find you can sell a \$3.00 jar of pennies for almost \$10.00.

This illustrates an important phenomenon known as the winner's curse. Your class stares at the jar and tries to guess how many pennies there are. Some students under guess – they may guess that there are only 100 or 200 pennies. They bid low. Others over guess – they may guess that there are 1,000 pennies or more. They bid high. Of course those who overestimate the number of pennies by the most bid the highest – so you make out like a bandit.

According to Nash equilibrium this should not happen. Everyone should rationally realize that they will only win if they guess high, so they should bid less than their estimate of how many pennies there are in the jar. They should bid a lot less – every player can guarantee they lose nothing by bidding nothing. So in equilibrium, they can't on average lose anything, let alone \$7.00.

QRE – by recognizing that there is a small probability that people aren't so rational – makes quite a different prediction. Let U denote the most possible profit from getting a large number of pennies at zero cost. Let u denote the least profit by getting a jar with no pennies at the highest possible bid. QRE says that the propensity of every action is at least $p = \exp(\lambda u)$ and that no action has a propensity greater than $P = \exp(\lambda U)$. Hence each person independently plays every action with probability at least p/MP where M is the number of possible bids. What happens as the number of bidders grows? Each bidder according to QRE has at least a p probability of making the highest possible bid. With many bidders it becomes a virtual certainty that one of the bidders will (unluckily for them) make this high bid, so with enough bidders, QRE assures the seller a nice profit.

6. Matching Pennies

Matching pennies is the classical zero sum game where each player chooses Heads or Tails and one wins if they match, the other if they do not. Goeree and Holt [2001] studied several variations of Matching Pennies in the laboratory. In the first variation the payoffs were 80 for the winner and 40 for the loser. The only Nash equilibrium is for players to randomize fifty-fifty and they did just that. The table below shows the theoretical Nash equilibrium of 50% and in parentheses the actual fraction of subjects that chose the corresponding row and column. As you can see it is quite close to 50%.

	50% (48%)	50% (52%)
50% (48%)	80,40	40,80
50% (52%)	40,80	80,40

Fifty-fifty is a particularly easy strategy to implement and the theory of mixed strategy equilibrium is peculiar in that it predicts that each player must randomize so as to make his opponent indifferent. This implies that in a mixed strategy equilibrium each player's play depends only on his opponents payoffs and not on his own.

To study randomization Goeree and Holt changed the payoffs by increasing (from 80 to 320) or decreasing (from 80 to 44) the payoff to Player 1 in the upper left corner. In theory this should change Player 2's equilibrium play, but Player 1 should continue to randomize fifty-fifty. The two tables below show the theoretical predictions of Nash equilibrium and in parentheses what actually happened: far from continuing to randomize fifty-fifty Player 1 played the row containing the highest payoff at least 92% of the time.

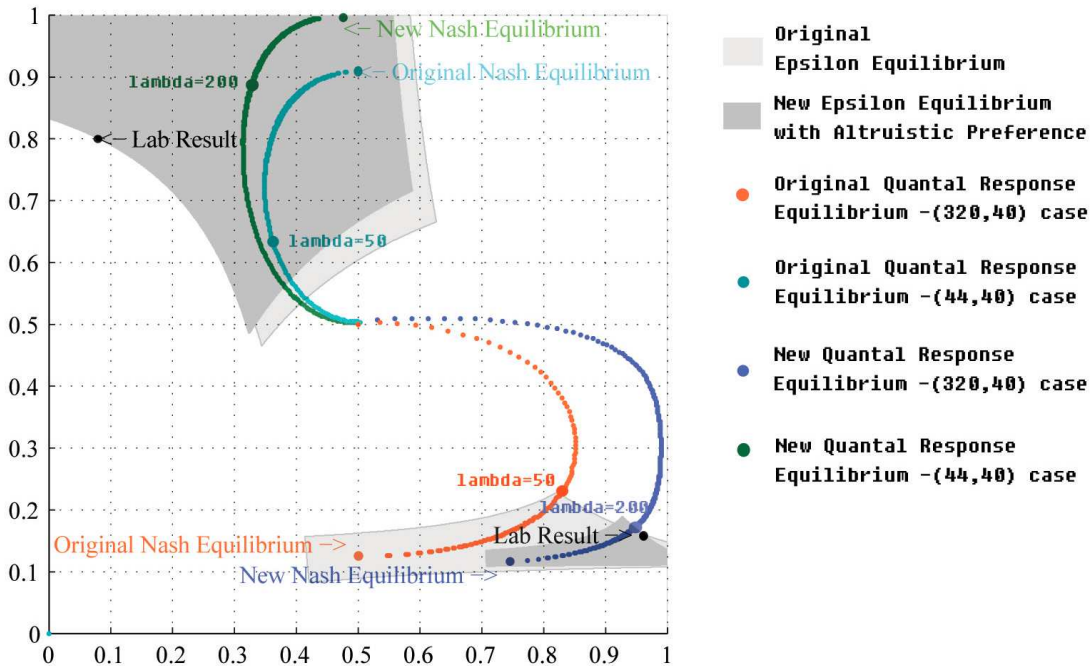
	12.5% (16%)	87.5% (84%)
50% (96%)	320,40	40,80
50% (4%)	40,80	80,40

	87.5% (80%)	12.5% (20%)
50% (8%)	44,40	40,80
50% (92%)	40,80	80,40

The theory here does about as badly as it can: the theory predicts equal probability between the two rows, but the actuality is that one row is played pretty much all the time. Notice however that this experiment involves players who are inexperienced in the sense that they only got to play the game once

The figure below is taken from Levine and Zheng [2010] and analyzes the game using the tools of altruism, QRE and ϵ -equilibrium. The vertical axis is the frequency with which Player 1 chooses the *Top* row; the horizontal axis the frequency with which Player 2 chooses the *Left* column. The laboratory results are shown by the black dots labeled *Lab Result* with the upper left dot corresponding to the second matrix – the 44 game, and the lower right dot corresponding to the first matrix – the 320 game. The theoretical prediction of Nash equilibrium – that Player 1 (and only Player 1) randomizes 50-50 – are labeled as *Original Nash Equilibrium*.

Several different ways of weakening the theory of selfish Nash equilibrium are considered. The first is by computing all the approximate equilibrium in which the losses are no greater than those actually suffered by the participants. This is the light gray



shaded region. The second is by computing the QRE corresponding to differing levels of λ . These are the light blue and red curves that begin at the respective Nash equilibria and – as λ declines – move eventually towards the completely random outcome where both players simply make each choice with equal 50% probability. The dark gray region and the green and dark blue curves also examine approximate and QRE – but do so under the hypothesis that some players are altruistic.

To understand what this diagram does and does not show, it is useful to start with QRE. One prediction of quantal response is a tendency toward the middle. For example in the 320 game Player 2 plays *Left* in Nash equilibrium 12.5% of the time. Quantal response says that errors in play will push that towards the middle – toward a fifty-fifty randomization, and indeed we see that in actuality 16% rather than 12.5% of Player 2’s play *Left*. This in turn has a substantial impact on the incentives of Player 1: with “too many” player 2’s playing *Left*, the best thing for Player 1 to do is to play *Top* and try to get the 320 – and again this is what we see participants do. We see it also in the diagram. As we vary the parameter of noisy choice away from Nash equilibrium and perfect best response we see that QRE play shifts towards to the right – towards the lab result with more Player 1’s playing *Top*. Similarly in the 44 game, “too many” player 2’s play *Right* – 20% rather than 12.5% – and this tilts the Player 1’s towards playing *Down*. Again, the initial effect of increasing the noise parameter is to move the QRE towards the lab result.

Eventually, when λ becomes too small, QRE approaches a pure fifty-fifty randomization. What the diagram also shows is that this happens “too soon” in the sense that play in the QRE “starts back” towards fifty-fifty before it gets to the laboratory result. That effect is much more pronounced in the 44 game than the 320 game.

Next consider altruism. This is potentially important in the 320 game since Player 2 by giving up 40 can increase the payoff of Player 1 by 280 – you don’t have to be that generous to take such an opportunity. This also can explain why “too many” Player 2’s play *Left*. If we assume a combination of errors due to quantal response and some altruistic players, it turns out we can explain the 320 game quite well, as the curve combining the two effects passes more or less directly through the laboratory result.

In the 44 game the situation is different. Even combining altruistic players with quantal response errors quantitatively we can explain only about half the laboratory result. Here the approximate equilibrium regions can help us understand what is going

on. Notice that in the 320 game the approximate equilibrium region while wide is not very tall. While there are many possible strategies by player 1 that are consistent with a relatively small loss, there are very few strategies by player 2: Player 2 must play *Right* with between about 10% and 20% probability. On the other hand, in the 44 game approximate equilibrium indicates we can say little beyond Player 1 should play *Top* more frequently than *Bottom* and Player 2 should play *Right* more frequently than *Left*. The reason for this is not hard to fathom In the 320 game incentives are relatively strong: by making a wrong choice players can lose between 40 and 280. In the 44 game by making a wrong choices player can lose between 4 and 40. Naturally when incentives are less strong the set of approximate equilibrium is larger and we are less able to make accurate predictions of how players will play.

7. Is Traditional Theory Useless?

The financial market meltdown in October 2009 has convinced many that markets are irrational, and rational models are doomed to failure. Only behavioral models recognizing the emotional “animal spirits” of investors can hope to capture the events that occur during a full blown financial panic. Most of this sentiment springs, however, from confusion about what rationality is and what rational models say.

Is it irrational to run for the exit when someone screams that the movie theater is on fire? Consider the following simple variation on the Prisoner’s Dilemma game. “Fire” has been shouted in the theater. You can rush for the exits, or proceed in an orderly way. If everybody rushes they all get 5; if everyone proceeds in an orderly way they all get 9. However, if I rush and nobody else does I am sure to get out so get 10; on the other hand if every rushes but me, I am doomed, so only get 0.

	<i>Everyone else</i>	
<i>You</i>	orderly	rush
<i>orderly</i>	9	0
<i>rush</i>	10	5

This game has a unique dominant strategy equilibrium: everyone rushes for the exits. Nowhere do we model the very real sick feeling of panic that people feel as they rush for

the exits. That is a symptom of being in a difficult situation, not an explanation of why people behave as they do.

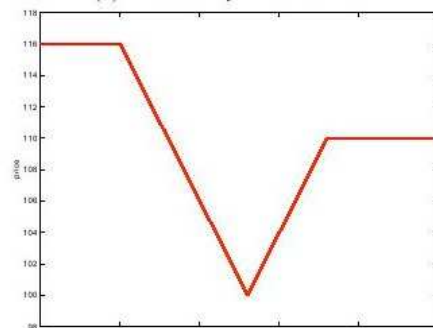
The situation in a market panic is similar. Suppose you turn on the television and notice the Chairman of the Federal Reserve Board giving a speech indicating that the financial sector is close to meltdown. It occurs to you that when this happens, stocks will not have much value. Naturally you wish to sell your stocks – and to do so before they fall in price, which is to say, to sell before everyone else can rush to sell. So there is a “panic” as everyone rushes to sell. Individual behavior here is rational – and unlike the rushing to the exits where more lives would be spared if the exodus was orderly, in the stock market there is no real harm if people rush to sell rather than selling in an orderly way.

In some circumstances people overdo it – and the price drops so much that it bounces right back up as soon as people get their wits back. Perhaps this is due to irrationality? Not at all – there is a beautiful paper written in 2009 by Lasse Pedersen

Panel (a): Minute-by-Minute Data from the Quant Event 2007.



Panel (b): Theoretically Predicted Price Path



analyzing the so-called “quant event” of August 3-14 2007, where prices did bounce. The first figure above shows the minute by minute real market price and the second figure shows prices computed from the theory. The key thing to understand is that the theory is

of pure rational expectations – irrationality, psychology, and “behavioral” economics do not enter the picture.

The same idea applies to bank runs. If you think your bank is going to fail taking your life savings with it, it is perfectly rational to try to get your money out as quickly as you can. Of course if everyone does that it pretty much guarantees the bank will fail. This is the classic Diamond and Dybvig [1983] model of bank runs – along with the some 3,639 follow-up papers. So far nobody has pointed out any facts or details about the 2009 crisis that is inconsistent with these models of rational behavior.

Economic theory, then, can understand panics. But can it predict them? No, it cannot and for good reason. One essential thing that is often overlooked in behavioral economics is that economics itself is not neutral in human behavior. People make mistakes. Because they are mistakes, if those mistakes are explained to them they will permanently change their behavior. Hence a theory of mistakes is always subject to the problem that if people understand the theory it becomes invalid. A related phenomenon occurs in markets. Suppose that we could forecast stock market crashes. A big computer program that all economists agreed was right would be run and it would tell us “Next week the stock market will fall 20%.” What would you do? Knowing the stock market will drop 20% next week, would you wait until next week to sell? Of course not, you would want to dump your stocks before everyone else did. And when everyone tried to do that the stock market would drop by 20% – but not next week, it would happen right now. You do not wait until you feel the flames before you rush for the theater exits.

Put another way, there is an intrinsic interaction between the forecaster and the forecast – at least if the forecaster is believed. Predicting economic activity is not like predicting the weather. Whether or not there is going to be a hurricane does not depend on whether or not we think there is going to be a hurricane. Whether or not there is going to be an economic crisis depends on whether or not we think there is going to be one. And this is why the economics profession came to adopt the rational expectations model. Unlike behavioral models – which treat economic activity like hurricanes – the rational expectations model captures the intrinsic connection between the forecaster and the forecast. In fact one description of a model of rational expectations is that it describes a world where the forecaster has no advantage in making forecasts over anyone else in the economy – which if people believe his forecasts will have to be the case.

The fact that we cannot predict crashes does not mean that we do not understand them. It is far from true that mainstream economics is blind “to the very possibility of catastrophic failures in a market economy” as Krugman wrote in a 2009 New York Times column. That is the same Krugman who in 1979 wrote a paper called “A Model of Balance-of-Payments.” This showed how under perfect foresight crises are ubiquitous when speculators swoop in and sell short. The paper is deficient in that it supposes that crises are perfectly foreseen and – as indicated above – this cannot lead to catastrophic drops in prices. However, the paper is not obscure, there having been some 2,354 follow-on papers, including Salant [1983]. Salant uses the tools of modern economics, in which the fundamental forces driving the economy are not perfectly foreseen, to show how rational expectations leads to speculation and unexpected yet catastrophic price drops.

Despite the fact that the idea of the Salant paper is integral to most modern economic models, it still never fails to surprise non-economists when market crises do occur. It has happened in England, in Mexico, in Argentina, Israel, Italy, Indonesia, Malaysia, Russia, and of course more than once in the United States.

8. Does Behavioral Theory Give Correct Answers?

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

The Naif at the Health Club

Consider the following facts from Della Vigna and Malmendier [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 the 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 a 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.

The Becker-DeGroot-Marschak Elicitation Procedure

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 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. In the face of such a procedure the best thing to do is not to buy low and sell high, but rather to state your true value.

This observation is fundamental to behavioral economics: it is sometimes called the endowment effect or the reference point effect. It argues that there is a reference point relative to expectations and gains and losses are measured against that reference point. The argument is that having the cup in your possession changes the reference point so

that losing the cup is a “loss,” while not having the cup in your possession means that getting the cup is a “gain.”

But is any of this true? 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 – and indeed base an entire theory of behavior on – 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.

9. Can We Understand Decision Making by Peering into the Brain?

A basic fact from computer science is that all Turing machines are equivalent: we can learn nothing of their capabilities by studying the specific way in which they are implemented. The human brain is also a general purpose computing device, capable of implementing a wide range of algorithms, ranging from simple choices between alternatives to sophisticated algorithms such as dynamic and linear programming. The types of decisions economists are interested in are generally implemented by means of sophisticated algorithms, and these decision procedures rarely occur within the brain. At a minimum paper and pencil are used, but of course we have access to computers as well. When hedge funds decide how to buy and sell, for example, they do so by means of computer programs that solve partial differential equations, not by some sort of impulsive seat of the pants hunch about what stock is about to go up.

There is another way to view this. Suppose you wanted to study Microsoft Word in order, say, to build a better word processor. Would you study the CPU of a PC? Would you study how the RAM is wired? Or the ASICS? Would you study the binary code? Surely you would do none of those things. Would you study the source code? Probably

not even that, most likely you would use the program, observe how it worked, and figure out how to build a program that did the same thing. That is what economists do. We observe human behavior and figure out models that behave the “way humans do.” We have no reason to believe that better understanding the wiring of the brain would improve our models any more than understanding the micro-code on an x86 chip would improve lead to improvements in word processors.

Where does this leave neuroeconomics? Of course there are one-time decisions that must be made quickly – for example the Naval Frigate problem. This occurred as a practical matter in 1989 when a U.S. Navy Frigate thinking it was under air attack accidentally shot down an Iranian passenger jet. Here the sailors on the frigate had only a few moments to make a quick decision based on imperfect information – shoot or not shoot? Understanding the kind of biases they may have had on the spur of the moment – ignoring certain information and attending to other information – can clearly improve performance in these sorts of circumstances.

The Naval Frigate problem, while important, is scarcely the kind of common decision problem that is at the heart of economic problems. There is, however, a broader contribution that neuroeconomics may hope to make to economics. Economists make the useful distinction between preferences and beliefs. This is the same distinction as that between goals and decision-making procedures. Given preferences – goals – economics does a good job of understanding how these are translated into action. We do not need neuroeconomics to come along and suggest that given their goals people employ some kind of primitive mechanical decision making procedure – not least because it is not true. On the other hand: economists struggle with the formation of preferences. How are goals formed? How are short-term impulses reconciled with longer-term prospects? How do we become angry? Humiliated? Why are we altruistic under some circumstances and spiteful in others? Here in studying preferences and goals – where economics is weak – peering into the brain may prove of ultimate use.

References

- Bolton, GE and A Ockenfels [2000]: “ERC: A Theory of Equity, Reciprocity, and Competition,” *American Economic Review*.
- Della Vigna, S. and U Malmendier [2006]: “Paying not to go to the gym,” *American Economic Review*.
- Diamond, D. W. and P. H. Dybvig [1983]: “ Bank Runs, Deposit Insurance, and Liquidity,” *The Journal of Political Economy*.
- 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*.
- Goeree, J. K. and C. A. Holt [2001]: “Ten Little Treasures of Game Theory and Ten Intuitive Contradictions,” *The American Economic Review*, **91**: 1402-1422.
- Gul, F and W Pesendorfer [2004]: “The Canonical Type Space for Interdependent Preferences”.
- Krugman, P. [1979] “A Model of Balance-of-Payments Crises” *Journal of Money, Credit and Banking*.
- 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.
- Levine, D. K. and J. Zheng [2010]: “The Relationship of Economic Theory to Experiments,” in *The Methods of Modern Experimental Economics*, ed. Guillaume Frechette and Andrew Schotter, Oxford University Press: Oxford

- 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*.
- Pedersen, Lasse [2009], “When Everyone Runs for the Exit,” Working Paper, NYU.
- 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*.
- 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*.
- Salant, S. W. [1983] “The vulnerability of price stabilization schemes to speculative attack” *The Journal of Political Economy*.