CHAPTER 2

THE RELATIONSHIP BETWEEN ECONOMIC THEORY AND EXPERIMENTS

DAVID K. LEVINE AND JIE ZHENG

INTRODUCTION

THE relationship between economic theory and experimental evidence is controversial. From reading the experimental literature, one could easily get the impression that economic theory has little or no significance for explaining experimental results. The point of this essay is that this is a tremendously misleading impression. Economic theory makes strong predictions about many situations and is generally quite accurate in predicting behavior in the laboratory. In most familiar situations where the theory is thought to fail, the failure is to properly apply the theory and not that the theory failed to explain the evidence.

That said, economic theory still needs to be strengthened to deal with experimental data: The problem is that in too many applications the theory is correct only in the sense that it has little to say about what will happen. Rather than speaking of whether the theory is correct or incorrect, the relevant question turns out to be whether it is useful or not useful. In many instances it is not useful. It may not be able to predict precisely how players will play in unfamiliar situations.[1] It buries too much in individual preferences without attempting to understand how individual

preferences are related to particular environments. This latter failing is especially true when it comes to preferences involving risk and time, as well as preferences involving interpersonal comparisons—altruism, spite, and fairness.

By way of contrast, in many circumstances equilibrium is robust to modest departures from assumptions about selfish and rational behavior. In these circumstances, the simplest form of the theory—Nash equilibrium with selfish preferences—explains the data quite well. In this case, as we shall explain, predictions about aggregate behavior are quite accurate. Predictions about individual behavior are better explained by a perturbed form of Nash equilibrium—now widely known as quantal response equilibrium.

Equilibrium Theory that Works

The central theory of equilibrium in economics is that of Nash equilibrium. Let us see how that theory works in a reasonably complex voting situation. The model is adapted from Palfrey and Rosenthal (1985). There are voters divided into two groups, namely, supporters of candidate A and supporters of candidate B. The number of voters is odd and divisible by three and can take on the values {3, 9, 27, 51}. Unlike the groups used by Palfrey and Rosenthal, the two groups are not equal in size; that is, group B is larger than group A. In the landslide treatment, there are twice as many members of B as of A. In the tossup treatment, there is one more voter in group B than in group A. The voters may either vote for their preferred candidate or abstain, and the rule is simple majority. The members of the winning group receive a common prize of 105, while those in the losing group receive 5. In case of a tie, both groups receive 55. Voting is costly: The costs are private information and are drawn independently and randomly on the interval [0, 55]. Players are told the rules in a common setting, and they get to play 50 times.

Computing the Nash equilibrium of this game is sufficiently difficult that it cannot be done by hand, nor is it possible to prove that there is a unique equilibrium. However, the equilibrium can be computed numerically, and grid searches show that there is only one equilibrium. The key to equilibrium is the probability of pivotal events: The benefit of casting a vote depends on the probability of being pivotal in an election. Thus a good test of Nash equilibrium is to compare the theoretical probability of a voter being pivotal—that is, of a close election—versus the empirical frequency observed in the laboratory. The graph in Figure 2.1 from Levine and Palfrey (2007) plots the theoretical probability on the horizontal axis and the empirical frequency on the vertical axis. If the theory worked perfectly, the points should align on the 45-degree line. They do. Despite the fact that both theoretically and from observing 50 data points it is no easy matter to infer the probability of being pivotal, the theory works nearly perfectly.

It deserves emphasis that when we speak of "theory" here we are speaking entirely of a theoretical computation. In finding the Nash equilibrium probabilities

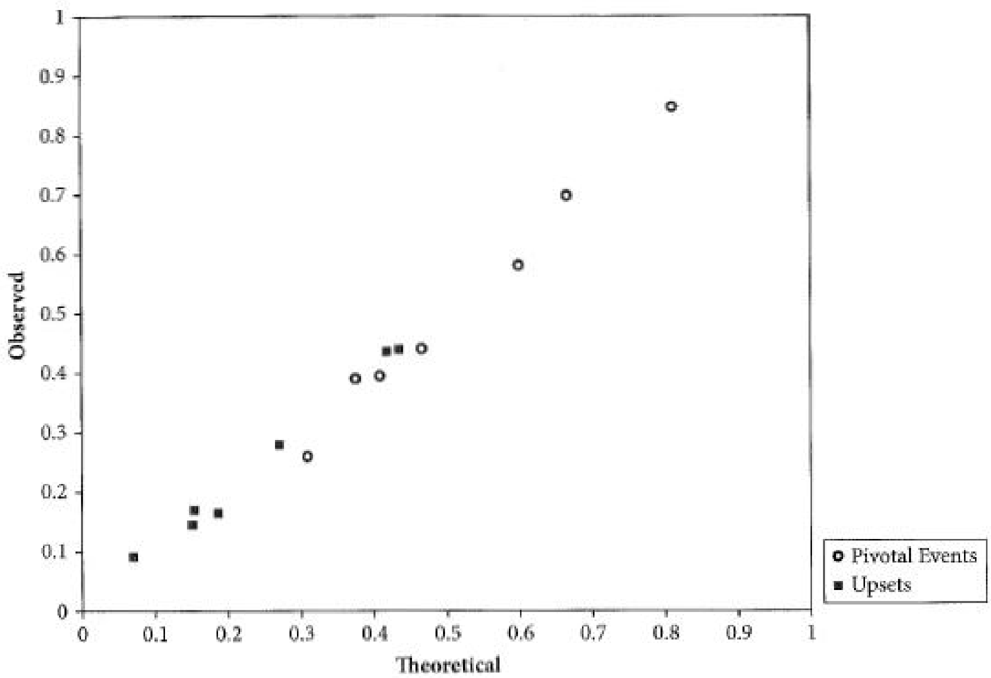
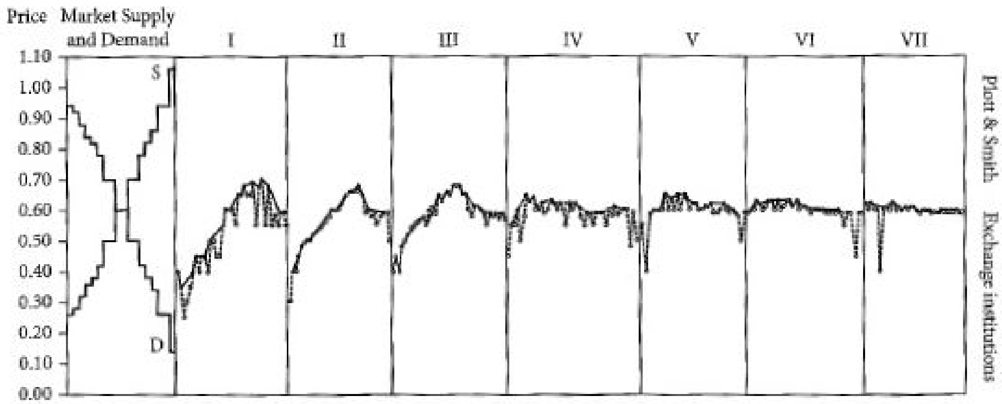


Figure 2.1. **The relationship between theoretical probability and empirical frequency for the complex voting example of Levine and Palfrey (2007).**

of being pivotal, no parameters are fit to the data: No estimation is done whatever. A pure computation is compared to live data, and the fit is nearly perfect.

The other central theory in economics besides Nash equilibrium is the competitive equilibrium of a market. In modern theory, this can be viewed as the Nash equilibrium of a mechanism in which traders reveal preferences to a market that then determines the equilibrium—with the exact details of the market clearing mechanism of no importance. Experiments on competitive equilibrium—generally in which the market clearing mechanism is a double oral auction in real time—have been conducted many times, dating back at least to the work of Smith (1962). The results are highly robust: Competitive equilibrium predicts the outcome of competitive market experiments with a high degree of accuracy, with experimental markets converging quickly to the competitive price. One typical picture is the history of bids in an experiment by Plott and Smith (1978) showing the convergence to the competitive equilibrium at a price of 60 (Figure 2.2). Again note that the competitive price of 60 is computed from purely theoretical considerations—no parameters are fit to the data.

This picture of data that nearly perfectly fits purely theoretical computations is true for a wide variety of experiments and is very much at odds with the viewpoint that experimental results somehow prove the theory wrong. Indeed the theory fits much better than models that must be estimated in order to fit noisy field data.



Sequences of bids and contracts. Experiment 2— (oral bid). ○ = unaccepted bid; ● = accepted bid (contract)

Figure 2.2. The history of bids in experiment 2 (oral bid) by Plott and Smith (1978).

EQUILIBRIUM THEORY THAT DOES NOT FAIL

Moving past theory that predicts accurately and well, there are a set of experiments in which equilibrium—especially the refinement of subgame perfection—apparently fails badly. One such example is the ultimatum bargaining game. Here one player proposes a division of $10 in nickels, and the second player may either accept or reject the proposal. If she accepts, then the money is divided as agreed upon. If she rejects, then the game ends and neither player receives any money. Subgame perfection predicts that the second player should accept any positive amount, and so the first mover should get at least $9.95. Table 2.1, with the data from Roth et al. (1991), shows that this is scarcely the case. Nobody offers less than $2.00 and most offers are for $5.00, which is the usual amount that the first player earns. Superficially, it would be hard to imagine a greater rejection of a theory than this. Moreover, like competitive market games, these results have been replicated many times under many conditions.

Despite appearances, theory is consistent with these results—it is the misapplication of the theory that leads to the apparent anomaly. First, the computation of the subgame equilibrium is based on the assumption that players are selfish—that they care only about their own money income. This assumption—which has nothing to do with equilibrium theory, but is merely an assertion about the nature of players' utility functions—is clearly rejected by the data. A selfish player would not reject a positive offer; this fact is the basis for calculating the subgame perfect equilibrium. However, the data clearly show that 5 out of 27 positive offers are rejected. The data—not to speak of common sense—show that many players find low offers offensive in the sense that they prefer nothing at all to a small share of the pie. A "theory" based on the assumption of selfish preferences will naturally

Table 2.1. Offers and Rejection Probabilities for the Ultimatum Bargaining Game

X($)	Number of 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 stakes games, round 10
Source: Roth et al. (1991)

fail to explain the data. However, there is nothing in the logic of rationality, Nash equilibrium, or subgame perfection that requires players to have selfish preferences.

In the mainstream theory of competitive markets, it is true that economists typically assume that people are selfish. This is not because economists believe that people are selfish—we doubt you could find a single economist who would assert that—but rather because in competitive markets it does not matter whether or not people are selfish because they have no opportunity to engage in spiteful or altruistic behavior. Consequently, it is convenient for computational purposes to model people in those environments as being selfish. That should not be taken to mean that this useful modeling tool should be ported to other inappropriate environments, such as bargaining situations.

Surprisingly, even the theory of selfish preferences does not do so badly as a cursory inspection of the data might indicate. Nash equilibrium—as opposed to subgame perfection—allows any offer to be an equilibrium: It is always possible that any lower offer than the one the first player makes might be rejected with probability one, while the current offer is accepted. Nash equilibrium rules out two less obvious features of the data. It rules out a heterogeneity of offers, and it rules out offers being rejected in equilibrium (if players are truly selfish). It is a mistaken view of the theory that leads to the conclusion that this is a large discrepancy. Any theory is an idealization. Players' exact preferences, beliefs, and so forth are never going to be known exactly to the modeler. As a result, the only meaningful theory of Nash equilibrium is Radner's (1980) notion of epsilon equilibrium. This requires only that no player loses more than epsilon compared to the true optimum—which in practice can never be known by the players. The correct test of the goodness of fit of Nash equilibrium in experimental data is not whether the results look like a Nash equilibrium, but rather whether players' losses (epsilon) are small relative to what they might have had.

The correct calculation of the departure of the facts from the theory, in other words, is to (a) determine how much money a player who had available the experimental data could have earned and (b) compare it to how much that player actually earned. To the extent that this is a large amount of money, we conclude that the theory fits poorly. To the extent that it is a small amount of money, we conclude that the theory fits well. This is regardless of whether the data "appear like" a Nash equilibrium or not. The key point is that allowing a small epsilon in certain games can result in a large change in equilibrium behavior. This large change does not contradict the theory of equilibrium—it is predicted by the theory of equilibrium.

For the ultimatum game, Fudenberg and Levine (1997) calculated the losses that players suffered from playing less than optimal strategies given the true strategies of their opponents. Out of the $10 on the table, players only lose on average about $1.00 per game.

This is not the end of the story, however. Nash equilibrium, at least as it is currently viewed, is supposed to be the equilibrium in which players understand their environment, including how their opponents play. It is supposed to be the outcome of a dynamic process of learning—indeed, it may accurately be described as a situation where no further learning is possible. This is important in the games in which the theory worked: In the voting experiment, players played 50 times and thus had a great deal of experience. Similarly, in the double oral auctions, players got to participate in many auctions and equilibrium occurs only after they acquire experience. In the ultimatum game, players got to play only 10 times. More important, in an extensive form game where players are informed only of the outcomes and not their opponents' strategies, players would have to engage in expensive active learning to achieve a Nash equilibrium; and without a great deal of repetition and patience, they have no incentive to do so. In ultimatum bargaining in particular, the first mover can only conjecture what might happen if she demanded more—in 10 plays there is relatively little incentive or opportunity to systematically experiment with different offers to see which will be rejected or accepted. If the game were played 100 times, for example, then it would make sense to try demanding a lot to see if perhaps the opponent would be willing to accept bad offers. In 10 repetitions such a learning strategy does not make sense.

A weaker theory than Nash equilibrium—but one more suitable to the ultimatum bargaining environment—is that of self-confirming equilibrium introduced in Fudenberg and Levine (1993). This asserts that players optimize given correct beliefs about the equilibrium path, but does not require that they know correctly what happens off the equilibrium path, as they do not necessarily observe that. This makes a difference when computing the amount of money players "lose" relative to the true optimum. As we observed in ultimatum bargaining, the first movers cannot know what will happen if they demanded more. So setting a demand that is too low is not a "knowing" error, in the sense that the player has no way to know whether it is an error or not. This leads us to compute not just the losses made by a player relative to the true optimum, but to compute how many of those losses are

"knowing losses," meaning that the player might reasonably know that he is making a loss. Self-confirming equilibrium is a theory that predicts that knowing losses should be low—but makes no prediction about unknowing losses.

For the ultimatum game, Fudenberg and Levine (1997) also calculated the knowing losses. On average, players lose only $0.33 per game, and this is due entirely to second players turning down positive offers—which as we noted has nothing to do with equilibrium theory at all. It is interesting to compare the impact of preferences (the spiteful play of the second players) versus that of learning (the mistaken offers of the first players). On average, players lose $0.33 due to having preferences that are not selfish, and on average they lose $0.67 because they lack adequate opportunity to learn about their opponents' strategies. The losses due to the deviation of preferences from the assumption of selfish behavior are considerably less than the losses due to incomplete learning.

The message here is not that theory does well with ultimatum bargaining. Rather the message is that theory is weak with respect to ultimatum bargaining—very little data in this game could be inconsistent with the theory. Rather, by applying the theory inappropriately, the conclusion was reached that the theory is wrong, while the correct conclusion is that the theory is not useful. Modern efforts in theory are quite rightly directed toward strengthening the theory—primarily by better modeling the endogenous attitudes of players toward one another as in Levine (1998), Fehr and Schmidt (1999), Bolton and Ockenfels (2000), or Gul and Pesendorfer (2004).

We can tell a similar tale of poorly applied subgame perfection in the other famous "rejection" of theory, the centipede game of McKelvey and Palfrey (1992) (Figure 2.3).

The extensive form of the game is shown below. There are two players, and each may take 80% of the pot or pass, with the pot doubling at each round. Backwards induction says to drop out immediately. In fact, as the empirical frequencies in the diagram show, only 8% of players actually do that. As in ultimatum bargaining, the evidence seems to fly in the face of the theory. Again, a closer examination shows that this is not the case.

In a sense, this centipede game is the opposite of ultimatum. In ultimatum the apparent discrepancy with theory was driven by the fact that second movers are

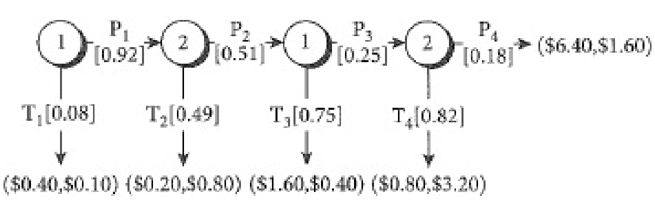


Figure 2.3. The centipede game of McKelvey and Palfrey (1992): Numbers in square brackets correspond to the observed conditional probabilities of play at each information set in rounds 6–10.

spiteful in the sense of being willing to take a small loss to punish an ungenerous opponent. In centipede the discrepancy is driven by altruism—by the willingness of a few players to suffer a small loss to provide a substantial reward to a generous opponent. The crucial empirical fact is that 18% of players will make a gift to their opponent in the final round. Note that it costs them only \$1.60 to give a gift worth \$5.60. These gifts change the strategic nature of the game completely. With the presence of gift-givers, the true optimal strategy for each player is to stay in as long as possible. If you are the first mover, stay in and hope you get lucky in the final round. If you are the second mover and make it to the final round, go ahead and grab then.

Most of the losses in centipede are actually suffered by players (foolishly misapplying subgame perfection?) who do not realize that they should stay in as long as possible, and so they drop out too soon. Overall losses were computed by Fudenberg and Levine (1997) to be about \$0.15 per player per game. However, if you drop out too soon, you never discover that there were players giving money away at the end of the game, so those losses are not knowing losses. The only knowing losses are the gifts by players in the final round. These amount to only \$0.02 per player per game. Note that as in ultimatum, failed learning is responsible for substantially greater losses than deviation in preferences from the benchmark case of selfishness.

Another important effort is to try to capture the insight of epsilon equilibrium—that when some players deviate a little from equilibrium play, this may greatly change the incentives of other players—without losing the predictive power of Nash equilibrium. The most important effort in that direction is what has become known from the work of McKelvey and Palfrey (1995) as quantal response equilibrium. This allows for the explicit possibility that players make random errors. Specifically, if we denote the utility that a player receives from her own pure strategy s_i and opponents mixed strategy σ_{-i} by $u_i(s_i, \sigma_{-i})$ and let $\lambda_i > 0$ be a behavioral parameter, we define the propensity with which different strategies are played by

$$p_i(s_i) = \exp(\lambda_i u_i(s_i, \sigma_{-i})).$$

Quantal response theory then predicts that the mixed strategies that will be employed are given by normalizing the propensities to add up to 1:

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

This theory, like Nash equilibrium, makes strong predictions. As $\lambda_i \longrightarrow \infty$, these predictions in fact converge to those of Nash equilibrium. One important strength of this theory is that it allows for substantial heterogeneity at the individual level. This is important, because experimental data are quite noisy and individual behavior is generally heterogeneous.

A good example of this is in the Levine and Palfrey (2007) voting experiment described in the first section. The aggregate fit of the theory was very good; but

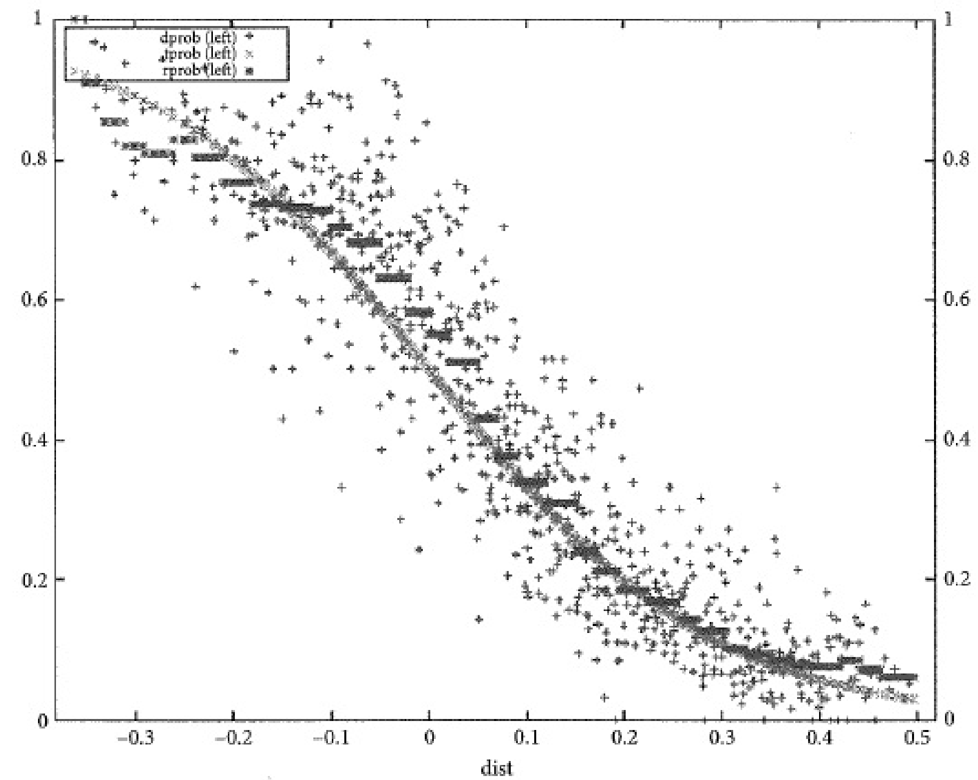


Figure 2.4. The relationship between turnout rate and the loss from participating for the complex voting example of Levine and Palfrey (2007).

at the individual level, the theory fits poorly. Figure 2.4 taken from that paper shows the empirical probability with which a voter participates as a function of the loss from participating. If the loss is positive, Nash equilibrium predicts that the probability of participation should be zero; if it is negative, the probability of participation should be one, and the data should align themselves accordingly. The individual data, represented by plus-signs and the aggregated data represented by the darker lines show that this is by no means true. When losses and gains are small, the probability of participation is relatively random—near 50%. As the loss from participating increases, the probability of participating decreases— but it hardly jumps from 1 to 0 as the threshold of indifference is crossed. However, the gradual decline seen in the data is exactly what is predicted by quantal response equilibrium. Quantal response predicts that when players are near indifferent, they effectively randomize. As incentives become stronger, they play more optimally. The downward sloping curve shows the best-fit quantal response function, where λ_i is estimated from the data. As can be seen, it fits the individual level data quite well.

A key idea here is that in the aggregate, quantal response equilibrium may or may not be sensitive to values of λ_i that are only moderately large. In some games,

such as the voting game, it makes little difference to aggregate behavior what λ_i is, since some voters over-voting makes it optimal for other voters to under-vote. Similarly in the market games, individual errors do not matter much at the aggregate level. The important thing is that we can always compute the quantal response equilibrium and determine how sensitive the equilibrium is to changes in λ_i.

A good illustration of the strength—and potential weakness—of quantal response equilibrium is the mixed strategy example of the asymmetric matching pennies game, described in Goeree and Holt (2001).[2] It is a simple simultaneous game where the row player chooses between *Top* and *Bottom* and the column player chooses between *Left* and *Right*. The payoff is (40, 80) when the outcome is (*Top*, *Right*) or (*Bottom*, *Left*), and it is (80, 40) when the outcome is (*Bottom*, *Right*). It would be symmetric if the payoff for the outcome (*Top*, *Left*) was (80, 40), but here we are interested in the asymmetric cases where the payoff for (*Top*, *Left*) is (320, 40) in one case (denoted "the (320, 40) case") and (44, 40) in the other (denoted "the (44, 40) case"). The data in Goeree and Holt (2001) show in the lab that 96% of row players play *Top* and that 16% of column players play *Left* in the (320, 40) case, with the fraction numbers 8% and 80% respectively for the (44, 40) case. It is obvious that these lab results are quite different from what the theory of Nash equilibrium predicts, where the fraction of row players playing *Top* should be 50% in both cases.

If we apply the theory of quantal response equilibrium to this mixed strategy example, the prediction power can be improved by a large degree. We do the calculations using each of the two alternative assumptions[3]: (1) the standard selfish preference assumption ($U_i = u_i$) and (2) the more realistic altruistic preference assumption ($U_i = \alpha u_i + (1-\alpha)u_{-i}$, where $\alpha \in [0, 1)$). In Figure 2.5, the horizontal axis represents the fraction of row players who play *Top* and the vertical axis represents the fraction of column players who play *Left*. Both Nash and quantal response equilibria are shown: The original equilibrium corresponding to the selfish case and the "new" equilibrium corresponding to altruistic preference with parameter $\alpha = 0.91$ are shown. The curves correspond to different quantal response equilibria with different values of λ. Note that we assume $\lambda_1 = \lambda_2 = \lambda$ since players are drawn from the same population.

By allowing players to make mistakes, as we can see from the graph, the theory of quantal response equilibrium gives a better prediction than Nash equilibrium does. This is especially true in the (320, 40) case with altruistic preference assumption: When $\lambda = 20$, the quantal response equilibrium is quite close to what the experimental data show. It is also worth noting from the graph that the improvement in results from applying quantal response equilibrium alone (for example, in the (320, 40) case, equilibrium shifted from (0.5, 0.13) to (0.82, 0.22)) is more than the improvement from assuming altruistic preference alone (respectively, equilibrium shifted from (0.50, 0.13) to (0.75, 0.12)). What remains mysterious is the (44, 40) case, where the lab result is poorly explained either by allowing people to make mistakes or by the preference of altruism.

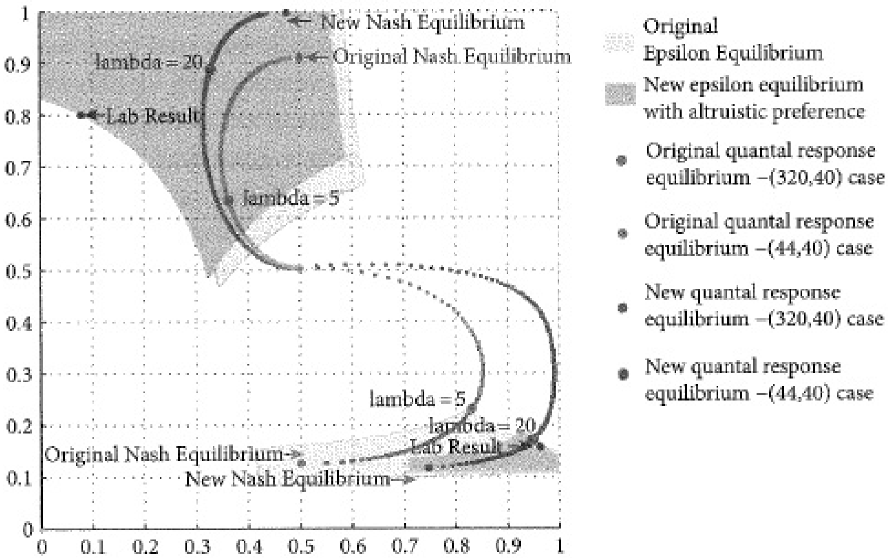


Figure 2.5. Quantal response equilibrium and ε equilibrium for the asymmetric matching pennies game of Goeree and Holt (2001).

We can also analyze ε equilibrium in this game. Under the selfish preference assumption, the laboratory data correspond to an value of \$0.07 per player per game for the (44, 40) case and \$0.06 for the (320, 40) case. The maximum possible amount that could be earned in each game is \$0.80 for the (44, 40) case and \$3.20 for the (320, 40) case. The ε values are \$0.05 and \$0.02 respectively under the altruistic preference assumption. In the (44, 40) case the set of possible equilibria is quite large: Pretty much any mixture in which the fraction of row players playing *Top* is less than 50% and the fraction of column players who play *Left* is greater than 50% is an ε equilibrium. In this sense it is not surprising that the lab result is far off the prediction of the selfish-rational theory. What is interesting is that the perturbations to payoffs that explain the laboratory result are neither due to errors (quantal response) nor due to altruism.

In the (320, 40) case, the set of equilibria is not so large. It predicts little about the row players' play—just that the row player should play *Top* more than 50% of the time. This must be the case, as the Nash equilibrium requires 50% play *Top*, and in the laboratory result 96% play *Top*, and of course both of these must lie in the equilibrium set. Note, however, that the play of the column player is predicted with a relatively high degree of precision: It lies in the range of 10–22%.

What Experiments Have Taught Us

Experimental economics has certainly taught us where the theory needs strengthening—as well as settling some long-standing methodological issues. For

example, the issue of "why should we expect Nash equilibrium" has always had two answers. One answer is that players introspectively imagine that they are in the shoes of the other player, and they reason their way to Nash equilibrium. This theory has conceptual problems, especially when there are multiple equilibria. It also has computational issues—for example, there is a great deal of evidence that the game in which commuters choose routes to work during rush hour is in equilibrium, although individual commuters certainly do not compute solutions to the game. Nevertheless, in principle, players might, at least in simpler games, employ a procedure such as the Harsanyi and Selten (1988) tracing procedure. Experimental evidence, however, decisively rejects the hypothesis that the first-time players are exposed to a game they manage to play a Nash equilibrium. As a result, the current view—for example, in Fudenberg and Levine (1998)—is that if equilibrium is reached, it is through learning. For example, the rush hour traffic game is known from the work of Monderer and Shapley (1996) to be a potential game, and such games have been shown, for example by Sandholm (2001), to be stable under a wide variety of learning procedures.

As Nash equilibrium cannot predict the outcome of one-off games, one area of theoretical research is to investigate models that can. The most promising models are the type models of Stahl and Wilson (1995): Here players are viewed as having different levels of strategic sophistication. At the bottom level, players play randomly; more sophisticated players optimize against random opponents; players who are even more sophisticated optimize against opponents who optimize against random opponents, and so forth. Experimental research, for example by Costa-Gomes et al. (2001), shows that these models can explain a great deal of first-time play, as well as the details of how players reason. The greatest lacuna in this literature is that it has not yet been well tied in to a theory of learning: we have a reasonable theory of first-time play and a reasonable theory of long-term play, but the in-between has not been solidly modeled.

The second area we highlighted above is the area of interpersonal preferences: altruism and spite. As mentioned, there are a variety of models including Levine (1998), Fehr and Schmidt (1999), Bolton and Ockenfels (2000), or Gul and Pesendorfer (2004), that attack this problem, but there is not as yet a settled theory.

There is one "emperor has no clothes" aspect of experimental research. This involves attitudes toward risk. The standard model of game theory supposes that players' preferences can be represented by a cardinal utility function. The deficiency in this theory was highlighted by Rabin's (2000) paradox

> Suppose we knew a risk-averse person turns down 50–50 lose \$100/gain \$105 bets for any lifetime wealth level less than \$350,000, but knew nothing about the degree of her risk aversion for wealth levels above \$350,000. Then we know that from an initial wealth level of \$340,000 the person will turn down a 50–50 bet of losing \$4,000 and gaining \$635,670.

The point here is that in the laboratory, players routinely turn down 50–50 lose $100/gain $105 gambles and even more favorable gambles. Yet this is not only inconsistent with behavior in the large, it is off by (three!) orders of magnitude. Roughly, the stakes in the laboratory are so small that any reasonable degree of risk aversion implies risk neutrality for laboratory stakes—something strongly contradicted by the available data.

There are various possible theoretical fixes, ranging from the prospect theory of Tversky and Kahneman (1974) to the dual-self approach of Fudenberg and Levine (2006), but it is fair to say that there is no settled theory and that this is an ongoing important area of research.

Conclusion

The idea that experimental economics has somehow overturned years of theoretical research is ludicrous. A good way to wrap up, perhaps, is with the famous prisoner's dilemma game. No game has been so much studied either theoretically or in the laboratory. One might summarize the widespread view as follows: People cooperate in the laboratory when the theory says they should not. *Caveat emptor*. The proper antidote to that view can be found in the careful experiments of Dal Bo (2005). The proper summary of that paper is as follows: Standard Nash equilibrium theory of selfish players works quite well in predicting the laboratory behavior of players in prisoner's dilemma games.

What experimental economics has done very effectively is to highlight where the theory is weak, and there has been an important feedback loop between improving the theory—quantal response equilibrium being an outstanding example—and improving the explanation of experimental facts.

NOTES

We are grateful to NSF grant SES-03-14713 for financial support, to Drew Fudenberg and Tom Palfrey for many conversations on this topic, to Colin Camerer for helpful comments, and to Guillaume Frechette for encouraging us to do this.

1. The theory sometimes can still make a good prediction even when players are not familiar with the game being played. See Camerer (2003) for examples.

2. Note, however, that players only got to play once, so no learning was possible.

3. We also did the calculation by assuming that a fraction $(1 - \beta)$ of people have altruistic preference and that the rest ($\beta \in (0, 1)$) of the people are selfish, but the result is not improved much from the case in which $\beta = 0$, which is equivalent to assumption 2.

REFERENCES

Bolton, G. E. and A. Ockenfels. 2000. ERC: A Theory of Equity, Reciprocity, and Competition. *American Economic Review* **90**:166–193.

Camerer, C. F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton, NJ: Princeton University Press.

Costa-Gomes, M., V. P. Crawford, and B. Broseta. 2001. Cognition and Behavior in Normal-Form Games: An Experimental Study. *Econometrica* **69**(5):1193–1235.

Dal Bo, P. 2005. Cooperation under the Shadow of the Future: Experimental Evidence from Infinitely Repeated Games. *American Economic Review* **95**(5):1591–1604.

Fehr, E. and K. M. Schmidt. 1999. A Theory Of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics* **114**(3):817–868.

Fudenberg, D. and D. K. Levine. 1993. Self-Confirming Equilibrium. *Econometrica* **61**(3):523–545.

Fudenberg, D. and D. K. Levine. 1997. Measuring Players' Losses in Experimental Games. *Quarterly Journal of Economics* **112**(2):507–536.

Fudenberg, D. and D. K. Levine. 1998. *The Theory of Learning in Games*, MIT Press.

Fudenberg, D. and D. K. Levine. 2006. A Dual Self Model of Impulse Control. *American Economic Review* **96**(5):1449–1476.

Goeree, J. K and C. A. Holt. 2001. Ten Little Treasures of Game Theory and Ten Intuitive Contradictions. *American Economic Review* **91**(5):1402–1422.

Gul, F. and W. Pesendorfer. 2004. The Canonical Type Space for Interdependent Preferences. Unpublished.

Harsanyi, J. C. and R. Selten. 1988. *A General Theory of Equilibrium Selection in Games.* Cambridge, MA: MIT Press.

Levine, D. K. 1998. Modeling Altruism and Spitefulness in Experiments. *Review of Economic Dynamics* **1**(3):593–622.

Levine, D. K. and T. R. Palfrey. 2007. The Paradox of Voter Participation: A Laboratory Study. *American Political Science Review* **101**:143–158.

McKelvey, R. D. and T. R. Palfrey. 1992. An Experimental Study of the Centipede Game. *Econometrica* **60**(4):803–836.

McKelvey, R. D. and T. R. Palfrey. 1995. Quantal Response Equilibria for Normal Form Games. *Games and Economic Behavior* **10**:6–38.

Monderer, D. and L. S. Shapley. 1996. Potential Games. *Games and Economic Behavior* **14**:124–143.

Palfrey, T. R. and H. Rosenthal. 1985. Voter Participation and Strategic Uncertainty. *American Political Science Review* **79**(1):62–78.

Plott, C. R. and V. L. Smith. 1978. An Experimental Examination of Two Exchange Institutions. *Review of Economic Studies* **45**:133–153.

Rabin, M. 2000. Risk Aversion and Expected-Utility Theory: A Calibration Theorem. *Econometrica* **68**(5):1281–1292.

Radner, R. 1980. Collusive Behavior in Noncooperative Epsilon-Equilibria of Oligopolies with Long but Finite Lives. *Journal of Economic Theory* **22**(2):136–154.

Roth, A. E., 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* **81**(5):1068–1095.

Sandholm, W. H. 2001. Potential Games with Continuous Player Sets. *Journal of Economic Theory* **91**:81–108.

Smith, V. L. 1962. An Experimental Study of Competitive Market Behavior. *Journal of Political Economy* **70**(2):111–137.

Stahl, D. O. and P. W. Wilson. 1995. On Players' Models of Other Players: Theory and Experimental Evidence. *Games and Economic Behavior* **10**:218–254.

Tversky, A. and D. Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science* **185**:1124–1131.