



Why Experiment in Economics?

Ken Binmore

The Economic Journal, Vol. 109, No. 453, Features (Feb., 1999), F16-F24.

Stable URL:

<http://links.jstor.org/sici?sici=0013-0133%28199902%29109%3A453%3CF16%3AWEIE%3E2.0.CO%3B2-T>

The Economic Journal is currently published by Royal Economic Society.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/res.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

WHY EXPERIMENT IN ECONOMICS?*

Ken Binmore

Experimental economics is now so popular that a Nobel prize is doubtless imminent for those who pioneered the subject. But one may ask what economists are doing in the laboratory at all. Would it not be better to leave laboratory experiments to psychologists who are trained to run them properly?

The answer to this question requires making the same distinction in social science that hard scientists make between physics and engineering. Physicists try to find out how the world works and engineers use whatever happens to be known at present to make things. A similar unacknowledged division exists in economics between what one might call scientific economists and policy advisers. It seems to me uncontroversial that laboratory experimentations for policy purposes—as in Plott's recent testing of the rules of the big American spectrum auction for the Federal Communications Commission—is not only firmly established as a tool for widening debate, but that it is an activity that can only sensibly be undertaken by economists who understand the institutions that are to be reformed. My concern is therefore with whether *scientific* economists are wasting their time in the laboratory.

I think the answer would be *yes* if economics were really as disastrous at predicting in the laboratory as the school of Kahneman and Tversky (1988) would have us believe. Indeed, I think there would be a good case for abandoning economics altogether if their wider claims were valid, since the rejoinder that economics somehow only works in the field seems pretty feeble to me.

However, we should not be so easily led by the nose. In the first place, economists should be aware how Kahneman and Tversky have been criticised in the experimental psychology literature for the failure of their results to withstand simple robustness tests (Gigerenzer, 1996). Within our own literature, we have Bohm (1993, 1994) reporting that preference reversals disappear when subjects have strong enough incentives, Loomes (1993) reporting no endowment effect under some conditions, and Cubitt *et al.* (1996) finding no common-ratio effect under others. But there is a much more important point, which militates in favour of economists running their own experiments. Under the circumstances in which the school of Kahneman and Tversky run experiments, we should not expect to find economic theory predicting well anyway!

I suppose that there are still economic theorists who believe that an invisible

* I am grateful to the Economic and Social Research Council and to the Leverhulme Foundation for funding this work through the Centre for Economic Learning and Social Evolution at University College London.

hand somehow makes people behave as if they were all omniscient superbeings able to think their way to equilibrium at a moment's notice. But most theorists nowadays agree that people get to equilibrium—insofar as they do—by an interactive process of trial-and-error learning. Taking this view requires accepting that they will not get to equilibrium at all if we put them in too complicated a situation, or fail to make paying attention to what is going on worth their while. My own experimental papers therefore insist that economic theory should only be expected to predict in the laboratory if the following three criteria are satisfied:

- The problem the subjects face is not only 'reasonably' simple in itself, but is framed so it seems simple to the subjects;
- The incentives provided are 'adequate';
- The time allowed for trial-and-error adjustment is 'sufficient'.

Bamboozling subjects with great arrays of numbers or asking them what they *would* do if \$100 were hanging on the outcome are therefore out. Even more to the point, so are all the experiments in which inexperienced subjects are asked to solve a problem that they have never seen before and will never see again.

I know that denying the predictive power of economics in the laboratory except under such conditions implies that we must also deny the predictive power of economics *in the field* when such conditions are not satisfied. But have we not got ourselves into enough trouble already by claiming vastly more than we can deliver? I am certainly tired at having fun poked at me by marketing experts for supposedly believing that economic consumer theory is relevant to the behaviour of customers buying low-cost items under supermarket conditions. How could customers find the time to research the value of the products on sale? Even if they could, the supermarkets would simply speed up the rate at which they differentiate their products and packaging.

To admit that optimising theories are sometimes misapplied is not to say that optimising theories never work at all. But this is not a position that seems to have occurred to those experimentalists whose professed aim is the debunking of economic theory. However, I will happily undertake to refute chemistry if you give me leave to mix my reagents in dirty test tubes. Equally, I will undertake to prove in the laboratory that young stockbrokers cannot learn their trade by denying my subjects access to the conventional wisdom that the stockbroking profession has built up over many years of interactive trial-and-error learning.

Just as we need to use clean test tubes in chemistry experiments, so we need to get the laboratory conditions right when testing economic theory. Scientific experimental economists can then settle down to the twofold role of consolidating economic theory in those contexts where it works fairly well, and providing a source of inspiration for revising the theory where it does not work so well. In situations where the theory does not work at all, we need to accept our limitations and look to disciplines like psychology for the answers.

By consolidating economic theory, I mean quantifying the words 'reason-

ably', 'adequate' and 'sufficient' in the three criteria suggested above. By interpreting the criteria severely enough, we can almost guarantee that an optimising theory will work. By interpreting them loosely enough, we can almost guarantee that an optimising theory will fail. But where does the boundary lie between these two extremes? My own experience in piloting experiments has not endowed me with any great confidence in my own ability to guess where the boundary lies in advance. In the sort of experiments in which I am interested, one can rely on the subjects' behaviour changing as they are paid more and gain experience, but it is hard to predict how fast their behaviour will change or how it will be affected by the manner in which the experiment is framed. It is to explore questions of this kind that a consolidating experimentalist runs experiments. He does not see his task as asking whether economics works or not. He already knows that sometimes it does and sometimes it does not. The task is to classify economic environments into those where the theory works and those where it does not. In the shady area in between, the data obtained from the classification effort will then serve to refine the theories that do not quite make it.

In the following two case studies, the first illustrates the need to replace the debunking motive by a consolidating motive. The second illustrates how experimentation can help in revising inadequate theories.

Case Study 1: Two-person, Zero-sum Games. The theory of mixed strategies in two-person, zero-sum games was invented by Von Neumann (1928). Given the reliance that modern economic theory places on the theory of games, one might therefore have imagined that Von Neumann's minimax theory would have been confirmed many times over in the laboratory. However, until recently, there were very few experimental studies that were directly relevant to the theory, and none of these studies were supportive. But it is not hard to see why. For example, in the study conducted by Estes (1957), the subjects were not even told that they were playing against an opponent. Even if they guessed that they were playing a game, they would still have to guess what its payoffs were! Other more intelligently designed studies, like that of Suppes and Atkinson (1960) required the subjects to read long typewritten instructions and then to participate in a slow pencil-and-paper implementation with low incentives.

The situation changed when O'Neill (1987) reopened the subject. But his positive findings were criticised by Brown and Rosenthal (1990), largely because O'Neill's data fail to show that the subjects were choosing independently in successive trials. The later studies of Rapaport and Boebel (1989) and Mookherjee and Sopher (1997) applied the same tests to their data, and so were also led to argue that their experiments refute the minimax theory.

But what is the point of debunking the minimax theory on such grounds? It is obviously true that people do not play minimax straight off. To do so, they would have to believe it to be common knowledge that everybody is cleverer than Borel, the great mathematician who preceded Von Neumann in studying zero-sum games but abandoned the subject because he thought the minimax

theorem was probably false. The only interesting question is whether people *learn* to play minimax in repeated trials against changing opponents. But, if people are learning, their behaviour in the current trial will *necessarily* be dependent on what happened in previous trials. In this case, the debunkers therefore succeeded in inventing a criterion for rejecting the theory that cannot be satisfied under the only conditions for which the theory might be valid.

Recently, Binmore *et al.* (1996) conducted a study that is sensitive to the learning issue. In particular, row and column players each saw a graph updated in real time that allowed them to compare a moving average of their own payoff with a moving average of the median subject in the same situation as themselves. Under such conditions, the minimax theory proved to be well supported.

Critics from the Kahneman-Tversky school will argue that we should not be surprised that people learn to optimise when placed in an environment with such a rich feedback structure. Tversky used to dismiss evidence of learning in economics experiments by saying psychology shows that subjects can be *taught* to adopt whatever behaviour an experimenter chooses to force upon them. It is certainly true that human beings are frightfully willing to obey a sufficiently determined authority figure (Milgram, 1975). However, being provided with information about how successful you are compared with other people in the same position as yourself is not the same as being told what to do by an authority figure. Our minimax experiment provides information about the average payoff of the median person in the same boat as yourself in an attempt to capture some of what is going on in real life when we learn how to do things well by copying people who are more successful than us. How else do we academics learn to do research if not by modelling ourselves on those we admire? How do young stockbrokers learn their trade if not by adopting the rules-of-thumb they see their seniors using?

I would certainly like to know whether the optimising behaviour observed in our minimax experiment would survive as the experimental environment is made progressively more adverse for learning. However, I would not have joined the Kahneman-Tversky school in denying the validity of orthodox economic reasoning if our subjects had failed to learn to optimise. The feedback environment we offered our subjects may be rich compared with the experiments of the Kahneman-Tversky school, but it is very sparse indeed compared with the environments within which real economic agents learn. For example, our subjects were not informed about the strategies that other people in the same position as themselves were using.

Case Study 2: The Ultimatum Game. In the Ultimatum Game (Güth *et al.*, 1982), player I proposes a division of a sum of money to player II. If player II refuses, both get nothing. According to the rational expectations reasoning incorporated in Selten's (1975) idea of a subgame-perfect equilibrium, player I should get all but a penny of the money on the table, since player II would be irrational to refuse a penny when the alternative

is nothing at all. But experiment does not confirm this prediction. Results in this much replicated experiment vary, but player I is most likely to make a proposal not too distant from fifty:fifty, and player II is more likely than not to refuse if offered less than a third of the money.

Most psychologists and sociologists take this result to confirm what they knew all along—that optimising theories have little or no application to humans, who mostly just unthinkingly operate whatever social norm happens to be relevant. In the case of the Ultimatum Game, so the story goes, a fairness norm gets triggered by the manner in which the experiment is framed in the laboratory (Kahneman *et al.*, 1986). Traditional economists prefer to save the optimising hypothesis by postulating that the players are not maximisers of money, but instead have exotic preferences that incorporate a spite component or a taste for fairness (Bolton, 1991).

One cannot say that either theory is wrong, because both can be made to fit the data very successfully. Subjects are also willing to embrace either or both when debriefed after the experiment. As a *description* of the phenomena they describe, each theory does quite well. But neither passes muster as an *explanation* of the data, because they leave the questions that really matter unanswered. The question for the first theory is: why is this particular fairness norm observed and not some other fairness norm? The question for the second theory is: why does player II develop exotic preferences in such situations?

I believe that to answer such questions requires using optimising theory in models of social evolution that study the formation of norms or the development of preferences. Such an approach makes it necessary to recognise that subjects' attitudes *change over time*. This fact was particularly evident in a two-stage Ultimatum Game studied by Binmore *et al.* (1985). At the first trial, subjects played as one would expect from the one-stage case. But the experience acquired at the first trial led subjects to move very markedly in the optimising direction at the second trial.

For a number of years, this result led to my social citation index doing very well as critics misrepresented our conclusions as the claim that backward induction will always be observed in all circumstances. This is a bizarre claim to attribute to me since I am equally notorious in the theoretical literature for denying that common knowledge of rationality implies backward induction in finite games of perfect information (Binmore, 1987, 1992, 1996a, 1997). In the experimental literature, people still quote Thaler's (1988) canard that our results on the two-stage Ultimatum Game are worthless because we told the subjects what to do. If so, why did they not do what they were told at the first trial? Why do our results fit neatly into the pattern found by later experimenters exploring other parameter values (Holt and Davis, 1993, p. 272)?

My own view is that the *Sturm und Drang* generated in the literature by the Ultimatum Game is largely misplaced. After all, it is because Selten expected pathological results in this case that he proposed the experiment to Güth in the first place. However, the game has succeeded in focusing attention on a very important theoretical issue: namely, the inadequacy of the literature on refinements of Nash equilibrium. As Samuelson (1994, 1997) argues, if we are

using equilibria to predict what people will do because we believe that some process of interactive learning will continue to operate until an equilibrium is reached, then we have no choice but to abandon the principle that weakly dominated strategies can be deleted *even once*, let alone successively, as in the backward induction algorithm. In fact, Binmore *et al.* (1993) show that slightly perturbed versions of the replicator dynamics converge readily on Nash equilibria of the Ultimatum Game that are not subgame perfect. That is to say, it was a theoretical mistake in the first place to argue that a sound optimising theory of human behaviour will necessarily single out the subgame-perfect equilibrium. It remains true that a Nash equilibrium which is robust against Selten's (1975) trembling-hand errors will necessarily result in play that follows the backward-induction path, but adopting a learning perspective forces one to abandon the naive assumption that it is adequate to compute an equilibrium for a fixed set of perturbations and then study the limiting case when the perturbations become negligible using the methods of comparative statics.

When people learn, the noise in the process is shaped endogenously along with the equilibrium strategies to which the process converges. In particular, the *rates* at which people learn not to evade different kinds of irrationality turn out to be crucial. In certain games, some kinds of irrationality are so damaging that they are eliminated at such a rate that equilibrium is reached before other less damaging kinds of irrationality have been affected much at all. Refinements of Nash equilibrium that are based on the absence of the latter irrationalities are then unlikely to predict human behaviour. But such theoretical failures are not guaranteed. In particular, nobody is saying that backward induction will *always* fail or *always* succeed. Instead, I am putting the consolidating view that it would be a good idea to use both experimental and theoretical tools to find out when and why backward induction predicts successfully and when and why it does not.

To make progress with such problems, we need to study models of interactive learning, but I think the current fashion for using econometric techniques to fit learning models to experimental data is hopelessly premature. The replicator dynamics¹ or fictitious play or one of the other simple models that can be taken down from the shelf certainly capture enough qualitative features of the data to make further research into learning models worthwhile, but it is not enough to identify broad statistical correspondences. We need a theory that successfully tracks the learning behaviour of individuals. I find it encouraging that theorists and experiments have now turned their attention to this question, but it will be disappointing if progress has to sit on the sidelines while we teach ourselves lessons that psychologists learned several decades ago.

Before leaving the Ultimatum Game, I want to return to the intuition that the observed behaviour is somehow tied up with resentment on the part of the responder at being made an unfair offer. Somehow, I now find myself on the

¹ Laslier and Walliser (1996) shows that much the same argument that Börgers and Sarin (1997) use to show that traditional reinforcement learning can be modeled using the replicator dynamics also applies to Roth and Erev's (1993) version of reinforcement learning.

wrong end of this debate too as a consequence of experimenters like Roth (Roth and Erev, 1995) switching from the fairness-norm camp to the learning camp, while theorists like Fudenberg and Levine (1998) have found their way to the exotic-preference camp through adopting learning models that apparently leave a responder in the Ultimatum Game with nothing to learn. Let me therefore offer my own trite reasons for why fairness and resentment matter in determining the basin of attraction in which subjects begin to play a bargaining game.

Kahneman and Tversky (1988) emphasise the importance of *framing*. When subjects enter a laboratory, I agree that cues offered in the way an experiment is framed trigger social norms that govern our behaviour in the games we play in real life. Such norms survive in real life because they coordinate our behaviour on *equilibria* of the games for which they are adapted. Since we almost never play a pure Ultimatum Game, we do not have a social norm adapted to its play. But we do have norms appropriate to games in which ultimata play a part. However, in these real-life games, it is seldom true that we will never interact again with our opponent or with one of the people observing how the game is played. A game-theoretic analysis will therefore treat such a real-life situation as a *repeated* game for which the folk theorem predicts that *many* equilibria will be available. As explained at great length in a recent book (Binmore, 1994, 1998), I believe that fairness norms evolved in the human species as an efficient solution to the equilibrium selection problem in such repeated games. In brief, they serve to *select* an equilibrium from a potentially infinite set of possible equilibria. But equilibria in repeated games also have to be *sustained* by the mechanism familiar from tit-for-tat that requires any deviations from unfair behaviour to be punished. This is why we get angry or resentful when treated unfairly.

This story so far accommodates both the fairness norm and the exotic preference explanations within a game-theoretic setting. Novices offer a fair amount because this is what their currently operative social norm recommends. Novices who are offered unfairly small amounts are programmed to feel resentful and so want to punish the proposer by refusing. But this behaviour *changes over time* as people dimly perceive that the norm they are using is not adapted to the problem with which they are faced. In the Ultimatum Game, people learn that it does not make much sense to get angry if offered too little, but the mavericks who initially make small offers learn much faster that it does not make sense to demand too much if one is nearly always refused.

The Ultimatum Game seems to be special only in having a structure that makes learning slow (Binmore *et al.*, 1993; Roth and Erev, 1995). The change in behaviour over time can be much more rapid in other bargaining games. The Nash Demand Game is a case in point. After being conditioned to coordinate on a variety of focal points, each group of subjects in the experiment of Binmore *et al.* (1993) had converged very closely to one of equilibria of the game after less than thirty trials. But different groups converged on different equilibria. When asked in a debriefing session what

was fair in the game they had just played, their median response was predicted very well by what actually happened in the game they played. I believe that such examples exhibit the evolution in miniature of a new fairness norm in the laboratory.

Such a breathless account is unlikely to convince anyone who has not previously had a chance to examine the data, and I offer it here only to protect myself when controversy arises in the future over who is responsible for abandoning considerations of fairness and resentment for the mechanical learning approach that now seems to have become the fashionable bandwagon. I was not to blame when backwards induction did not work—nor will I be to blame when it turns out that *both* learning and fairness need to be modelled when predicting bargaining data.

Conclusion

To return to the question with which we started: Would it not be better to leave laboratory experiments to psychologists who are trained to run them properly? Nobody doubts that we have a great deal to learn from psychologists about laboratory technique and learning theory, but recent history would nevertheless suggest that the answer is a resounding *no*. Our comparative advantage as economists is that we not only understand the formal statements of economic theory, but we are also sensitive to the economic environments and institutions within which the assumptions from which such statements are deduced are likely to be valid. Just as chemists know not to mix reagents in dirty test tubes, so we know that there is no point in testing economic propositions in circumstances to which they should not reasonably be expected to apply.

University College London

References

- Binmore, K. (1987). 'Modeling rational players, I and II.' *Economics and Philosophy*, vols. 3 and 4, pp. 179–214 and 9–55.
- Binmore, K. (1992). 'Foundations of game theory.' In (J.-J. Laffont, ed.) *Advances in Economic Theory: Sixth World Congress of the Econometric Society*, Cambridge: Cambridge University Press.
- Binmore, K. (1994). *Playing Fair: Game Theory and the Social Contract I*. Cambridge MA: MIT Press.
- Binmore, K. (1996). 'A note on backward induction.' *Games and Economic Behavior*, vol. 17, pp. 135–7.
- Binmore, K. (1997). 'Rationality and backward induction.' *Journal of Economic Methodology*, vol. 4, pp. 23–41.
- Binmore, K. (1998). *Just Playing: Game Theory and the Social Contract II*. Cambridge MA: MIT Press (forthcoming).
- Binmore, K., Gale, J. and Samuelson, L. (1995). 'Learning to be imperfect: the Ultimatum Game.' *Games and Economic Behavior*, vol. 8, pp. 56–90.
- Binmore, K., Shaked, A. and Sutton, J. (1985). 'Testing noncooperative game theory: a preliminary study.' *American Economic Review*, vol. 75, pp. 1178–80.
- Binmore, K., Swierzbinski, J., Hsu, S. and Proulx, C. (1993). 'Focal points and bargaining.' *International Journal of Game Theory*, vol. 22, pp. 381–409.
- Binmore, K., Swierzbinski, J. and Proulx, C. (1996). 'Does minimax work? an experimental study', ELSE Discussion paper, UCL.
- Bohm, P. (1993). 'Preference reversal, real-world lotteries, and lottery-interested subjects'. *Journal of Economic Behavior and Organization*, vol. 22, pp. 327–48.

- Bohm, P. (1994). 'Behaviour under uncertainty without preference reversal.' *Empirical Economics*, vol. 19, pp. 185–200.
- Bolton, G. (1991). 'A comparative model of bargaining: theory and evidence.' *American Economic Review*, vol. 81, pp. 1096–1136.
- Börgers, T. and Sarin, R. (1997). 'Learning through reinforcement and replicator dynamics.' *Journal of Economic Theory*, vol. 77, pp. 1–14.
- Brown, J. and Rosenthal, R. (1990). '"Testing the minimax hypothesis: a re-examination of O'Neill's game experiment."' *Econometrica*, vol. 58, pp. 1065–81.
- Brown and Rosenthal (1990) author to complete p. 7.
- Cubitt, R., Starmer, C. and Sugden, R. (1996). 'Dynamic choice and the common ration effect: an experimental investigation.' Discussion Paper, University of East Anglia.
- Estes, W. (1957). 'Of models and men.' *American Psychologist*, vol. 12, pp. 609–17.
- Fudenberg, D. and Levine, D. (1998). *Theories of Learning in Games*. Cambridge MA: MIT Press.
- Gigerenzer, G. (1996). 'On narrow norms and vague heuristics: a reply to Kahneman and Tversky.' *Psychological Review*, vol. 103, pp. 582–91.
- Güth, W., Schmittberger, R. and Schwarze, B. (1982). 'An experimental analysis of ultimatum bargaining.' *Journal of Behavior and Organization*, vol. 3, pp. 367–88.
- Holt, C. and Davis, D. (1993). *Experimental Economics*. Princeton NJ: Princeton University Press.
- Kahneman, D., Knetsch, J. and Thaler, R. (1986). 'Fairness and the assumptions of economics.' *Journal of Business*, vol. 59, pp. 258–300.
- Kahneman, D. and Tversky, A. (1988). 'Rational choice and the framing of decisions.' In *Decision Making*. Cambridge: Cambridge University Press.
- Laslier, J.-F. and Walliser, B. (1996). 'A behavioral learning process in games.' CERAS Research Paper, Paris.
- Loomes, G. (1993). 'Experimental tests of endowment and ambiguity effects.' ESRC Award Report.
- Milgram, S. (1975). *Obedience to Authority*. New York: Harper Colophon.
- Mookherjee, D. and Sopher, B. (1997). 'Learning and decision costs in experimental constant-sum games.' *Games and Economic Behavior*, vol. 19, pp. 97–132.
- O'Neill, B. (1987). 'Nonmetric test of the minimax theory of two-person zero-sum games.' *Proceedings of the National Academy of Sciences*, vol. 84, pp. 2106–9.
- Rapoport, A. and Boebel, R. (1989). 'Mixed strategies in strictly competitive games: a further test of the minimax hypothesis.' Working paper, University of North Carolina.
- Roth, A. and Erev, I. (1995). 'Learning in extensive-form games: experimental data and simple dynamic models in the medium term.' *Games and Economic Behavior*, vol. 8, pp. 164–212.
- Samuelson, L. (1994). 'Does evolution eliminate dominated strategies?' In (A. Kirman, K. Binmore and P. Tani, eds.) *The Frontiers of Game Theory*. Cambridge MA: MIT Press.
- Samuelson, L. (1997). *Evolutionary Games and Equilibrium Selection*. Cambridge MA: MIT Press.
- Selten, R. (1975). 'Reexamination of the perfectness concept for equilibrium points in extensive-games.' *International Journal of Game Theory*, vol. 4, pp. 25–55.
- Suppes, P. and Atkinson, R. (1960). *Applications of a Markov Model to Multiperson Interactions*. Stanford CA: Stanford University Press.
- Thaler, R. (1988). 'Anomalies: the ultimatum game.' *Journal of Economic Perspectives*, vol. 2, pp. 195–206.
- Von Neumann, J. (1928). 'Zur Theorie der Gesellschaftsspiele.' *Mathematische Annalen*, vol. 100, pp. 295–320.