What can we learn from experimental economics? Let me put my cards on the table. I am convinced that experiments have the potential to make a significant contribution to knowledge in economics. As I see it, experiments provide an exciting test-bed for economic theory, a vehicle for examining whether the predictions of theory hold; and where they do not, the results of experiments can often provide useful data from which to develop new theories. As a practising experimentalist, it would be extremely tempting to use this space to promote my art by presenting examples of ‘success’ stories in experimental economics. For example, there are reassuring bed-time stories that economists can tell their children at night about a happy on-going relationship between economic theory and experimentation. Perhaps the most obvious choice would be the remarkable success of equilibrium predictions in ‘market experiments’. While it would be misleading to suggest that every prediction of economic theory has worked in simulated market environments, they often do, and successes have been particularly striking in cases where there is, as yet, no well-developed theory to explain how agents get to the predicted equilibria. I could tell this story, but I will not: besides, it has already been told by others rather better placed to tell it well. I intend to focus on one area of experimental research where there are some clear and persistent conflicts between theory and evidence. The area concerned is the (descriptive) theory of choice under risk. In my view the theory in this area has advanced during the last two decades, at least in part as a consequence of experimentation. At the same time, experimental evidence seems to challenge some of the most basic assumptions of economic theorising. A good part of my concern here is to ask what we should make of this state of affairs.

Like it or not, economists have a bad reputation for being relatively unmoved by facts about the world, at least ones that do not fit conventional theoretical presuppositions. If the reputation is well deserved we should make efforts to mend our ways; if it is not, then we should be at pains to demonstrate the error of that view. Either way, a serious engagement with experimental evidence seems warranted. To pat ourselves on the back by focusing on the areas of correspondence between theory and evidence, brushing aside more
troublesome features of the data, would be a dubious response akin to a research chemist dismissing persistent laboratory anomalies on the grounds that the Bunsen burner still works fine. On the other hand, in some cases where theory and (experimental) evidence conflict, there may be sound reasons for rejecting the evidence. For instance, if experimental subjects are not sufficiently well-motivated, or if they do not understand the experimental scenario, their behaviour may be a product of carelessness or confusion and hence, irrelevant to an assessment of economic theory. As there will always be room for argument over whether subjects understood a given experiment, or cared enough about the experimental rewards and so on, it is impossible to be sure that the results of any particular experiment are reliable. But a crucial feature of the experimental paradigm is that it offers the possibility of replication. If interesting results can be reproduced in independent trials, if they can be shown to be robust to variants in experimental design which, say, vary incentive levels and so on, then we may have the basis for more confidence. Many experimentalists, I take it, would agree that while you cannot infer much with confidence from a single experiment, you can learn valuable things from a programme of experimental research. In my view, we have been learning facts about the world by undertaking experimental research but I have begun to wonder whether economists really want to know at least some of these facts. I think the point can best be made by example.

1. A Brief History of An Experimental Research Programme

Experimental methods struck a rich seam of applications in decision making under risk. The project was stimulated by some now famous counter examples to expected utility theory (EUT) invented by Maurice Allais (1953). The original Allais examples invited people to respond to hypothetical choices between simple gambles involving large sums of money. The first experiments using these examples seemed to indicate that individuals’ choices systematically violate the independence axiom of EUT. At first, data accumulated slowly, but by the end of the 1970s a growing body of evidence had prompted a number of theorists to develop alternative models, and by the mid 1980s a wide variety of such theories had been developed. Since these new theories had been designed to explain what was known about EUT violations, fresh experiments were needed to assess their comparative performance and so began a new phase of experimentation aimed at testing competing theories. The general conclusion to emerge from this phase of the programme was that some theoretical approaches showed promise, but each had limitations too. Consequently, the search for a better theory continued with some researchers

3 See Allais and Hagen (1979).
4 See Machina (1987) and Weber and Camerer (1987) for reviews of the theory and evidence for the relevant period.
5 Examples in this vein include Camerer (1989) and Battalio et al. (1990).
investigating possible refinements of existing approaches while others suggested new theoretical strategies. Further experimental evidence bearing on this second generation of theories is currently emerging.

Such a programme of research looks like a prime example of good scientific practice in operation: an on-going dialogue between theory and evidence has been unfolding in which proposed theories confront data from laboratory settings specifically designed to test them; where the theories fail, efforts are made to refine or replace them with better alternatives. Moreover, judged against one widely accepted test of whether a scientific discourse is progressing – that is, whether it has prompted the discovery of novel empirical phenomena – the research programme can claim at least some success. One example relates to the testing of regret theory developed by Bell (1982), Fishburn (1982) and Loomes and Sugden (1982). Regret theory implies that choices may be systematically non-transitive in some quite specific choice contexts and violate monotonicity in others. Since both properties are commonly regarded as desiderata for any satisfactory theory, these were bold predictions. No doubt many economists would have expected them to be proven false, yet subsequent experiments appeared to confirm these and other predictions of regret theory.

Even so, a critic might question whether significant progress has been made in the hunt for a superior predictive theory by pointing to other features of the programme. For example, it remains the case that no single theory comes close to organising the experimental data and EUT remains industry standard for theorists and applied economists working outside of the specialist literature on risky decision making. Moreover, basic puzzles which motivated the development of new theories, remain unexplained. An epilogue to the story about regret theory is illustrative of this last point. Regret theory generated a series of novel predictions which turned out to be supported by some of the subsequent experimental evidence. Good news it seems, but here is the rub: further testing suggests that regret theory is not the correct explanation for the new phenomena whose discovery it prompted. But, since these new phenomena run counter to every theory so far proposed by economists, the net effect of this branch of research has been to increase the number of as yet unanswered questions. It now seems clear that the theories developed so far are, at best, only partial explanations of the data and there is much going on in actual choice behaviour which we do not yet understand.

6 This has been identified as an important criterion of theory appraisal in a number of methodological approaches, particularly those in the Popperian tradition. See for example Lakatos (1978).
7 Monotonicity implies that if an individual is offered a choice between two gambles, where one stochastically dominates the other, the dominated option will not be chosen.
9 Experiments reported by Starmer and Sugden (1997) suggest that regret theory is probably not the correct explanation for the form of intransitivity which it prompted the discovery of. Similarly, results reported by Starmer and Sugden (1993) and Humphrey (1995) appear to undermine the regret theory explanation for violations of monotonicity.

© Royal Economic Society 1999
One possible reaction to this would be to urge researchers to extend the search for better theories in the hope that some super-theory will emerge with more impressive predictive ability. Another quite different response would be to conclude that the programme as a whole has reached a blind alley. If twenty years of quite intensive research effort has failed to produce a theory with demonstrably improved predictive content, perhaps we should accept that EUT is as good a theory as we are likely to find and devote our attention to new and potentially more fruitful questions? If the latter conclusion seems tempting, however, consider this. One thing we have learned for sure as a consequence of this programme of research is that EUT is descriptively false. Mountains of experimental evidence reveal systematic (i.e., predictable, not random) violations of the axioms of EUT, and the more we look, the more we find. This is not good news for the general economist, but there it is.

2. Does the Evidence Matter (Much)?

It is hard to deny that EUT is a key building block of economic theory, so to the extent that we wish to interpret it, or economic theories which rely on it, as predictive theories, the evidence looks worrying. Nevertheless, I can think of at least two sorts of argument which might be used to down-play its significance. One would be to suggest that although the theory is violated, EUT provides a close enough approximation to actual choices for it to be reliable in practical applications. Another would be to suggest that violations of the theory may be artifacts of the laboratory environment which might not generalise to economically meaningful settings.

What case could be made for the view that, despite evidence of violation, EUT is a close enough approximation to actual behaviour for practical purposes? One way to make the case would be to point to evidence of its predictive success in practical applications. If there were some broad-based body of evidence which seemed to show that, on the whole the theory worked quite well in say, applied contexts, that would be reassuring. As far as I can see, however, there is no such corpus: the evidence bearing directly on the predictive efficacy of EUT derives largely from experimental testing and that evidence often points to high failure rates with anything from 25 to 50% of choices violating EUT. But perhaps the argument could be made another way. If violations of EUT were empirically important phenomenon, should we not then expect to observe anomalies regularly, in everyday market behaviour, which parallel laboratory findings? Take monotonicity for example. Experiments have revealed violations of the principle, but who has ever seen corresponding behaviour in real markets like, say, people buying tickets for the National Lottery then subsequently swapping them for tickets with poorer odds of winning? Since we see no evidence of any such transactions, isn’t monotonicity a reasonably safe assumption? Cunning as the argument is, it is not compelling. To see why, consider this evidence produced by Tversky and Kahneman (1986). They presented subjects with the following choice problem:

© Royal Economic Society 1999
It is very easy to see that option $B$ dominates option $A$ since, for every colour, the prize for option $B$ is always at least as good as the prize for option $A$ and in some cases it is better. Kahneman and Tversky presented this problem to 88 subjects and found that all of them chose $B$. Now consider this slightly modified version of the above problems:

Consider the following pair of lotteries, described by the percentage of marbles of different colors in each box and the amount of money you win or lose depending on the color of a randomly drawn marble. Which lottery do you prefer?

<table>
<thead>
<tr>
<th>Option $A$</th>
<th>Option $B$</th>
<th>Option $C$</th>
<th>Option $D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>90% white</td>
<td>6% red</td>
<td>1% green</td>
<td>1% blue</td>
</tr>
<tr>
<td>$0</td>
<td>win $45</td>
<td>win $30</td>
<td>lose $15</td>
</tr>
<tr>
<td>90% white</td>
<td>6% red</td>
<td>1% green</td>
<td>1% blue</td>
</tr>
<tr>
<td>$0</td>
<td>win $45</td>
<td>win $45</td>
<td>lose $10</td>
</tr>
<tr>
<td>90% white</td>
<td>6% red</td>
<td>1% green</td>
<td>3% yellow</td>
</tr>
<tr>
<td>$0</td>
<td>win $45</td>
<td>win $30</td>
<td>lose $15</td>
</tr>
<tr>
<td>90% white</td>
<td>7% red</td>
<td>1% green</td>
<td>2% yellow</td>
</tr>
<tr>
<td>$0</td>
<td>win $45</td>
<td>win $10</td>
<td>lose $15</td>
</tr>
</tbody>
</table>

Fig. 1

It is very easy to see that option $B$ dominates option $A$ since, for every colour, the prize for option $B$ is always at least as good as the prize for option $A$ and in some cases it is better. Kahneman and Tversky presented this problem to 88 subjects and found that all of them chose $B$. Now consider this slightly modified version of the above problems:

Fig. 2

Options $C$ and $D$ are stochastically equivalent to $A$ and $B$ respectively, the only difference being a minor change in the presentation which ‘simplifies’ the options by assigning each prize to a single colour. This modification, however, also makes it more difficult to detect the fact that the first option is dominated by the second and, when this choice problem was presented to another group of 124 subjects, the majority (58%) chose the dominated option $C$.

This evidence supports two conclusions: first, individuals are not so stupid that they will choose options which are *patently* dominated by other alternatives in their choice set; on the other hand, they do not have *generally* monotonic preferences and may violate monotonicity in circumstances *where it is not obvious that they are doing so*. Here, then, is the catch. If monotonicity only tends to fail in cases where it is difficult spot the violation, we have reason to doubt
whether those cases would be obvious to the casual eye of the general economi
tist. Moreover, in the absence of theories which explain, and hence allow us to predict, violations of principles like monotonicity we can hardly expect to know where to look for such anomalies in market contexts (if they occur). Consequently, to conclude that EUT violations are probably rare, on the grounds that contrary evidence does not present itself in an obvious form, in the market place, seems little more than arm-waving.

If ‘reasonable approximation’ arguments are unpersuasive, what basis might there be for thinking that behaviour in the laboratory is artificial in some sense? There are manifold reasons why a given experiment might generate results which are unrepresentative of ‘normal’ market behaviour. These include considerations already mentioned such as whether subjects were appropriately motivated, whether they understood the instructions, the reward mechanisms and so on. Issues of this sort raise specific concerns about particular experiments which can be explored through further experimentation. In fact, very large literatures have developed in precisely this way. One such literature relates to the so-called preference reversal phenomenon.

3. Preference Reversal: Awkward Facts from Experimental Labs

The preference reversal phenomenon – first reported by psychologists Lichtenstein and Slovic (1971) and Lindman (1971) – has typically been observed in experiments involving two types of task. The first requires the subject to choose between two gambles: one (often called the ‘$-bet’) offers a small chance of winning a ‘good’ prize; the other (the ‘P-bet’) offers a larger chance of winning a smaller prize. The second task requires the subject to assign monetary values, typically reservation prices, to each gamble. In terms of standard theory, the two tasks amount to different ways of eliciting the preference ordering of the P- and $-bets. As such, both tasks should produce the same ranking of options for a given individual, yet repeated studies have found this apparently puzzling result: those who choose the P-bet in the choice task often place higher valuations on the $-bet.

In the first study of the phenomenon by economists, Grether and Plott (1979) argued that preference reversals might be explained by weaknesses in the designs of the psychologists’ experiments. In their report, they are disarmingly frank about their objectives: their aim was to ‘discredit the psychologists’ works as applied to economics’ (p. 623). They designed new experiments in the expectation that, once appropriate controls were implemented, the phenomenon would evaporate. It did not, and the authors confess some surprise at being forced to conclude that ‘Taken at face value the data are simply inconsistent with preference theory and have broad implications about research priorities within economics . . . It suggests that no optimisation principles of any sort lie behind even the simplest of human choices’ (ibid).

10 A similar argument could be made in relation to other principles such as transitivity.

© Royal Economic Society 1999
Subsequent researchers still remained sceptical, however, and sought to check the result further by introducing, for example, higher incentives and other controls designed to reduce the possibility of subject error. While some studies along these lines (e.g., Pommerehne et al., 1982; Reilly, 1982) reported reductions in the frequency of reversals, the phenomenon persisted.

Preference reversal provides an illustration of the way in which experimental/theoretical research programmes evolve. When confronted with observations contrary to basic theoretical assumptions, economists have quite rightly argued that apparent failures of theory might be explained in terms of other hypotheses far less, if at all, challenging to standard theory. Often these arguments have implied potential weaknesses in particular experimental designs. Plausible as many of these conjectures were, they did not undermine the experimental approach, but rather suggested new lines of experimental enquiry which were subsequently pursued. It now seems hard to deny that preference reversals are robust: the phenomenon has been repeatedly observed in this type of experiment, and has persisted in the face of sustained attempts to make it go away.

I suspect that preference reversal constitutes one of those facts which economists do not really want to know. In two reviews of the preference reversal literature, Tversky and Thaler (1990) and Hausman (1992) have concluded that it cannot be adequately explained in terms of any model of complete and stable preferences. If they are right, then in spite of much effort to demonstrate to the contrary, it seems that if we are to explain it at all, the explanation will require us to give up one of the most fundamental presuppositions of economic theory. There is a broader context here. It is noteworthy that the psychology literature has generated a number of hypotheses which could explain preference reversal and several commentators (Hausman, 1992; Tversky and Thaler, 1990; and Camerer, 1995) have argued that economists should pay more attention to these theories. But as Hausman notes, economists have shown little enthusiasm for models which depart from the traditional explanatory mould. Historically a number of rather famous voices, including Simon (1957) and Selten (1990), have urged economists to take account of psychological theories of choice. Since there is evidence which is much more readily explained in terms of psychological hypotheses – preference reversal is merely one of many examples which could be cited – what explains the reluctance of economists to pursue these ideas. Is this evidence of a blunt refusal to face the facts?

It might be argued that even if preference reversal remains a puzzle, it has little more than curiosity value. That, however, is hard to square with the attention it has received from economists: as Tammi (1997) points out, many of those who have devoted significant time and energy to finding a resolution consistent with standard economic theory have explicitly acknowledged its importance. To down-play its significance, for no better reason than it being difficult to resolve would seem an unduly cynical reaction. But, there may be more legitimate reasons for thinking that preference reversals, and evidence of anomalies more generally, have relatively marginal significance for economists. It could be argued that there are, in fact, theoretical reasons for expecting
economic theories to work better (in terms of predicting actual behaviour) in real market environments, than they will in certain types of experimental settings. Let me sketch one such argument.

4. The Evolution Argument

Economic theories assume optimising behaviour on the part of individual agents. But identifying optimal solutions, even in relatively simple settings, can often seem quite demanding both analytically and computationally. Given this, is the assumption that agents behave as if optimising, in anything but the simplest of settings, at all plausible? One way to defend the use of optimising assumptions in accounts of market behaviour is to argue that optimal behaviour might evolve in some relevant settings. Hence, although agents facing new decision contexts may fail to optimise, if similar decisions are faced repeatedly, if there is feedback on the consequences of actions which can facilitate learning and sufficiently high incentives to promote such learning, actual behaviour might gravitate toward optimal solutions over time.

The flip side of this argument implies that economic theories will work less well in environments where agents either make one-off choices in unfamiliar environments, or make repeated choices in the absence of appropriate feedback and incentives. Now, it is typical of experiments in decision making under risk that the subjects are required to make a series of one-off choices in quite unfamiliar contexts and typically, the only feedback about the ‘success’ of choice strategies comes in the form of a payoff to a single task, randomly selected at the end of such experiments. Could it be that failures of theories like EUT observed in laboratory contexts reflect essentially ‘naive’ behaviour which is unrepresentative of ‘evolved’ market behaviour? While this line of argument suggests an interesting possibility, it falls far short of any justification for being dismissive of the evidence.

The claim that EUT (and other economic theories) are only intended to apply in appropriately evolved environments – though no doubt held by some – would be a controversial claim and one which would, perhaps quite severely, limit the scope of economic theorising. Moreover, Ken Binmore’s plea that we should confine applications of economics to those areas where it can be reasonably expected to work, strikes me as bordering on circular: if only those phenomena which, a priori, we expect to be consistent with standard notions of optimisation are allowed to count as economics, and anything else is sociology or psychology, successful prediction is not much to write home about. Others, I take it, would wish to interpret economic theory more generally. Indeed, as a general methodological principle there is surely a good case for interpreting theories broadly until there is a reason to conclude otherwise. It is possible that economic theories do work better in evolved environments. But that is best regarded as an empirical question and it would seem rather dubious practice to simply assume it as part of a defence strategy designed to immunise theories against dis-confirming evidence. Traditionally, economists have aspired to finding general theories of behaviour and against this backdrop, the
discovery of evidence indicating failure in certain domains of interest is surely to discover a limitation. After all, many economically relevant choices are made relatively infrequently with little potential for feedback and learning (e.g. major investment decisions). The same is true of many of the most important life choices facing individuals (e.g. decisions with respect to education, employment, health, house buying and so on). Unless we are inclined to regard all such choices as economically irrelevant, there is a case for being interested in theories which apply to infrequent decisions and the evidence which bears directly upon them.

That said, there is evidence that the predictions of economic theory sometimes work better in repeated choice contexts where feedback mechanisms are present. The preference reversal literature again provides one example;\textsuperscript{11} the literature from market experiments provides another.\textsuperscript{12} But in my view, it would be quite wrong to use such evidence as a stick with which to beat the data from experimental settings which focus on one-off choices. If there are differences between one-off and long-run decisions, it seems natural to ask why. That is an interesting research question and one which, incidentally, can be usefully investigated by experimental methods.\textsuperscript{13} The results of market experiments show that sometimes experimental markets converge to predicted outcomes, and sometimes they do not. Part of the difficulty in understanding why such differences exist between alternative market institutions derives from the fact that we have a relatively underdeveloped understanding of what determines initial decisions and the dynamics of adjustment. It could be that the development of satisfactory theories capable of explaining these facets of behaviour will elude us for some time to come. It could even be that short run behaviour is too ephemeral to ever predict with any degree of confidence, but let us not kid ourselves by pretending that these issues are irrelevant for questions of general interest to the economics profession.

5. Conclusions

In my view, experimental economics provides one useful arena in which to pursue an ongoing dialogue between theory and evidence. I have focused on just one programme of research concerned with the development of better theories of choice under risk and I think it fair to say that interesting developments have emerged. This programme has stimulated a stream of novel

\textsuperscript{11} For example, Chu and Chu (1990) provide experimental results which show that the incidence of preference reversal can be reduced if not eliminated by exposing subjects who have committed a reversal to arbitrage mechanisms designed to exploit the inconsistency. This and other related evidence is reviewed by Camerer (1995) and Tammi (1997).

\textsuperscript{12} See for example the review by Smith (1989, p166) in which he argues that `The economist’s maximising paradigm often performs well in predicting the equilibrium reached over time in experimental markets, but this theory is not generally able to account for short-run dynamic behaviour, such as the contract price paths from initial states to final steady states.’

\textsuperscript{13} Of course there are limits to just how long a run can be simulated in laboratory conditions, but existing evidence suggests that behaviour can change substantially, and in some cases converge on equilibrium predictions, well within such constraints.

© Royal Economic Society 1999
theoretical ideas and generated a much richer body of evidence for assessing and developing new theories. As such, the dialogue would seem to be healthy and progressive. On the other hand, there is not much evidence of this work currently percolating beyond the specialist literature. I do not wish to imply any criticism of the broader profession on this score: it seems entirely reasonable to adopt the position that, until a demonstrably superior alternative emerges, EUT is the best game in town. My aim has been to argue that, even if it is the best game in town, there are worrying questions about its empirical validity. My arguments are primarily aimed at those who would seek to dismiss or down-play the significance of evidence against theories like EUT by questioning the reliability or relevance of experimental data.

Economists have displayed a good deal of enthusiasm for engaging with at least some phenomena — like preference reversal — which challenge basic assumptions of economic theory. But while much effort has been directed towards finding explanations consistent with conventional theory, there has been little engagement with alternative modes of explanation, even when conventional routes appear to have failed. Admitting failure in such cases would be an honourable position — a recognition that there are phenomena which, at least currently, defy explanation of the conventional variety — but beating the data, or the methods which have generated it, until they become silent, would not.

I think there are genuine questions about whether laboratory anomalies, like those observed in experiments discussed above, will prove to be very important in the long run development of economic theory. They may not. It could turn out that, for most practical purposes, they can be ignored. We may discover a quite general tendency for economic theories to work better in real market environments than they do in the laboratory. If so, perhaps laboratory experimentation may one day be a thing of the past. I don’t know. But then that is just the point. I don’t know and I don’t believe any of us have a sound basis for making definite judgements about such issues at this stage. So let’s experiment and see what we can find out.

University of East Anglia

References


© Royal Economic Society 1999


© Royal Economic Society 1999