A REALITY CHECK FOR GAME THEORY

David J. Butler
University of Western Australia

Colin F. Camerer (2003). **Behavioral Game Theory: Experiments in Strategic Interaction.** Princeton University Press: Princeton, New Jersey, p. 544. £48.95/ $75.00. ISBN 0-691-09039-4.

There is a tendency for some economists who are less familiar with game theory (GT) to comment that it has never fulfilled its early promise and that it essentially has little to offer as a tool to understand, predict and guide human, social and economic behaviour. For instance in Sylvia Nasar's book '*A Beautiful Mind*', she recounts the remarkable controversy that surrounded the award of a nobel prize for GT, including one committee member apparently asking for 'a single major example that game theory had any empirical validity whatsoever' (p. 371). An excellent reply to such sceptics would be to present them with a copy of Colin Camerer's new book, '*Behavioral Game Theory*'. This is not to say that his book catalogues an unbroken string of successes for the theory, as we shall see neither GT nor human behaviour emerges unscathed from the body of work reported therein.

Camerer's book is an up-to-date and comprehensive look at a relatively new but fast-growing area of economics concerned with using experimental methods to investigate strategic interaction. Grouped into nine chapters, he summarizes several hundreds of studies on different aspects of GT, from mixed-strategy equilibrium to dominance-solvable games to signalling games, carefully synthesizing what has been learned and frequently drawing attention to what has still to be discovered. One virtue of this book is simply that it has brought into one volume the remarkable breadth and depth of research that has gone on in this area. Even those who, like myself, thought we were keeping at least one eye on this literature may discover entire bodies of work that we were unaware of.

There are naturally limits to its scope; in the preface, he mentions some omissions such as cooperative games, and deliberately skimpy treatment, e.g. public goods games and auctions. However, he also tends to avoid related experiments published in the psychology and biology journals. This is a reasonable decision, given that boundaries must be drawn somewhere; however, I will later question whether one or two of the claims he makes are as a consequence less justified than they might be.

Before we look at some of the many interesting experiments and findings reported in the chapters of this book, I will begin with a discussion of some general themes and issues that I think deserve extra attention. I follow that

0950-0804/05/01 0137–11 JOURNAL OF ECONOMIC SURVEYS Vol. 19, No. 1

section with a selection of studies that interested me, as well as his occasional errors, in the order of the chapters in the text, along with some other observations. Due to the large number of different games discussed, I will mostly assume the reader has basic familiarity with at least the more well-known ones.

1. Broad Issues

The first and perhaps most fundamental issue we need to confront is how should we interpret and use the experimental data generated by the behavioural GT project? For example, can we conclude that a game-theoretic concept is refuted if data drawn from an experiment designed specifically to showcase its predictive power find no such evidence? Or does it refute the rationality assumption for the human behaviour that generated the data? In chapter 1, Camerer goes to some trouble to explain why the uncovering of empirical regularities in human strategic interaction is needed to inform the future development of GT. What should we make of this use of the data and the interpretations on which it rests?

Many economists see GT as simply a set of answers to mathematical questions and such answers can be neither disproved nor improved upon by observing the behaviour of student subjects in some experiment. But this view surely misses the point. Although it is true no abstract mathematical object can be refuted by any experiment, economics is not interested in any particular mathematical object for its own sake; they are a part of our subject matter only if they can assist us in our projects. Hence, the 'necessary truths' defence of GT does not explain why we in economics should make any more use of those truths than of other mathematical truths that have not been co-opted into economics. In the light of this, we might instead ask: can experimental data lead us to conclude that some game-theoretic concept does not have the relevance, or context-independent predictive power, for the study of human strategic interactions that economists had previously ascribed to it? Can such data guide us in seeking out more useful mathematical concepts rather as biologists look to nature when modelling biological facts?

One line of defence for the continued use of empirically dubious game-theoretic concepts is that we may believe GT describes a benchmark of optimal play, even if human behaviour often falls short of the ideal. In these cases, we may be able to use the concepts in a normative and prescriptive sense, to improve our performance. An example of this might be a constant-sum game such as tennis, in which an informed coach could use insights drawn from mixed-strategy equilibrium to raise the win rate of the player who employs him. Even in these cases it is important to remember that if others are not playing optimally we can often do better than our notional equilibrium strategy by exploiting their suboptimality. This point is vividly illustrated in the beauty contest game, as described in chapter 5.

Alternatively, we may not accept that the empirically flawed concepts have normative validity, choosing instead to stand by our actions. The convention in nearly all of traditional economics is to assume that people are both rational and self-interested. Applied to GT, this means assuming players have unlimited reasoning powers and utilities that are a simple function of their own dollar

payoffs. Using these assumptions for human decision-making, experimentalists apply the concepts of GT to predict the outcomes of human strategic interactions. But if the perfect rationality and self-interest assumptions are inappropriate, perhaps we can justify our violation of the game-theoretic predictions.

For instance, in the ultimatum game, the prediction of GT coupled with these two behavioural assumptions alone is that responders will accept any positive share of the pie from the proposer, rather than reject the offer and have both parties receive nothing. Or, in the beauty contest game (to be explained later), the prediction of GT is that players will reason their way to the unique subgame perfect equilibrium, even though many levels of reasoning are required of players to support this prediction. As Camerer explains in chapters 1, 2 and elsewhere, these predictions have repeatedly failed in experiments. This is surprising in the sense that these assumptions regarding human behaviour have been very successful in most other areas of economics, going all the way back to Adam Smith's remarks on the irrelevance of the benevolence of the butcher and the baker for the workings of the invisible hand.

Should we then place the blame on the assumptions of perfect rationality and selfish preferences, while arguing GT's concepts are still necessarily true? The difficulty with this approach is that it does not help us decide how to apply GT to human behaviour. We seem to be saying that the concepts (coupled with the standard auxiliary assumptions) are very useful descriptively and normatively, except when they are not, in which cases we need different auxiliary assumptions. But when a theory is demonstrated not to predict in a context-independent way, its usefulness diminishes, a fate already befallen Expected-Utility Theory.

A better response is to acknowledge that complete context independence is unrealistic, but try to extend the domain of GT by incorporating into it a small number of critical extra factors suggested by the data. Suppose that, in some contexts, we suspect that our predictions are failing because players care about the intentions they ascribe to the other players. Psychological game theory (PGT) was developed (Geanakoplos *et al*., 1989) for such a case, as it allows utilities to depend not just on money but also on a player's beliefs regarding the intentions of other players. In equilibrium, these beliefs are correct, and a Psychological Nash Equilibrium (PNE) will exist. Camerer discusses PGT in chapter 2.

A development such as PGT not only acknowledges that behaviour is not invariant across contexts where beliefs about intentions do and do not matter, but also builds upon the framework of GT by adding the tools to predict or explain outcomes in those contexts where intentions matter. GT's predictions then become consistent with known empirical regularities (summarized in chapter 2), and its new concepts, such as PNE, can be a source of insight and understanding regarding human strategic behaviour and its outcomes. Influential papers that have built upon this framework include Rabin (1993) and, more recently, Dufwenberg and Kirchsteiger (2004); Camerer discusses these at length in pp. 105–110.

An implication of this type of extension to GT is that two games with identical payoffs from the standard perspective may lead to different behaviour if one and

only one game leads players to care about intentions. For example, one-shot ultimatum games and prisoner's dilemma games may cause players to transform utilities as in PGT when the other party is believed to be another person. But, if told in advance that the other party is a computer making preprogrammed choices, we may now disregard the intentions of that player and behave much as standard GT would predict. This is consistent with experimental findings, including a few experiments which use fMRI scans of player's brains, as they make their choices under these different conditions, which find significant differences in the activity levels of brain regions concerned with social interactions (e.g. McCabe *et al.*, 2001; Rilling *et al.*, 2002).

An advantage of allowing data from experiments to drive theoretical advances in GT is that it makes GT much more empirical than has previously been the case, at least in economics. Such a change of direction has been advocated by other researchers as the only real way forward if GT is to provide new understanding about our social world (e.g. Sugden, 2001). Camerer is clearly also in this camp.

A second area of controversy concerns the role of natural selection in generating the behavioural deviations from GT. Camerer excludes experimental games reported in the biology journals (such as *Evolution and Human Behaviour*) from his book and has some moderately hostile things to say about the usefulness of looking to evolution for explanations of data or new hypotheses regarding human strategic interactions (p. 116). I find his claims too sweeping; indeed, one could say the idea players may exhibit concerns for reciprocity and fairness in certain games rather than selfishness is itself a prediction from evolutionary rather than economic theory. One interesting study from the biology literature that draws on kin selection (Segal and Hershberger, 1999) predicted that rates of cooperation in prisoner's dilemma games played between non-identical twins would be lower than that for identical twins. The study found strong support for this hypothesis, suggesting evolutionary explanations do have a role to play here.

Part of the difficulty of using evolutionary logic is that economists differ in what they think evolutionary arguments do and do not predict. Camerer notes (p. 68) that, in dictator and ultimatum games, young children exhibit self-interested behaviour, becoming fair-minded only as they grow older. He argues that this finding is crucial for implying that fairness norms are not innate. But it is not clear that this shows any such thing: it is well known from the evolution of life histories that particular genes switch on and/or switch off at different phases of human development. For example, language (Pinker, 1994), puberty, etc. indeed the entire developmental process is itself under genetic control. More generally, a key adaptive feature of humans is their behavioural flexibility in the face of a rapidly changing social and natural environment. When behaviour varies across cultures, the variation often itself has an adaptive function; numerous examples are given in Barrett *et al.* (2002).

Before we leave this controversial topic, recent research (Brosnan and DeWaal, 2003) found that females of another cooperative primate, the brown capuchin monkey, exhibit social emotions similar to human responders in an ultimatum game in analogous circumstances. Their work suggests a sense of fairness might

be selected for among intelligent social species rather than being a unique product of human culture. Culture may shape the raw material we are already endowed with, but it is not by itself the source of our sense of fairness.

A third issue is the growing use in this literature of quantal response equilibrium, or QRE. While Nash Equilibrium assumes a player selects the best response from among his options with certainty, QRE allows a degree of noise into the process. If, given their beliefs over other player's strategies, options with greater expected values are selected with a larger probability, we have a statistical equivalent of Nash Equilibrium called QRE. A parameter λ in a logit response function captures a player's sensitivity to differences in the expected values of her alternatives. If $\lambda = 0$, she cares nothing for differences in the expected values of her options and chooses randomly, and if λ is large, her behaviour converges back on the Nash Equilibrium.

Camerer introduces QRE in a brief but useful appendix to chapter 1 on the rudiments of GT and makes reference to it in subsequent chapters. Given that it is now well known from tests of expected utility theory that human players make errors in choices, QRE is obviously an important development for experimental GT. But is it fair to conclude (as on p. 239) that error rather than other causes such as cooperativeness can mop up behavioural deviations from GT, allowing the estimated λ to vary dramatically in magnitude so as to achieve the best fit with the data (in chapter 3 alone, estimates of λ vary from 0.248 to 3.24)?

Referring back to the experiments where subjects played prisoner's dilemma games against both humans and computers, is it reasonable to say our error rate rises when we know we are playing people or is it not really error but social utility at work? Kiyonari *et al*. (2000) in their study of one-shot dilemma games varied both the financial incentives of subject's decisions and the degree to which a social exchange heuristic would be triggered. They found that rates of cooperation rise as the incentives are increased and fall as the social context (and incentives) is reduced. They suggest that choice patterns such as these indicate the error of cooperation is the result of an adaptive heuristic. If such non-Nash behaviour in one-shot games is simply error (see for example Goeree and Holt, 2001), that error behaves most counter-intuitively: increasing when the degree of realism and the magnitude of the consequences rise.

Perhaps an error term can be combined with some social preferences theory to jointly explain the data? More fundamentally, is the concept of error used in QRE correct? Continuing the parallel to the experimental literature on expected utility theory, the λ in QRE is akin to the Hey and Orme (1994), or Fechner concept of error, which is not the only possibility. Other approaches based on random and/or hazy preferences are also possible (e.g. Loomes and Sugden, 1995) but have not been incorporated into behavioural GT so far.

2. Chapter Highlights

Camerer opens chapter 1 by listing more than a dozen seemingly distinct real-world problems united by the power of GT to explain and predict their outcomes.

He points out that, in this task, GT is often right but also often wrong. It goes wrong because too much theorizing has been based on introspection and assumption rather than observation, an imbalance of theory and facts he hopes to help redress. One of his examples is of ultimatum bargaining, which we saw earlier. He claims negative reciprocity lies behind a responder's rejection of unfair offers even at substantial cost to themselves, drawing an intriguing analogy to jilted boyfriends harassing former girlfriends.

On the question of whether this behaviour results from cultural standards or from evolution, not surprisingly he prefers the cultural standards story. However, evidence from psychobiology (e.g. Panksepp, 1998) shows that the rage system in the mammalian brain is closely linked to centres in the cortex that anticipate rewards. If an anticipated reward is not forthcoming, anger or aggression follows swiftly. Given these primitive connections, perhaps even reciprocity seems too cognitively advanced a cause for responder actions.

After a helpful appendix to chapter 1 on the rudiments of GT, a second appendix discusses experimental methodology. This is a more controversial subject as nearly everyone has their own ideas on good and bad practice; however, Camerer's views here are probably as close to the mainstream as any. One amusing example of poor reasoning is given in the section on incentives in this appendix. He relates (p. 40) an anecdote from the reality TV finale of a series of Survivor. One defeated contestant had to decide which of the two finalists he should vote for, and asked them both to choose an integer from 1 to 9, his vote going to the player choosing closest to his own (secret) number. The first player guessed 7, to which the second player responded with a guess of 3, an option clearly dominated by 6. As the prize at stake was $1 million, in expected value terms this was an extraordinarily costly mistake, suggesting powerful incentives do not necessarily improve decision quality.

Chapter 2 focuses on the huge literature on dictator, ultimatum and trust games, about which I have already commented. Camerer also discusses some theories of social preference that try to account for these data, such as Rabin (1993) on p. 106. It is interesting to note that the transformations of payoffs involved in Rabin's model do not conform to those apparently reported by subjects in the Kiyonari *et al*. (2000) study. Whether this is due to flaws in their methodology or weaknesses in Rabin's conception of fairness is a question worthy of investigation. A couple of errors also creep into chapter 2, the most prominent on p. 45 is his formalization of the structure of public goods games. His equations in the main text are inconsistent with those in the footnote on the same page; worse, they are both wrong.

Chapter 3 looks at the evidence on the often counter-intuitive predictions of mixed strategy equilibria, which turns out to be surprisingly supportive. Several experiments looked at versions of hotelling's location game, with two, three and four firms selecting a point along a line [0, 100] at which they should locate. Support was found for the well-known 2-player equilibrium of side-by-side at 50, but more interestingly, also for the less well-known 3-player equilibrium of randomization over the interval [25, 75], with no choices below 25 or above 75.

Although the data did not match the prediction perfectly (p. 144), it is surprisingly close (Collins and Sherstyuk, 2000). The 4-player equilibrium predicts two clusters of firms, one at 25 and the other at 75. Hück *et al.* (2002) found reasonable support for this prediction, although a disequilibrium cluster at 50 was also found.

Applications of mixed strategy equilibrium to sports such as tennis and soccer are also discussed towards the end of chapter 3, this time using field experiments. The Walker and Wooders (2001) paper on top tennis players found good support for most predictions of the theory across a number of high-level matches, although there was some evidence of over-alternation in the direction of serves. As an American, Camerer is to be congratulated for at least trying to explain penalty kicks in soccer (p. 146), even though on this occasion his attempt was quite inaccurate. Despite this, the research on soccer found decent support for the predictions of mixed strategy equilibrium (Palacios-Huerta, 2001), confirming the related evidence from tennis matches.

In chapter 4, experiments on structured and unstructured bargaining are discussed, with generally more mixed results than in the previous chapter. In unstructured bargaining, one problem subjects have when trying to strike a deal is that they often believe solutions that favour themselves are fair, making agreement with each other harder. For instance, in bargaining over lottery tickets, Roth and Malouf (1979) found that when two players have unequal prizes, arguments for a 50 : 50 division of the tickets nearly always come from the high-prize player. More generally, if multiple focal points exist, players tend to push for the one under which they personally would do better.

For structured bargaining, one interesting study by Johnson *et al.* (2002) used software called MOUSELAB to record the information subjects seek when bargaining in alternating-offer experiments, and how long they spend at each stage of the process, to infer what approach the subject is using. By requiring subjects to click on boxes to reveal the current and future pie sizes, the authors could see whether subjects engaged in backward induction to locate the subgame perfect equilibrium. A significant minority did not even open the second- and third-round boxes, making backward induction impossible to implement. However, after receiving training on backward induction, outcomes improved significantly. This suggests that it is the unfamiliarity of subjects with the necessary concepts that explains the divergence between prediction and reality. But, given that few people in reality have the necessary conceptual tools required, perhaps we need two separate sets of models and predictions: one for the small minority of trained subjects and another for everybody else? Alternatively, maybe there is a much larger niche for economists in training people how to think strategically than we assume.

Chapter 5 presents a number of interesting findings on dominance-solvable games. These games require subjects to iteratively eliminate any (strongly or weakly) dominated strategies and assume other players will do likewise, until a unique subgame perfect equilibrium is achieved. We referred earlier to beauty contest games which are a very useful tool for investigating how deeply subjects actually reason. Briefly, each of n players must simultaneously choose a number in [0, 100]. There is a fraction p of the average number chosen, which defines the

target for each person to aim for to win the prize. If $p = \frac{2}{3}$, and the average number chosen in some experiment is 60, then the winner is the player closest to 40.

It should be apparent (although it is not always) that any number in [67, 100] is a dominated option, chosen only by zero-step players; if we believe that others realize this also (one-step players), the highest possible average is 66, giving a target of 44. If everyone thinks everyone realizes this (two-step), then the target is really $\frac{2}{3} \times 44$, or about 29. If every player can follow this logic to its conclusion, and believes everyone else can too, the unique solution is to select 0. In reality, most players are one- or two-step players. A few subjects do in fact select 0, but they never win the prize, being literally too smart for their own good! To win, a subject needs to use only one more level of reasoning than everyone else. Camerer appears somewhat inconsistent with his definitions here: on p. 210, he defines a two-step player as choosing in [29, 44], but by p. 217, he switches definitions (now starting from the mid-point of 50), so that a choice of 22 suggests a two-step player. Nonetheless, the results he presents in this section are powerful and sobering, suggesting that even for very intelligent subject pools we are heavily cognitively constrained, or at the least, believe our fellows are.

The centipede game is also popular with experimentalists. A simple version has two players moving sequentially, each of whom must decide to take (T) or pass (P). The size of the pie doubles each time a player chooses P, but ends when a player chooses T. The twist is that the share of the total pie each player would get alternates between 20 and 80% with each move. Backward induction shows equilibrium play is to take at the first opportunity, but frequently some degree of tacit cooperation is achieved. A four-move version (McKelvey and Palfrey, 1992) finds players rarely take in the early stages, but the cooperation unravels towards the end of the game, once again suggesting many people use a couple of steps of iterated dominance only.

Other versions have used more players, more steps and higher incentives. Rapoport *et al.* (2003) used very high stakes (potentially thousands of dollars) in a 3-player centipede game and found more support for Nash behaviour. But a key difference in their design is that the terminal node gave all three players 0, which would work against subjects adopting a 'best for all if we pass' philosophy, as no player now wishes to pass until the end. Camerer's discussion of this study (p. 221) is therefore incorrect to state that 'subjects could have made thousands of dollars if they passed to the end...' and his conclusion that 'a sufficient condition for Nash behaviour seems to be three players and high stakes' is also called into question. It would be interesting, and potentially very expensive, to redo this experiment with the standard progression of payoffs rather than a terminal node of zeroes!

Chapter 6 investigates theories of how players learn in games, looking at evolutionary dynamics, reinforcement learning and belief learning among others. Camerer and Ho (1999) offer Experience-Weighted-Attraction (EWA) as a more general theory incorporating reinforcement and belief learning as special cases. One problem I have with the models tested in this chapter is that they appear to me to start from rather implausible premises. For example, reinforcement learning is

grounded in the long-discredited behaviourist psychology of people like B.F. Skinner. Additionally, the evolutionary approach described here assumes that players are born with an unchanging strategy, the more successful ones increasing in relative frequency over time. This has little to do with how an evolutionary psychologist, for example, would view human learning. Overall, the evidence reported in this chapter strikes me as inconclusive, although EWA performs better across a wider range of games than the simpler theories. I hope future learning models will draw upon some more cogent and up-to-date psychology than the blank-slate theories discussed here.

Chapter 7 looks at coordination games as a solution to the problem of multiple equilibria. This is one of the most engaging chapters in this book because a country's history is often involved in the selection of one equilibrium over another. It also highlights the importance of experimental methods in GT research, as mathematics alone often cannot tell us which outcome should or will occur. Memorable anecdotes on the origins of conventions (pp. 338–340) include the standard width of railway tracks, the geographical concentration of certain industries in specific locations and why some countries drive on the left while others drive on the right.

One story I like is the story of driving in Bolivia. Although Bolivians normally drive on the right and hence have steering wheels in the left of the vehicle, they switch sides on mountain roads. This is because those roads are narrow and dangerous with no easy view of the cliff edge from the driver's seat. By switching sides, both sets of drivers gain a clearer view of the most dangerous sections. Fortunately, there are road signs that warn drivers the convention is about to be reversed; however, it does seem that Bolivia settled on the wrong choice of equilibrium back in the mists of time.

A general finding from this literature is that, for many non-rational reasons, players often succeed in coordinating their actions, if they have a common conception of a more prominent or focal choice. Where there are clashing focal points, coordination failure is more common. Allowing communication between players can be some help both when it allows them to focus on just one equilibrium and when it provides assurance that they will behave as they say. Interestingly, this means that, in some games, one-way communication works better than two-way, while in others, the reverse is true.

Chapter 8 looks at signalling, screening and reputation games. In a signalling game, one player takes some action to signal his type or future intentions to uninformed players if the costs of the signal are outweighed by the benefits he accrues from the receivers of the signal believing the message. A credible signal is affordable only by the type that sends it, leading to separating equilibrium. Pooling equilibrium often results if both types can afford the signal. This phenomenon occurs in the business world when a firm offers a very generous warranty or guarantee on its product, to overcome unjustified suspicion regarding its quality.

Experimentally, however, many players fail both to draw the full logical inferences of how their actions reveal their private information to others and to

rapidly infer the types of other players from their actions. Given that real players are then boundedly rational, Camerer refers to a study by Cooper *et al*. (1997) who note that this fact provides a role for the use of redundant signals in these games, to clarify and reinforce messages.

The book closes with a useful summary of the main findings of the earlier chapters and an appendix describing the design details of 90 of the experimental studies discussed previously. Also, in chapter 9, he offers a wide-ranging list of his top 10 open research questions in behavioural GT, divided into five where the answers are becoming clearer such as how we value the payoffs of others and five where much is still to be done, such as what games subjects think they are playing. He ends with the reasonable and, I think, realistic hope that eventually the behavioural label can be dispensed with as mainstream GT inexorably absorbs the many lessons from this reality check and fulfils at last its initial promise. Judging from the research documented in this excellent book, we may not have long to wait.

Acknowledgements

I thank Martin Dufwenberg and participants in his behavioural game theory seminar in 2004 at the University of Arizona for inspiring me to read and discuss Colin Camerer's book. A number of the points I raise were prompted by listening to and reflecting upon the comments of Martin and his students. However, I remain solely responsible for any errors or misinterpretations that may be contained herein. I also thank the Economics Department at Arizona where I began this review.

References

Barrett, L., Dunbar, R. and Lycett, J. (2002). *Human Evolutionary Psychology*. London: Palgrave.

Brosnan, S. and DeWaal, F. (2003). Monkeys reject unequal pay. *Nature* 425: 297–299.

Camerer, C. and Ho, T. (1999). Experience-weighted attraction learning in normal-form games. *Econometrica* 67: 827–874.

Collins, R. and Sherstyuk, K. (2000). Spatial competition with three firms. *Economic Inquiry* 38: 73–94.

Cooper, D., Garvin, S. and Kagel, J. (1997). Adaptive learning versus equilibrium refinements in an entry limit pricing game. *Economic Journal* 107: 553–575.

Dufwenberg, M. and Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior* 47: 268–298.

Geanakoplos, J., Pearce, D. and Stachetti, E. (1989). Psychological games and sequential rationality. *Games and Economic Behavior* 1: 60–79.

Goeree, J. and Holt, C. (2001). Ten little treasures of game theory and ten intuitive contradictions. *American Economic Review* 91: 1402–1422.

Hey, J.D. and Orme, C. (1994). Investigating generalisations of expected utility theory using experimental data. *Econometrica* 62: 1291–1326.

Hück, S., Müller, W. and Vriend, N. (2002). The East end, the West end, and King's Cross: on clustering in the four-player Hotelling game. *Economic Inquiry* 40: 231–240.

Johnson, E., Camerer, C., Sen, S. and Rymon, T. (2002). Detecting failures of backward induction: monitoring information search in sequential bargaining experiments. *Journal of Economic Theory* 104: 16–47.

Kiyonari, T., Tanida, S. and Yamagishi, T. (2000). Social exchange and reciprocity: confusion or a heuristic? *Evolution and Human Behavior* 21: 41–427.
Loomes, G. and Sugden, R. (1995). Incorporating a stochastic element into decision theories. *European Economic Review* 39: 641–648.
McCabe, K., Houser, D., Ryan, L., Smith, V. and Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Science* 98: 11832–11835.
McKelvey, R. and Palfrey, T. (1992). An experimental study of the centipede game. *Econometrica* 60: 803–836.
Palacios-Huerta, I. (2001). *Professionals Play Minimax*. Brown University Working paper.
Panksepp, J. (1998). *Affective Neuroscience: the Foundations of Human and Animal Emotions*. New York: Oxford University Press.
Pinker, S. (1994). *The Language Instinct: the New Science of Language and Mind*. London: Allen Lane.
Rabin, M. (1993). Incorporating fairness into game theory. *American Economic Review* 83: 1281–1302.
Rapoport, A., Stein, W., Parco, J. and Nicholas, T. (2003). Equilibrium play and adaptive learning in a three-player centipede game. *Games and Economic Behavior* 43: 239–265.
Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G. and Kilts, C. (2002). A neural basis for social cooperation. *Neuron* 35: 395–405.
Roth, A. and Malouf, M. (1979). Game-theoretic models and the role of information in bargaining. *Psychological Review* 86: 574–594.
Segal, N. and Hershberger, S. (1999). Cooperation and competition between twins: findings from a prisoner's dilemma game. *Evolution and Human Behavior* 20: 29–51.
Sugden, R. (2001). Ken Binmore's evolutionary social theory. *Economic Journal* 111: F213–F243.
Walker, M. and Wooders, J. (2001). Minimax play at Wimbledon. *American Economic Review* 91: 1521–1538.