Review TRENDS in Cognitive Sciences Vol.7 No.5 May 2003

Behavioural studies of strategic thinking in games

Colin F. Camerer

Division HSS, 228-77 California Institute of Technology, Pasadena, CA 91125, USA

Game theory is a mathematical language for describing strategic interactions, in which each player's choice affects the payoff of other players (where players can be genes, people, companies, nation-states, etc.). The impact of game theory in psychology has been limited by the lack of cognitive mechanisms underlying game-theoretic predictions. 'Behavioural game theory' is a recent approach linking game theory to cognitive science by adding cognitive details about 'social utility functions', theories of limits on iterated thinking, and statistical theories of how players learn and influence others. New directions include the effects of game descriptions on choice ('framing'), strategic heuristics, and mental representation. These ideas will help root game theory more deeply in cognitive science and extend the scope of both enterprises.

In game theory, a 'game' is a complete specification of the strategies each 'player' has, the order in which players choose strategies, the information players have, and how players value possible outcomes ('utilities') that result from strategy choices. Game theory is applied at many levels: players can be genes, people, groups, companies or nation-states. Strategies can be genetically coded instincts, bidding methods, a legal defense, corporate practices, or a wartime battle plan. Outcomes can be anything players value – food or reproduction, buying a jewelry box on the internet, an acquittal, company profits or a 'regime change'.

Game theory began as mathematics and continues largely in that tradition, solving puzzles about how idealized players will behave, or which strategies evolve [1]. Game theory has become standard in many social sciences, but has little influence in psychology because many of its principles are not well grounded cognitively. This article is about an emerging approach – 'behavioural game theory' – which uses experimental evidence to inform mathematical models of cognitive limits, learning rules and social utility [2,3]. This approach is similar to other recent work on judgment and decision, but specialized to strategic interaction. Behavioural game theory could lead to a synthesis with cognitive science that will be of interest to both communities.

The best-developed elements of behavioural game theory are: theories of limited strategic thinking; theories of learning; and social preference (or utility) functions. After describing these topics, some promising ideas that are less well-developed will be mentioned.

Corresponding author: Colin F. Camerer (camerer@hss.caltech.edu).

Limited strategic thinking

The mathematical core of game theory is what players think other players will do. In most theories, this reasoning is iterated (A guesses what B will do by guessing what B will guess A will do, ad infinitum) until mutually consistent responses – an 'equilibrium' – is reached. For games that are new to players, a more plausible model is that players use a limited number of steps of iterated reasoning. The 'p-beauty contest (pBC) game' is a good tool for measuring steps of thinking [4,5]. In a pBC players simultaneously choose a number between zero and 100. The player whose number is closest to 2/3 of the average number wins a fixed prize. This game models situations like entering a rapidly growing industry or political race, undercutting a competitor's price (but not by too much), or deciding when to sell stocks before an anticipated crash in the market. In each case the goal is to be ahead of the pack, but not too far ahead.

The Nash equilibrium in the pBC is zero – the only number that is a best-response if others are also choosing it – which results from infinitely many steps of iteration. However, in most experimental-subject pools the average number is in the range 20–40 (see Box 1). Some new theories suggest precisely how limited thinking may work in pBC games and a wide variety of other games (see Box 2).

Inferring thinking from cognitive measures

Steps of thinking can also be inferred from direct cognitive measures. Suppose, for example, that players can view the payoffs they will get by opening boxes on a computer screen with a mouse click. Players' attention can then be measured directly. Consider a bargaining game in which players alternate offers over how to divide a sum of money: Player 1 makes the first offer; if it is rejected Player 2 makes a counteroffer, and so on. Because of impatience or costly delay, the amount available to divide shrinks after each rejected offer. Game theory assumes that selfish players look ahead to all possible future bargaining periods, then 'backward induct' to see what they should offer and accept in the *first* period, to get Player 2 to accept the offer rather than reject and make a counteroffer. In an experiment where subjects had to click open a box to see just how much the amount would shrink, most subjects did not look ahead and work backwards (see Fig. 1) [8]. In fact,

 1364-6613/03/$ - see front matter doi:10.1016/S1364-6613(03)00094-9

Box 1. *p*-beauty contests

The *p*BC game got its name from a metaphor used by economist John Maynard Keynes, who likened the stock market to a beauty contest in which people care about which contestant *other people* consider to be beautiful (and further iterations of expectations), rather than who is *truly* most beautiful. (His metaphor is an apt one for the 'tech stock' boom and bust of the late 1990s, when investors self-reportedly cared little about actual company profits as long as others were optimistic.) The *p*BC game has been run with many subject pools. The average number chosen is usually 25–40 with a large standard deviation (around 20) (Fig. I). This result is robust across countries (Germany, US, Singapore) and ages (high-school students to 70-year-old, well-functioning adults). The lowest averages, 15–20, come from subject pools with unusual analytical skill (Caltech students), training in game theory, and from people entering newspaper contests who are self-selected for their motivation and knowledge [5].

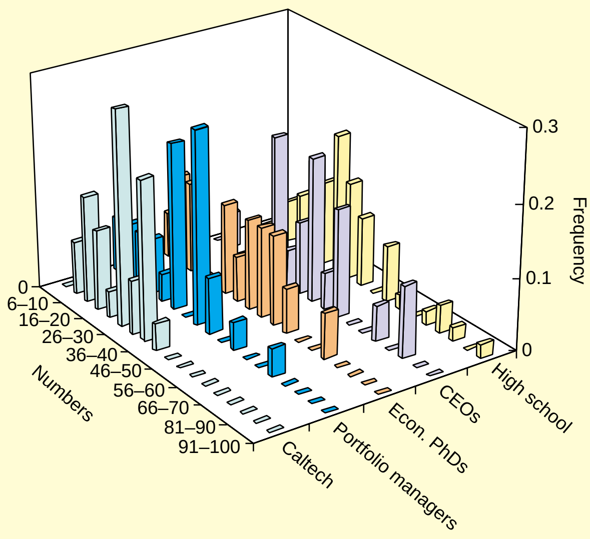

Fig. I. Numbers (front left axis) are choices from 0 to 100 in *p*BC games (number bins are 0, intervals of five numbers beginning 1–5, and intervals of ten numbers above 70, beginning 71–100). Frequencies of different number choices are on the vertical axis. The number closest to 2/3 times the average wins $20. The equilibrium prediction is to zero, but most subject pools average 20–35. Subject pools are (left to right): Caltech undergraduates, stock market portfolio managers, economics PhD students, CEOs, and high school students. Different subject pools have different patterns of choices, with lower averages in subject pools with greater analytical skill.

Box 2. Models of limited iterated thinking

New theories formalize limits on strategic thinking. In 'cognitive hierarchy' (CH) theory, players use a series of recursive thinking steps [5]. The proportion of players stopping after each step is given by a one-parameter Poisson distribution with mean τ (i.e. the average number of thinking steps). CH theory assumes that higher-step players choose the strategy with the highest expected payoff, given their perceived distribution of what lower-step thinkers will do. In the *p*BC game (Box 1), players doing zero steps of thinking randomize equally across all numbers. One-step players think they are playing zero-step players, so they expect an average of 50 and choose 2/3 of 50, or 33. Two-step players think they are playing a mixture of zero- and one-step players. If two-step players guess the relative proportions of lower-level thinkers accurately, their best response is 26 (for $\tau = 1.5$). Further iterations converge rapidly because Poisson frequencies of higher-level thinkers drop off sharply (probably owing to working memory constraint). Best-fitting estimates of τ for the five subject pools shown in Fig. I (Box 1) are 3.0, 2.8, 2.3, 1.0 and 1.6. A related two-parameter model of 'noisy introspection' [6] assumes that players are less and less confident about higher iterations of thinking, and choose according to a stochastic utility or 'softmax' response rule (i.e. nearly-best responses are chosen almost as often as best responses). An even richer approach allows many possible decision rules and uses statistics to recover which rules appear to be most common (usually one-step thinking does [7]).

in 10–20% of the trials players did not open the second- and third-round boxes at all so they could not possibly be computing the equilibrium. Attention measures also show that players do not look at payoffs other players might have earned, even when it might help them guess what those players will do in the future [9]. Other findings are that heuristic one- or two-step rules (see Box 3) are common [10], and that players' mental models often under-represent the payoffs of others [10].

Learning

Even though strategic thinking is limited, behaviour can approximate equilibrium predictions surprisingly well if people can learn over time (or through imitation or some other adaptive process). For example, Fig. 2 shows data from a *p*BC game repeated 10 times, with feedback after each round. After several rounds, most subjects chose numbers close to the equilibrium of zero.

Many statistical rules of how learning occurs have been developed to explain experimental data (and field data [11,12]). One approach is belief learning ('fictitious play'), in which players use the history of play by others to form beliefs about what others will do, and respond accordingly [13]. Another approach is reinforcement, in which players repeat previous strategies if they yielded good payoffs [14] (perhaps spreading reinforcement to similar strategies [15], and learning more slowly when reinforcement is variable [16]). Both models fit data paths better than equilibrium theories (which predict no learning) and are insightful. In a hybrid model, experience-weighted attraction (EWA) learning, players put a partial weight δ on counterfactual imagination of foregone payoffs from strategies they did not choose [17] (see Box 3). EWA learning is more robust than belief or reinforcement models because it predicts as well as those models do, but also more accurately when those models are too fast or too slow compared with human learning. Another hybrid is

Box 3. Models of learning in games

In the hybrid EWA model [17], the numerical attraction of strategy *j* after period t is updated according to $A_t(j) = [\phi A_{t-1}(j) + \delta\pi_t(k)]/[\phi(1-\kappa)+1]$, where $\pi_t(k)$ is the actual payoff or foregone payoff from strategy *k*, and the weight δ is set to 1.0 if strategy *j* is chosen. Attractions map onto choice probabilities using a stochastic choice ('softmax') rule. The weight ϕ decays previous attractions; low ϕ represents memory decay or discarding old history in a changing environment. If $\kappa = 1$ attractions cumulate and players converge sharply; if $\kappa = 0$ attractions are weighted averages. Fictitious play belief learning corresponds to $\phi = \delta = 1, \kappa = 0$; simple reinforcement corresponds to $\delta = 0$. The values of the parameters δ, ϕ and κ can be estimated from data or derived from functions that adapt to experience.

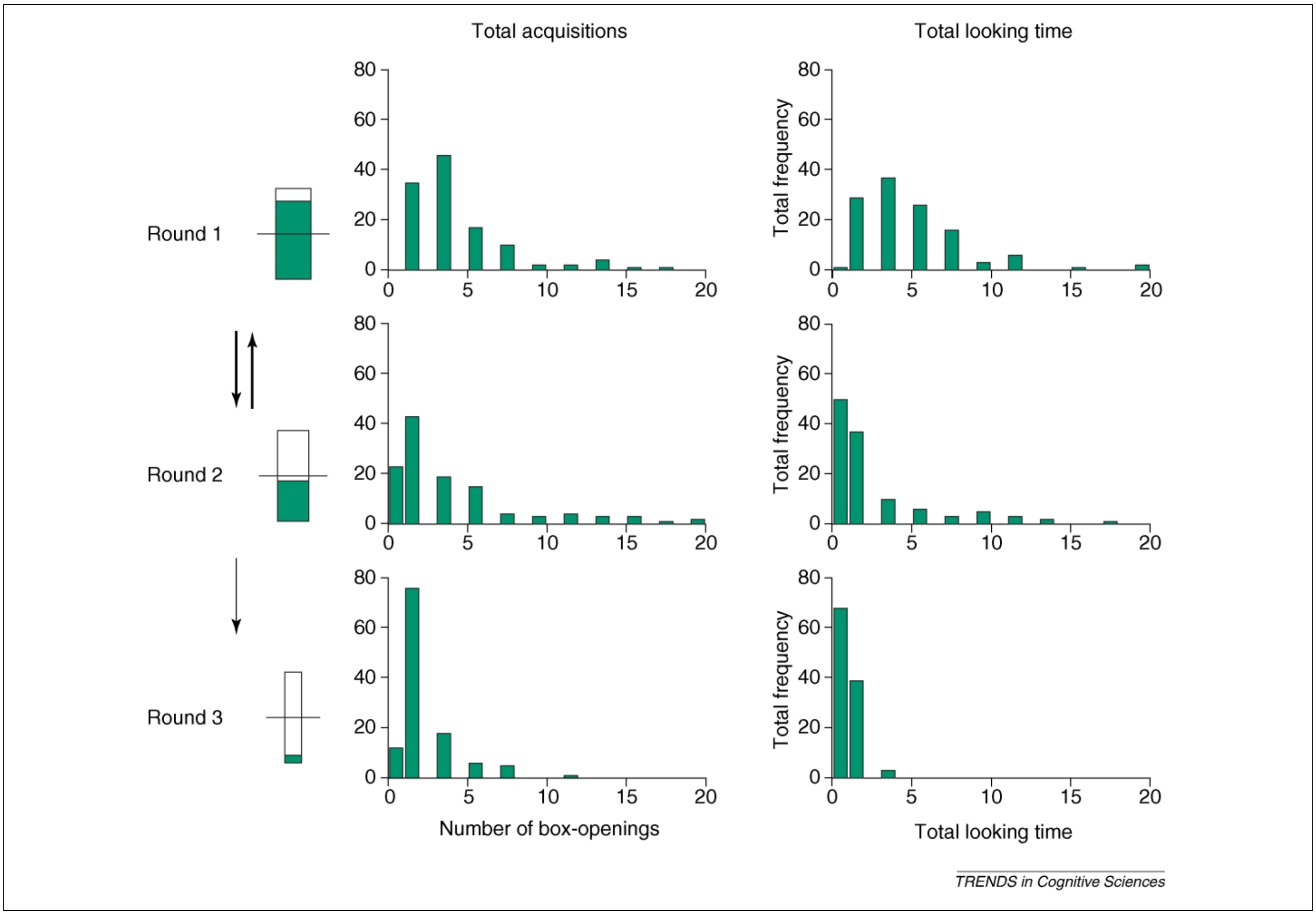

Fig. 1. Each box represents one stage in three-stage bargaining with alternating offers (Round 1 at top). Icons (left) show the relative time for which each box is opened (indicated by shaded portion), the relative number of times each box is opened (indicated by box width), and the relative number of forward (down arrow) and backward (up arrow) transitions (indicated by arrow thickness). In stage 1 (top box) players divide $5; in stage 2 they divide $2.50; in stage 3 they divide $1.25 (and if the third-stage offer is rejected they earn nothing). The icons show that most players look longest and most often at the first stage, and do not 'backward induct' by looking at the third stage then working backward (there is no arrow pointing upward from box 3 to box 2). Histograms show the frequencies of number of box-openings (middle column) and total time boxes are open (right column). Note that in 20% of trials box 2 is not opened at all (zero box-openings) and in 10% of trials box 3 is not opened.

'rule learning': players shift weight towards learning rules that give higher payoffs [18].

These learning theories are much like those used to study animal and human learning in other domains. What makes learning in games special is that 'sophisticated' players realize they are playing against other people who are also learning, and adjust their behaviour accordingly (like rugby football quarterbacks who learn to throw a ball 'ahead of the receiver' – to where a receiver will be in a couple of seconds). Sophistication is particularly important when players are matched together repeatedly – like workers in firms or rival companies. For example, suppose 2/3 of the average in one period of the *p* BC game was 27. A player who learns might choose 2/3 of 27 in the next period, or 18. However, a sophisticated player who anticipates that others are adjusting in this way will choose 2/3 of 18, or 12; and the sophisticated player will win more often. Figure 2 shows the fit of an EWA learning model that includes a large proportion of sophisticated players. Estimates also show that as players gain experience with the game, the degree of sophistication rises – they learn that others (like themselves) are learning [19].

Social preferences

Behavioural game theory is useful for studying social motives that occur in strategic interactions, such as altruism, fairness, trust, vengeance, hatred, reciprocity and spite. An important part of behavioural game theory is to build precise models of how these forces work, derived from data and other considerations (e.g. evolutionary stability).

A familiar game is the prisoners' dilemma (PD) In a PD, players are collectively better off if they all 'cooperate', but players privately prefer to 'defect', whether others cooperate or not. (Contributing to 'public resources', which benefit each player less than their private contribution, are a close relative of PD.) Many experiments have shown that players cooperate in a one-shot PD about half the time and contribute about half of their endowment in public resources games [20, 21]. Players who cooperate typically say they expect others to cooperate, which is consistent with the idea that cooperation is reciprocal (or 'conditional') rather than simply altruistic or rooted in moral principle (and also consistent with attribution theory [22]). Giving players a chance to punish low contributors, at a cost to themselves, raises group contributions close to the optimal level at which everyone contributes [23]. Cooperation also rises sharply when

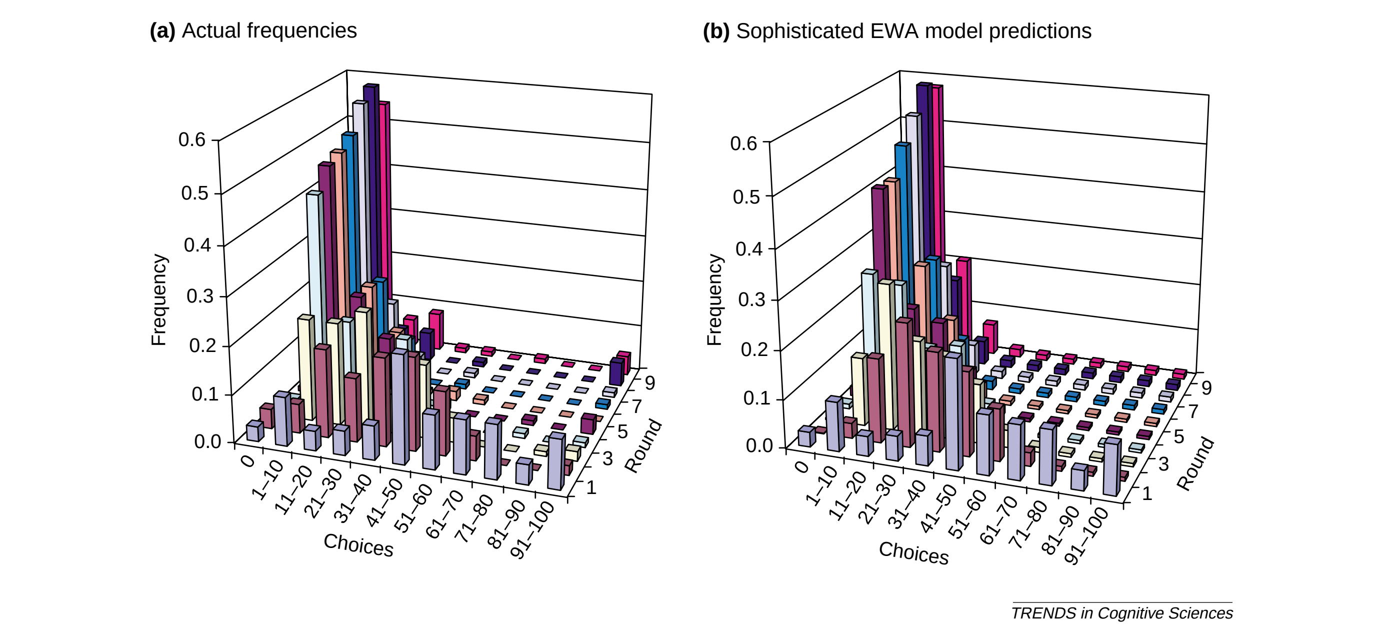

Fig. 2. (a) Frequencies (vertical axis) of number choices (front axis) made by Singapore students in *p*BC games across 10 rounds (right axis), with $p = 0.7$ and 0.9. (b) Averaged predictions of a model of 'sophisticated learning' across 10 rounds. The model assumes that 22% of subjects learn according to a hybrid EWA rule, and an estimated 78% are 'sophisticated' and realize that 22% are learning and 78% are sophisticated. It can be seen that there is a close fit between the model and the experimentally derived frequencies.

players have a chance to talk about what they plan to do, even if they won't see each other again after talking [21], which suggests that promising or empathy are important influences on behaviour.

Ultimatum bargaining

More evidence of social motives comes from ultimatum (or 'take-it-or-leave-it') bargaining [3]. In an ultimatum game a Proposer makes a one-time offer to a Responder, who can accept it or reject it; if she rejects the game ends and they both get nothing. Players who are selfish, and think that others are too, will offer the least they can, and take anything they are offered. Contrary to this prediction, the average offer is usually 30–50% and offers of less than 20% are rejected half the time (see Box 4).

Other games: dictators, trust, and competition

Rejections by Responders in ultimatum games show negative reciprocity or vengeance (the willingness to sacrifice money to punish others who were unfair). Other games reveal other motives. In dictator games, a Proposer dictates an allocation the Responder must accept. As the Responder cannot reject the offer, the Proposer's offer measures pure altruism rather than strategically offering enough to avoid rejection. In dictator games, Proposers offer less than in ultimatum games: around 15% of the stakes (and many subjects offer nothing). However, the average offer varies widely with contextual labels and with other variables (e.g. if the Proposer knows more about the Responder's personal characteristics she gives more).

Many social scientists are interested in trust (and the broader concept of 'social capital'), which underlies a healthy and productive society, and is strongly correlated with economic growth [27]. Game theory gives a crisp way to define trust and measure it. In one trust game [28], an investor is endowed with a sum, typically $10, and invests as little or much as she likes. The amount invested is tripled (representing a return on social investment) and given to an anonymous Trustee. The Trustee can pay back as much of the tripled sum as she likes to the Investor, or keep it all. The amount invested measures trust; the amount repaid measures trustworthiness.

Trust games model opportunities to gain from investment with no legal protection against theft by a business partner or government, or employers prepaying a wage but depending on workers to work hard even if part of their effort is not observed [23]. If selfish players play only once, Trustees will never pay back any money; rational selfish Investors should anticipate this and invest nothing. In fact, in one-shot games Investors typically risk about half their money, and Trustees pay back slightly less than was risked. Trustee repayment shows positive reciprocity. Initial fMRI evidence shows that cooperative behaviour

Box 4. Ultimatum bargaining

In an ultimatum game a Proposer makes a one-time offer to a Responder, who can choose to accept it or reject it. Ultimatum offers of nearly 50%, and substantial rejection rates by Responders, have been observed in many societies, even at high stakes (up to $400 in the US, and sums with even larger purchasing power in poorer countries [24]). A remarkable comparison across 15 simple small-scale societies, in remote places like Papua New Guinea, shows that the tendency to share equally is positively correlated with (a) the degree of cooperation in a society, and (b) the amount of integration people have into market trading (e.g. selling crops or cows at a village center) [25]. Furthermore, self-interest predictions that Proposers will offer very little, and Responders will take anything, do appear to hold up in certain populations such as small-scale societies with little social fabric (e.g. the Machiguenga in Peru [26]).

in these games is correlated with activity in the limbic system, and in prefrontal cortex – thought to be a location of 'theory of mind' [29]. When trust games are repeated, players are typically very trusting and trustworthy in early periods, but trust typically breaks down in the last period or two. This pattern shows that early on, trust and trustworthiness are partly 'strategic' trust (designed to maintain goodwill and elicit future benefits), and strategic trust evaporates as the amount of future benefit shrinks towards the end of an interaction [30].

Competition has a strong effect in many of these games. For example, if two or more Proposers make offers in an ultimatum game, and a single Responder accepts the highest offer, then the only game-theoretic equilibrium is for the Proposers to offer all the money to the Responder (the complete opposite of the prediction in the ultimatum game with one Proposer). In experiments, this predicted competition does occur rapidly, resulting in a very unfair allocation – almost no earnings for Proposers [31].

Mathematical theories of social Preferences

An obvious way to explain these phenomena, suggested long ago [32], is that a player's utility for allocations includes both her own earnings and a weight α on the other player's earnings [33]. There are consistent behavioural correlations of revealed α with psychometric scales within a game (e.g. people high in 'Machiavellianism' are less trustworthy [34]). However, the predominant value of α varies systematically across situations. Ultimatum responders who reject unfair offers act as if $\alpha < 0$, players who reciprocate trust act as if $\alpha > 0$, and offering everything to another side in the face of competition is consistent with $\alpha = 0$. Systematic variation across games means that α is not solely a personality trait. The challenge is then to find a single social utility function that explains typical cross-game behavioural differences without switching parameter values.

Three main theories have been proposed. (1) In 'inequality-aversion' theories, players prefer more money *and* prefer that allocations be equal; therefore, players will sacrifice money to make outcomes more equal [35,36]. Utility functions of this sort can explain all the patterns mentioned above but cannot explain why players reject 'unfair' (Proposer-generated) offers more often than 'uneven' ones [37]. (2) In 'me-min-us' Rawlsitarian theory, players care about their own payoff, the minimum payoff, and the total of all payoffs [38]. This theory cannot explain ultimatum rejections (because rejecting an offer lowers all three components) but can account for other patterns which inequality aversion cannot. (3) In reciprocity theories, player A forms a judgment about another player B's 'kindness' – did B's choice help A, or harm her? [39] Kindness is numerically scaled so that helping is positive and harming is negative. Players who care about the product of their own kindness *and* the other player's kindness will reciprocate both positively and negatively (because behaving negatively towards a person who is negatively kind – or 'mean' – creates positive utility). Reciprocity theories can explain the correlation between cooperation and expectations of cooperation by PD.

Many studies are now comparing these different theories. Elements of all three will probably be useful for different purposes.

New directions

Acknowledging limits on strategic thinking and monetary self-interest allows scope for individual differences in analytical skill or motivation, personality traits, gender, race, and so on, to affect behaviour. Many such variables have been studied (e.g. Box 2 shows that high-IQ college students do more steps of thinking than others, and psychometric measures correlate with cooperativeness [34]), but these effects are often modest compared with the effects of different game structures.

Framing effects of game description

An example is framing: how does the explicit description of a game influence choice? For example, players in ultimatum games divide *less* evenly when the game is described as a buyer–seller interaction, or when the Proposer earns the right to make an offer by winning a preplay contest [40]; they divide *more* evenly when the game is described as making claims to a common resource [41]. These description changes appear to evoke different shared social norms for what divisions are fair (as in equity theories in social psychology).

Framing effects are particularly important in games where players have a common interest in coordinating their actions, because the way strategies are described can focus attention on psychologically prominent 'focal points'. Coordination games are an embarrassment for standard theory because it is hard to derive mathematical rules that pick out the one of many equilibria that is obvious (and usually played). Suppose two players can simultaneously choose Red or Blue. They earn $10 if they both choose Red and $5 if they both choose Blue. They will surely choose Red. But both choosing Blue is also a Nash equilibrium, which should be chosen according to standard theory. Behavioural theories explain the obvious choice of Red by assuming that players implicitly act as a team [42], or players use a 'Stackelberg heuristic': they act as if they are going first, but others will figure out what they are likely to have chosen and 'follow' them [43,44].

Mental representation

A related direction is mental representation. Theorists analyze games in the form of matrices or trees but players presumably construct internal representations that might barely resemble matrices or trees. Just as people do not represent explicitly false propositions in mental models of logic, players appear to under-represent payoffs of others in their mental models of games. Games with mixed motives, and with conflicting rankings of outcomes across players ('biorders'), are also difficult to represent [45]. Limits on representation are particularly important when games are quite complex, with many players and strategies, unfolding over time (like diplomatic maneuvering or planning a business strategy). Ideas from multi-agent machine learning suggest new approaches to modeling people [46], and empirical models of how players behave may inspire new machine learning algorithms.

Individual choice heuristics in games

Another direction for behavioural game theory is exploring when systematic deviations from rationality observed in individual choices also occur in games. For example, it is well-known that people expect random series to 'even out' more rapidly then they do; this too often leads to alternating strategies when people play games that require unpredictable randomization [47]. When game payoffs are described as losses from a reference point, players take longer to choose and take more risk [48], are less cooperative [49], and pass up more mutually-beneficial trades [50], compared with 'gain-framed' games with equivalent final payoffs. Measured beliefs about what others will do in games sum to less than one, which mathematically corresponds to a pessimistic reluctance to take action when important information is missing ('ambiguity-aversion') [51]. In competitive games mimicking entry into new businesses, subjects are overconfident (they all think they are more skilled than average, and as a results, lose money as a group) and they neglect the number and skill of likely competitors [52].

Conclusion

Behavioural game theory has progressed rapidly since the term was coined 10 years or so ago [53]. It extends the cognitive plausibility and empirical accuracy of game theory, expressing ideas in mathematical models that permit rapid progress. Further innovation will be helped by data from cognitive science, such as measurement of response times, information acquisition, and findings from fMRI. It will also be aided by studies of special populations (such as children, people with autism, and people in small-scale societies), and by input from cognitive scientists about mental representation and strategic heuristics, especially when games are complex. Many questions remain about the neural mechanisms underlying strategic thinking and heuristics, learning, and social utility (see Box 5. Questions for Future Research). Better theories of how people behave will also help in the design of robust economic institutions (like auctions and incentive structures in companies) [54,55]. The future is, as always, unpredictable, but the closer links that are being forged between game theory and cognitive science should extend the scope of both fields of enterprise.

Box 5. Questions for Future Research

- What neural mechanisms correspond to aspects of strategic thinking?
- Are ultimatum rejections and other apparent expressions of social preference due to emotions, learned heuristics, evolved modules, or combinations of these and other mechanisms?
- Which models of social preference are best-suited to which games?
- Are there reliable cross-game individual differences in social preferences?
- How do players learn in games when there is little available information?
- Do groups behave differently from individuals in games?
- What do mental representations of games look like?

Acknowledgements

This work was supported by NSF grant SES-0078911. The author wishes to express thanks to Yuval Rottenstreich, the *TICS* Editor and the anonymous referees for very helpful comments.

References

1 Weibull, J. (1995) *Evolutionary Game Theory*, MIT Press
2 Crawford, V. (1997) In *Advances in Economics and Econometrics: Theory and Applications* (Kreps, D. and Wallis, K., eds) pp. 206–242, Cambridge University Press
3 Camerer, C.F. (2003) *Behavioural Game Theory: Experiments on Strategic Interaction*, Princeton University Press
4 Nagel, R. (1998) In *Games and Human Behavior: Essays in Honor of Amnon Rapoport* (Budescu, D., Erev, I. and Zwick, R., eds) Lawrence Erlbaum Assoc., Inc.
5 Camerer, C. *et al.* (2003) Models of thinking, learning and teaching in games. *Am. Econ. Rev.*
6 Goeree, J. and Holt, C. A theory of noisy introspection. *Games Econ. Behav.* (in press)
7 Stahl, D.O. and Wilson, P. (1995) On players' models of other players: theory and experimental evidence. *Games Econ. Behavior* 10, 218–254
8 Johnson, E.J. *et al.* (2002) Detecting backward induction in sequential bargaining experiments. *J. Econ. Theory* 104, 16–47
9 Johnson, E. and Camerer, C. In *The Psychology of Economic Decisions*, (Brocas, I and Castillo, J., eds.), Oxford University Press
10 Costa-Gomes, M. *et al.* (2001) Cognition and behavior in normal-form games. *Econometrica* 69, 1193–1235
11 Weisbuch, G. *et al.* (2000) Market organisation and trading relationships. *Econ. J.* 110, 411–436
12 Ho, T. and Chong, J.K. A parsimonious model of SKU choice. *J. Mark. Res.* (in press)
13 Cheung, Y.W. and Friedman, D. (1997) Individual learning in normal form games: some laboratory results. *Games Econ. Behavior* 19, 46–76
14 Erev, I. and Roth, A. (1998) Predicting how people play games: reinforcement learning in experimental games with unique, mixed-strategy equilibria. *Am. Econ. Rev.* 88, 848–881
15 Sarin, R. and Vahid, F. Strategy similarity and coordination. *Econ. J.* (in press)
16 Erev, I. *et al.* (1999) The effect of adding a constant to all payoffs: experimental investigation, and implications for reinforcement learning models. *J. Econ. Behav. Organ.* 39, 111–128
17 Camerer, C. and Ho, T. (1999) Experience-weighted attraction learning in normal-form games. *Econometrica* 67, 827–874
18 Stahl, D. (2000) Rule learning in symmetric normal-form games: theory and evidence. *Games Econ. Behavior* 32, 105–138
19 Camerer, C.F. *et al.* (2002) Sophisticated experience-weighted attraction learning and strategic teaching in repeated games. *J. Econ. Theory* 104, 137–188
20 Ledyard, J. (1995) In *Handbook of Experimental Economics* (Kagel, J. and Roth, A., eds) Princeton University Press
21 Sally, D. (1995) Conversations and cooperation in social dilemmas: a meta-analysis of experiments from 1958 to 1992. *Rationality and Society* 7, 58–92
22 Kelley, H.H. and Stahelski, A.J. (1970) Social interaction basis of cooperators' and competitors' beliefs about others. *J. Pers. Soc. Psych* 16, 66–91
23 Fehr, E. and Gachter, S. (2000) Fairness and retaliation: the economics of reciprocity. *J. Econ. Perspect.* 14, 159–181
24 List, J. and Cherry, T. (2000) Learning to accept in ultimatum games: evidence from an experimental design that generates low offers. *Exp. Econ.* 3, 11–31
25 Henrich, J. *et al.* (2001) In search of homo economicus: behavioral experiments in 15 small-scale societies. *Am. Econ. Rev.* 91, 73–78
26 Henrich, J. (2000) Does culture matter in economic behaviour? Ultimatum game bargaining among the Machiguenga of the Peruvian Amazon. *Am. Econ. Rev.* 90, 973–979
27 Knack, S. and Keefer, P. (1997) Does social capital have an economic payoff? a cross-country investigation. *Q. J. Econ.* 112, 1252–1288

28 Berg, J. *et al.* (1995) Trust, reciprocity, and social history. *Games Econ. Behavior* 10, 122–142
29 McCabe, K. *et al.* (2001) A functional imaging study of cooperation in two-person reciprocal exchange. *Proc. Natl. Acad. Sci. U. S. A.* 98, 11832–11835
30 Camerer, C. and Weigelt, K. (1988) Experimental tests of a sequential equilibrium reputation model. *Econometrica* 56, 1–36
31 Roth, A.E. *et al.* (1991) Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh and Tokyo: an experimental study. *Am. Econ. Rev.* 81, 1068–1095
32 Edgeworth, F. (1881) *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences*, Routledge Kegan
33 McClintock, C. (1972) Social motivation – a set of propositions. *Behav. Sci.* 17, 438–454
34 Gunnthursdottir, A. *et al.* (2002) Using the Machiavellian instrument to predict trustworthiness in a bargaining game. *J. Econ. Psychol.* 23, 49–66
35 Fehr, E. and Schmidt, K.M. (1999) A theory of fairness, competition and cooperation. *Q. J. Econ.* 114, 817–868
36 Bolton, G.E. and Ockenfels, A. (2000) ERC: a theory of equity, reciprocity, and competition. *Am. Econ. Rev.* 90, 166–193
37 Blount, S. (1995) When social outcomes aren't fair: the effect of causal attributions on preference. *Organ. Behav. Hum. Decis. Process.* 63, 131–144
38 Charness, G. and Rabin, M. (2002) Understanding social preferences with simple tests. *Q. J. Econ.* 117, 817–869
39 Rabin, M. (1993) Incorporating fairness into game theory and economics. *Am. Econ. Rev.* 83, 1281–1302
40 Hoffman, E. *et al.* (1994) Preferences, property rights and anonymity in bargaining games. *Games and Econ. Behavior* 7, 346–380
41 Larrick, R.P. and Blount, S. (1997) The claiming effect: why players are more generous in social dilemmas than in ultimatum games. *J. Pers. Soc. Psychol.* 72, 810–825
42 Sugden, R. (1993) Thinking as a team: toward an explanation of nonselfish behavior. *Soc. Philos. Policy* 10, 69–89
43 Colman, A. and Stirk, J.A. (1998) Stackelberg reasoning in mixed-motive games: an experimental investigation. *J. Econ Psychol.* 19, 279–293
44 Weber, R. and Camerer, C.F. First-mover advantage and virtual observability in ultimatum bargaining and 'weak-link' coordination games. *Games Econ. Behavior.* (in press)
45 Warglien, M. *et al.* (1999) I modelli mentali dei giochi: focalizzione e rappresentazioni selettive. *Sistemi Intelligenti* 11, 85–112
46 Warsh, W.E. *et al.* Analyzing complex strategic interactions in multi-agent games. *Game Theoretic Decis. Theoretic Agents Workshop*, AAAI-02 (in press)
47 Rapoport, A. and Budescu, D.V. (1992) Generation of random series in two-person strictly competitive games. *J. Exp. Psychol. Gen.* 121, 352–363
48 Camerer, C.F. *et al.* (1993) In *Frontiers of Game Theory* (Binmore, K., Kirman, A. and Tani, P., eds) pp. 27–48, MIT Press
49 Andreoni, J. (1995) Warm-glow versus cold-prickle: the effects of positive and negative framing on cooperation in experiments. *Q. J. Econ* 110, 1–21
50 Bazerman, M. *et al.* (1985) Integrative bargaining in a competitive market. *Organ. Behav. Hum. Perform.* 35, 294–313
51 Camerer, C.F. and Karjalainen, R. (1994) In *Models and Experiments on Risk and Rationality* (Munier, B. and Machina, M., eds) pp. 325–358, Kluwer Academic Publishers
52 Camerer, C.F. and Lovallo, D. (1999) Overconfidence and excess entry: an experimental approach. *Am. Econ. Rev.* 89, 306–318
53 Camerer, C.F. (1990) In *Insights in Decision Making: Theory and Applications* (Hogarth, R., ed.), pp. 311–336, University of Chicago Press
54 Plott, C.R. (1987) In *Laboratory Experimentation in Economics: Six Points of View* (Roth, A.E., ed.), pp. 187–213, Cambridge University Press
55 Roth, A.E. (2002) The economist as engineer: game theory, experimental economics and computation as tools of design economics. *Econometrica* 70, 1341–1378