P000153

social preferences

Behaviour in a variety of games is inconsistent with the traditional formulation of egoistic decision-makers; however, the observed differences are often systematic and robust. In many cases, people behave as if they value the outcomes accruing to other reference agents. In reaction, behavioural economists have offered and tested a variety of formulations (such as inequality aversion and reciprocity) that capture the social nature of preferences.

For the longest time economists reacted allergically to preference formulations that allowed for anything but material self-interest (cf. Binmore, Shaked and Sutton, 1985). The reaction was well founded: by adding elements to the agent's utility function, potentially one allows economic theory to explain everything and, therefore, nothing. Any behaviour can be explained by assuming it is preferred. However, this strong position has sometimes made economics seem out of touch with the world economists try to explain. Even economists care about the outcomes achieved by others, in addition to their own outcomes. Moreover, they also care about how those outcomes are achieved. Only in 1982, however, was the weakness of taking material self-interest for granted demonstrated by Werner Güth and his coauthors, who showed that economic theory failed in the simplest of decision settings (Güth, Schmittberger and Schwarze, 1982), the ultimatum game. In this game a first mover offers a share of a monetary 'pie' to a second mover who either accepts the proposal, in which case it is divided as proposed, or rejects the proposal, in which case both players earn nothing. Since then this game has become the workhorse of experimenters intent on exploring carefully the extent to which people behave in ways that are contrary to their material self-interest.

While it is interesting to document the fact that people consider the outcomes of others when they make choices in experimental games, there are at least two other particularly compelling aspects of the research that has developed since the 1980s. First, these deviations from self-interest can be replicated, and have been, both inside and outside the laboratory. Replication suggests that these behaviours are not just errors or flukes, and therefore, although self-interest is a convenient modelling assumption, it should not be used as the basis for policy formulation. Second, this research illustrates that there is a difference between theory failing because of a false assumption and its failing because of flawed logic. Research shows that people do use economic reasoning, but that they, or most of them, are not narrowly self-interested.

The original results of the ultimatum game provided the impetus for a large body of research. Initially, some researchers were convinced that the explanation was not a concern for others but simple error (for example, Binmore, Shaked and Sutton, 1985). However, this explanation was soon swept aside by volumes of evidence from a variety of games that suggested that the payoffs of other players entered into the strategic choices of experimental participants (see the reviews of Bowles, 2004; or Sobel, 2005). Despite all this research, a precise definition of social preference has not been settled upon. In most cases, 'social preference' is defined loosely as *a concern for the payoffs allocated to other relevant reference agents in addition to the concern for one's own payoff*. (A largely separate branch of research has focused on altruism and warm glow motives for giving to others, especially in the context of public goods provision. This work is discussed elsewhere in the dictionary.)

Within the standard outcome-oriented definition, research has focused on identifying the more pro-social preferences for altruism and inequality aversion while considerably less attention has been given to their opposites, spite and eminence. The evidence from the hundreds of ultimatum games conducted since 1982 suggests that, on the second-mover side of the game, few people are willing to accept the low offers associated with the subgame perfect equilibrium prediction. In fact, offers of less than 20 per cent of the pie are routinely rejected, and as offers increase they are more likely to be accepted (Camerer, 2003). Turning down positive offers is clearly against one's material self-interest, but it is consistent with aversion to unequal payoffs (inequality aversion). As the stakes increase, the probability of a rejection falls, but even when the pie is as large as three months expenditures the rejection rate is not zero (Cameron, 1999).

Interpreting the motivation of the first mover in the ultimatum game is not as straightforward, though. One hypothesis is that proposers offer half the pie because they are inequality averse. We cannot, however, distinguish this reasoning from that of completely selfish, but astute, proposers who anticipate that low offers will be rejected and offer half because they know it will be accepted. The dictator game evolved to identify the motives of first movers (Forsythe et al., 1994). The dictator game is played just like the ultimatum game except for one very important design change: second movers are passive recipients of whatever they are allocated. In other words, they cannot reject offers. If the enlightened self-interest hypothesis is correct, we would expect to see first movers allocating nothing in the dictator game. This is not the case. Although allocations in the dictator game are susceptible to changes in the presentation of the game (Hoffman et al., 1994; Eckel and Grossman, 1996), it is common for people to allocate positive amounts. In fact, it is common for the behaviour of non-student participants in the two games to be indistinguishable (Carpenter, Burks and Verhoogen, 2005) suggesting that many people prefer equal outcomes.

There is some question as to whether the simple outcome-oriented definition of social preference is sufficient. An example illustrates why. Instead of offers being generated by other participants, imagine second movers in the ultimatum game being assigned offers randomly by a computer programme. If inequality aversion is a sufficient description of the motivations of participants, this change should have no impact on behaviour. However, it does: responders are much less likely to reject computer-generated offers than offers that come from real proposers (Blount, 1995). This indicates that people are also interested in the process and intentions that generate outcomes. The definition of social preference should perhaps be expanded accordingly to *a concern for the payoffs allocated to other relevant reference agents and the intentions that led to this payoff profile in addition to the concern for one's own payoff*.

Expanding the definition of social preference to include a process component allows us to also classify reciprocity – treating only kind acts with kindness – as a social preference. Pure reciprocity, however, is more elusive than inequality aversion because one needs to show that outcomes and intentions matter. Only a few experiments have been conducted to show that intentions matter, but the results are compelling. For example, imagine two binary choice versions of the ultimatum game (Falk, Fehr and Fischbacher, 2003). In game A, the proposer can decide between claiming the lion's share of a ten-dollar pie (8, 2) and sharing the pie equally (5, 5). In game B, the first option is the same (8, 2) but the second is even worse for the second mover because the proposer demands the whole pie (10, 0). Inequality aversion predicts that the (8, 2) offer will be rejected at the same rate in the two games

because the other offer is irrelevant – the decision-maker should focus only on the outcome presented. Reciprocity, on the other hand, suggests that one would be much less likely to reject (8, 2) in game B because it is the kinder of the two offers. Indeed, people are almost five times more likely to reject the (8, 2) offer in game A. An alternative approach is to compare the response of participants to different outcome allocations after another participant has made a kind or unkind act to the response when there is no initial move by another participant (Charness and Rabin, 2002). Reciprocity is identified by the subtraction of the first outcomes and intentions experiment from the second baseline inequality-aversion experiment.

In the trust (or investment) game, a first mover decides how much to send to a second mover. Any amount sent is multiplied by $k>1$ before it reaches the second-mover. The second mover then decides how much to send back. Because of the multiplication, sending money is socially efficient yet a first mover should send money only if she trusts the second mover to send back at least enough to cover the investment. The standard interpretation is that the first mover must expect the second mover to be motivated by reciprocity before it makes sense to invest in the partnership (Berg, Dickaut and McCabe, 1995). However, one can just as easily invoke inequality aversion to explain the fact that people tend to send back more when they receive more (Cox, 2004). The same problem exists with the related experiments developed to test for the notion of gift exchange in the labour market context (for example, Fehr and Schmidt, 1999).

Other, more indirect, evidence for reciprocity and the more nuanced definition of social preference comes from the experimental literature on voluntary contributions to public goods. In these settings participants are given an endowment and asked to decide how much to contribute to a 'group project'. The incentives are of a social dilemma; contributing nothing is a dominant strategy but contributing everything is socially efficient. Playing the public goods game in strategic form asks participants to decide how much they want to contribute conditional on the contributions of others. Half the participants are conditionally cooperative in that they generate contribution schedules that are increasing in the contributions of others (Fischbacher, Gächter and Fehr, 2001). The fact that people condition their contributions according to those of others suggests that intentions and reciprocity matter.

To identify reciprocity separately from inequality aversion one may employ a design in which the two forces pull in different directions. Imagine that one can punish free riders in the public goods game: a participant can impose a penalty p at a cost c. In most cases people punish despite it being dominant to free ride on the punishment done by others (that is, punishment is just a second-order public good), and this tends to stabilize contributions (Fehr and Gächter, 2000). However, in most cases $p>c$, which means cooperators reduce the inequality between themselves and the free rider by punishing. To isolate the role of, in this case negative, reciprocity one can allow $p<c$, which actually increases the inequality. Although they do it less often, people punish when the sanction delivered is lower than the cost, and this is a nice demonstration of reciprocity (Carpenter, 2007).

Several attempts have been made to organize the evidence on social preferences into parsimonious, but flexible, utility functions. One of the most successful outcome-oriented approaches is the Fehr and Schmidt (1999) specification, perhaps because it is relatively easy to work with. Here the utility of player i increases in her own payoff, x_i, but decreases in any difference between her payoff and the payoffs of other relevant players. For two-player games this is just:

$$u_i(x_i, x_j) = \left\{ \begin{array}{ll} x_i - \alpha_i(x_j - x_i) & \text{if } x_i < x_j \\ x_i - \beta_i(x_i - x_j) & \text{if } x_i \geq x_j \end{array} \right\}$$

where α_i is player i's degree of inferiority aversion and β_i is her degree of superiority aversion. It is natural to expect $\alpha_i > \beta_i$.

While this utility function is a good first approximation because it has been shown to be consistent with much of the experimental data (if one is willing to make assumptions about the distribution of α's and β's in the population) it is limited in two ways. First, as one can see in Figure 1, the predictions can be coarse. It is not hard to graph the indifference curves associated with the Fehr–Schmidt specification, but if one superimposes a budget constraint on the indifference mapping there are just two predictions: keep it all or give away half unless the constraint has exactly the same slope as the indifference curve, in which case any amount between nothing and half is possible.

The fact that intentions play no role is a second problem faced by all the outcome-oriented approaches. A trade-off does, however, exist because incorporating intentions makes the specifications considerably harder to work with. The outcome- and process-oriented specifications evolved from the notion of *psychological games*, which posits that utility will depend on both outcomes and beliefs (Geanakoplos, Pearce and Stacchetti, 1989). Beliefs are important because emotional responses are often triggered by expectations about how one should be treated. Perhaps the specification that is easiest to work with is the Charness–Rabin utility function, which incorporates a term θq to capture reciprocal motivations:

$$u_i(x_i, x_j) = (\rho r + \sigma s + \theta q)x_j + (1 - \rho r - \sigma s - \theta q)x_i.$$

The parameters r and s indicate which of the two players has the advantage ($r = 1$ if $x_i > x_j$, $s = 1$ if $x_j > x_i$ and $r = s = 0$ otherwise) and the parameters ρ and σ represent outcome-oriented preferences (Charness and Rabin, 2002). To recover the Fehr–Schmidt specification we simply assume $\sigma < 0 < \rho < 1$ and $\theta = 0$. Reciprocity and intentions are at work if $\theta > 0$ because we set $q = -1$ if player j has misbehaved and $q = 0$ otherwise.

Why should economists care about social preferences? By ignoring social preferences, economists have incompletely characterized many important interactions (Fehr and Fischbacher, 2002). Because many people are motivated by notions of fairness and reciprocity, social preferences can hinder the dynamics of competition that are assumed to drive equilibria, especially in the context of labour markets. For example, wages may never fall to the competitive equilibrium level because bosses understand that workers are reciprocally motivated. By lowering the wage, the boss also lowers morale and productivity (Bewley, 1999). Likewise, the economic theory of collective action is only narrowly applicable because it fails to realize that most people are predisposed to cooperate, but hate being taken advantage of (Andreoni,

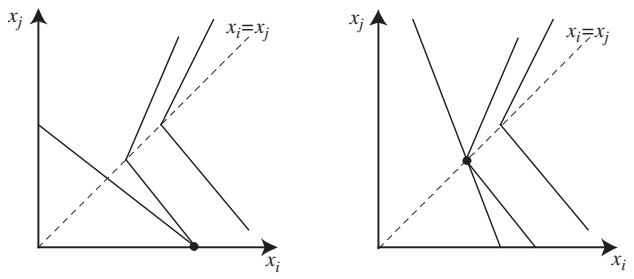

Figure 1

1988). Designing incentives is ultimately more challenging when one accounts for the heterogeneity of social motivations identified in economic experiments.

Future research on social preferences is likely to extend in a number of interesting directions. Experimenters have begun to move from the laboratory to the field to identify the preferences of more representative samples and to investigate the external validity of these preferences (that is, what important behaviours and outcomes do social preferences correlate with?). Within the laboratory it will be interesting to better isolate the role of outcomes versus the role of intentions, to examine the co-evolution of preferences and institutions, and to examine the difference between social preferences and social norms. Is it the case, for example, that norms dictate how one should treat others regardless of whether the prescribed behaviour is consistent with one's underlying preferences?

Jeffrey Carpenter

See also

< xref = A000240 > altruism in experiments;
< xref = E000020 > economic man;
< xref = P000350 > public good experiments;
< xref = xyyyyyy > trust in experiments.

Bibliography

Andreoni, J. 1988. Why free ride? Strategies and learning in public good experiments. *Journal of Public Economics* 37, 291–304.

Berg, J., Dickaut, J. and McCabe, K. 1995. Trust, reciprocity and social history. *Games and Economic Behavior* 10, 122–42.

Bewley, T. 1999. *Why Wages Don't Fall During a Recession*. Cambridge, MA: Harvard University Press.

Binmore, K., Shaked, A. and Sutton, J. 1985. Testing noncooperative bargaining theory: a preliminary study. *American Economic Review* 75, 1178–80.

Blount, S. 1995. When social outcomes aren't fair: the effect of causal attribution on preferences. *Organizational Behavior & Human Decision Processes* 62, 131–44.

Bowles, S. 2004. *Microeconomics: Behavior, Institutions and Evolution*. Princeton: Princeton University Press.

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

Cameron, L. 1999. Raising the stakes in the ultimatum game: experimental evidence from Indonesia. *Economic Inquiry* 37, 47–59.

Carpenter, J. 2007. The demand for punishment. *Journal of Economic Behavior & Organization* 62, 522–42.

Carpenter, J., Burks, S. and Verhoogen, E. 2005. Comparing students to workers: the effects of social framing on behavior in distribution games. In *Field Experiments in Economics. Research in Experimental Economics*, eds. J. Carpenter, G. Harrison and J. List. Greenwich, CT and London: JAI/Elsevier.

Charness, G. and Rabin, M. 2002. Understanding social preferences with simple tests. *Quarterly Journal of Economics* 117, 817–70.

Cox, J.C. 2004. How to identify trust and reciprocity. *Games and Economic Behavior* 46, 260–81.

Eckel, C. and Grossman, P. 1996. Altruism in anonymous dictator games. *Games and Economic Behavior* 16, 181–91.

Falk, A., Fehr, E. and Fischbacher, U. 2003. On the nature of fair behavior. *Economic Inquiry* 41, 20–6.
Fehr, E. and Fischbacher, U. 2002. Why social preferences matter – the impact of non-selfish motives on competition, cooperation and incentives. *Economic Journal* 112, 1–33.
Fehr, E. and Gächter, S. 2000. Cooperation and punishment in public goods experiments. *American Economic Review* 90, 980–94.
Fehr, E. and Schmidt, K. 1999. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 114, 769–816.
Fischbacher, U., Gächter, S. and Fehr, E. 2001. Are people conditionally cooperative? Evidence from a public goods experiment. *Economic Letters* 71, 397–404.
Forsythe, R., Horowitz, J., Savin, N.E. and Sefton, M. 1994. Fairness in simple bargaining experiments. *Games and Economic Behavior* 6, 347–69.
Geanakoplos, J., Pearce, D. and Stacchetti, E. 1989. Psychological games and sequential rationality. *Games and Economic Behavior* 1, 60–79.
Güth, W., Schmittberger, R. and Schwarze, B. 1982. An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization* 3, 367–88.
Hoffman, E., McCabe, K., Shachat, J. and Smith, V. 1994. Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior* 7, 346–80.
Sobel, J. 2005. Interdependent preferences and reciprocity. *Journal of Economic Literature* 43, 392–436.

Index terms

altruism
collective action
dictator game
fairness
free rider problem
gift exchange
inequality aversion
labour markets
psychological games
public good experiments
reciprocity
self-interest
social dilemma
social norms
social preferences
subgame perfection
trust game
ultimatum game

Index terms not found:

free rider problem
public good experiments
subgame perfection
trust game