CBESS Discussion Paper 11-06

Mutual advantage, conventions and team reasoning

by Robert Sugden*

* School of Economics, University of East Anglia, Norwich NR4 7TJ, United Kingdom, r.sugden@uea.ac.uk

Abstract
This paper proposes a conception of mutual advantage as a motivation for cooperative behaviour. This motivation is contrasted with the 'emotional' reciprocity that is represented in current theories of social preferences. The paper explores parallels between mutual advantage and Humean analyses of convention, and between mutual advantage and theories of team reasoning.

JEL classification codes
C72, D03, D63

Keywords
Mutual advantage; reciprocity; team reasoning

CBESS
University of East Anglia
Norwich NR4 7TJ
United Kingdom
www.uea.ac.uk/ssf/cbess

1. Introduction

The text for this paper is taken from the final paragraph of Antonio Genovesi's *Lectures on Commerce, or on Civil Economy* (2005 [1765-67]) . These lectures were delivered at the University of Naples, where Genovesi was the world's first professor of economics. Having taught his students how a commercial society works, he concludes:

> Here the idea of the present work. [We should study civil economy] to go along with the law of the moderator of the world, which commands us to do our best to be useful to one another. (p. 890; translation by Bruni and Sugden)

The key idea here, and a recurring theme throughout the lectures, is that in a well-ordered commercial society the motivation for economic relationships is *mutual assistance*, or *being useful to one another*. By viewing economic relationships as reciprocal, Genovesi assimilates them to other relationships of civil society.

This way of seeing markets is different from the one that most economists have taken. The usual view is encapsulated in one of the most famous passages in Adam Smith's *Wealth of Nations*:

> It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interest. We address ourselves, not to their humanity but to their self-love, and never talk to them of our own necessities but of their advantages. Nobody but a beggar chooses to depend chiefly upon the benevolence of his fellow-citizens. (1976 [1776], pp. 26–27)

On Smith's account, when individuals engage in market transactions, they are motivated by self-love, although subject to the constraints of justice. There is an explicit contrast between self-love on the one side and 'benevolence' and 'humanity' on the other. The same contrast appears in an even more famous passage from the same book:

> By ... directing that industry in such a manner as may be of the greatest value, [the merchant] intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention. Nor is it always the worse for society that is was no part of it. I have never known much good done by those who affected to trade for the publick good. (1976 [1776], p. 456)

The socially useful merchant, we are told, is motivated by 'his own gain'; this is contrasted favourably with the ironic picture of the benevolent merchant who intends 'the publick good'. Of course, Smith does not deny the reality of benevolence; his *Theory of Moral Sentiments* (1976 [1759]) offers a psychological and sociological explanation of both justice and benevolence. Nevertheless, the market is understood as a system of 'natural liberty' in which

individuals act on self-love. Taken as a whole, the workings of that system tend to produce socially beneficial outcomes; and the explanation of how this comes about involves the assumption that economic agents pursue their own interests. Further, as the remark about the beggar encourages us to think, transactions which express self-interest on both sides are compatible with individual independence and dignity in a way that transactions of benevolence may not be.

The main features of Smith's account of market motivation are now part of the conventional wisdom of economics (perhaps with the neoclassical proviso that economic agents are assumed only to be 'non-tuistic' – that is, not to take any interest in the welfare of those people with whom they trade). Against this background, Genovesi's text stands out as suggesting a different understanding of economic relationships, as motivated neither by self-interest nor benevolence, but by joint intentions for mutual advantage.

In a recent paper, Luigino Bruni and I have presented a sketch of what it might mean to view market relationships in this way (Bruni and Sugden, 2008). In this paper, I offer some further thoughts – still regrettably provisional – about how this sketch might be developed into a more formal theory. I shall suggest that this account of mutual advantage can be integrated with two other lines of thought on which I have been working for many years – the idea that individuals can reason 'as teams', and the idea that norms of cooperation can emerge and reproduce themselves spontaneously.

Intuitively, team reasoning seems a promising way of representing a motivation directed at mutual advantage, since team reasoning itself is neither self-interested nor benevolent; instead, it represents individuals as *reasoning together* about the achievement of common goals. However, the theory of team reasoning, at least in the form proposed by Bacharach (2006), has developed as an offshoot of classical game theory and has inherited that theory's focus on rational agency (while allowing rational agency to be attributed to groups). As a co-editor of Bacharach's unfinished magnus opus, I was drawn into his way of thinking about team reasoning, but my own inclinations are to be more sceptical about the role of reason in economic life.

Those inclinations are better expressed in my work on the evolution of conventions and norms (Sugden, 2004) , which draws on ideas from David Hume (1978 [1739–40]) and David Lewis (1969). One of the main themes of this work is that processes of social evolution tend to select conventions that can be construed as embodying principles of

reciprocity. These conventions require individuals to play their respective parts in practices that are mutually beneficial, but assessments of benefit are made relative to customary expectations which may themselves be conventional. Thus, the moral codes that form spontaneously around such conventions may not be capable of rational reconstruction as the imaginary outcomes of ideally rational deliberation. I would like to find a way of understanding the motivation of mutual advantage as conventional in this sense.

2. Emotional reciprocity

To explain the meaning of mutual advantage as a motivation, and how it differs from other motivations that are commonly assumed in economic analysis, I will use the Trust Game as my leading example. Social theorists have been discussing this game for centuries: versions of it can be found in Thomas Hobbes's *Leviathan* (1651/ 1962, Chs 14–15) and in David Hume's *Treatise of Human Nature* (1740/ 1978, pp. 520–521). Its modern incarnation as an experimental paradigm is due to Joyce Berg, John Dickhaut and Kevin McCabe (1995).

[Figure 1 near here]

I will discuss the simple version of the game shown in Figure 1. There are two players, A and B. A moves first, choosing between *hold* and *send*; *send* is interpreted as an act of trust. If A chooses *send*, B chooses between *return* and *keep*; *return* is interpreted as the repayment of A's trust. The payoffs to each combination of actions are shown at the end of each path through the game tree, A's payoff first. These are specified in 'material' units, such as money; there is a presumption that, other things being equal, players prefer larger payoffs to smaller ones, but whether this is their *only* motivation is left open. Notice that B gets a higher payoff from (*send*, *keep*) than from (*send*, *return*), and so has a temptation not to repay A's trust if it is given. If A expects that B will not repay, his payoff-maximising strategy is *hold*. However, both benefit from the combination of trust and repayment: (*send*, *return*) gives both A and B a higher payoff than *send*.

This game has been the subject of many experiments. The typical finding is that substantial proportions of A-players choose to send, and that of those B-players who have been sent to, substantial proportions choose to return. It is natural to say that the B-players are revealing some motivation to repay trust, even though doing so is contrary to their immediate self-interest, and that A-players are revealing either a motivation to engage in trust

or an expectation that B-players will repay trust (or both). The problem for economic theory is to explain exactly what these motivations are.

The earliest economic theories of non-selfish behaviour represented the relevant motivations as *altruistic*. An altruistic player derives utility both from her own material payoffs and from those of her co-player. In the Trust Game, a sufficiently altruistic B will return, and an A (whether altruistic or self-interested) who expects this will send. A more recent variant of this kind of theory assumes that individuals are averse to inequality in the distribution of payoffs (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). If B is sufficiently averse to inequality, she will return, and if A expects this, he will send. However, this kind of explanation fails to capture the relational nature of trust. If B was motivated only by altruism or inequality aversion, it would not matter to her what had happened in the game before it was her turn to move – or, indeed, whether the game had any previous moves. But the idea of *repaying* trust presupposes a concept of trust itself, and that must refer to A's action.

Theorists of social preferences often try to deal with this issue by taking account of players' *intentions*. The intuitive idea is that B's motivation to return is a response to the intentions that are revealed in A's decision to send. In the literature of social preferences, most analyses of intentions use a conceptual scheme first proposed by Matthew Rabin (1993). In this scheme, as applied to a two-player game, an action by one individual (say A) is categorised as 'kind' or 'unkind' towards the other (B) in terms of that action's effects on the two players' payoffs. Developing this idea in general raises some difficult problems, but the essential concept of kindness is closely related to altruism. A's action is deemed to be kind to B if it increases B's payoff at some cost to A; thus, an act of kindness by A is prima facie evidence that A derives utility from B's payoffs. Rabin's hypothesis is that individuals are motivated by kindness towards people who are being kind to them, and by unkindness towards those who are being unkind to them. This hypothesis is now widely used in behavioural economics to represent the idea that individuals are concerned about reciprocity.

Strictly, Rabin's theory applies only to normal-form games. Viewed as a general problem, extending the theory to games in which players move sequentially is not straightforward (for one possible solution, see Dufwenberg and Kirchsteiger, 2004). But it seems that any reasonable extension of Rabin's theory will have the following implication for the Trust Game: there is no equilibrium in which the path (*send*, *return*) has probability 1.

To see why not, suppose that A chooses *send*, believing that B will choose *return* with probability 1. Then B's choice is between returning, which induces the payoff vector (1, 1), and keeping, which induces (–1, 3). Reasons of self-interest point towards the latter, but we need to take account B's judgments about A's intentions. Given that A expected B to return, his choice was between (0, 0) and (1, 1). According to Rabin's definitions, choosing (1, 1) is from this menu was neither kind nor unkind. Given the underlying logic of the theory, this conclusion seems unavoidable: A's behaviour was entirely consistent with what might seem to be the most natural default hypothesis, namely that he has neither positive nor negative concern about B's payoff. Given that A's choice of *send* is at the zero point of the kindness/ unkindness scale, B receives neither utility nor disutility from being kind or unkind towards A. Thus, B will be motivated only by material payoffs, and will choose *keep*. So A's belief about B is false. This establishes that (*send*, *return*) is not an equilibrium.

Rabin recognises that his theory has this implication, and that this is unsatisfactory. Discussing a normal-form version of a Trust Game, he suggests that, contrary to the implications of his model, 'it seems plausible that cooperation would take place'. His response to this problem is to say, reasonably enough, that his model is not intended to represent all psychological factors that can affect behaviour in games; theorists may need to consider modelling 'additional emotions' (p. 1296). Still, what Rabin is saying is that the particular kind of *emotional reciprocity* that appears in his model does not provide an adequate explanation of trust. Since reciprocity is surely integral to the concept of trust, the implication is that there is more to reciprocity than returning kindness for kindness and unkindness for unkindness.

Why does the hypothesis of emotional reciprocity fail to explain trust? The fundamental reason, I suggest, is that the concept of emotional reciprocity rests on the presupposition that the paradigm form of moral motivation is altruism; and altruism is not a reciprocal concept. In the first economic theories of non-selfish behaviour, individuals were simply assumed *to be* altruistic. In Rabin's theory, and in the theories that have developed from it, individuals are sensitive to one another's intentions; but when one player assesses the moral status of her co-player's intentions, the question she asks is whether the co-player is altruistic. As the analysis of the Trust Game shows, this kind of scrutiny of intentions can be fatal for the achievement of cooperation.

A genuinely cooperative practice, such as the combination of sending and returning in the Trust Game, is beneficial to both parties. When such a practice is well-established, each

party expects the other to participate in it. In many cases, given this expectation, one party's participation in the practice is in the interests of both – as, in the Trust Game, is A's participation in the practice of sending and returning. But does the fact that A benefits from his participation discharge B from the moral obligation to participate too? To think in this way is to take a childish view of the world. It is to think that social life is structured around unilateral acts of kindness and unkindness – of giving to others and taking from others – and that as moral agents, individuals scrutinise one another's intentions, unilaterally rewarding those that are good and punishing those that are bad. As Smith is pointing out in his remarks about the butcher, the brewer, the baker and the beggar, these are not the right moral categories for a society based on cooperation between free and equal individuals.

3. Mutual advantage

To understand trust, we need a concept of *cooperative* intention that is not grounded in altruism, but instead is reciprocal all the way down. As a first approximation, here is a different way of thinking about the Trust Game.

In that game, it is uncontroversial that if A and B both acted on self-interest, with mutual knowledge that they were so motivated, the outcome would be (0, 0). So we can treat that outcome as a non-cooperative default. Relative to that benchmark, the combination (*send*, *return*) is a practice that benefits both players. Further, it benefits them equally in material terms (each gains one unit relative to the benchmark). At least if we restrict attention to pure strategies, no other feasible combination of actions benefits both players. If one asks the question 'What would be a fair cooperative solution to the Trust Game?', it seems obvious that the answer is (*send*, *return*). As another way of thinking about this question, imagine that (as in the scenarios considered in cooperative game theory) the two players were able to make binding agreements about which strategies they would play. In this scenario, A has the sole power to bring about the outcome (0, 0), while A and B together have the joint power to bring about any of the three outcomes. If they were at all rational, they would surely see the logic of settling on (1, 1), the game's unique core solution.

So suppose that A and B both recognise that (*send*, *return*) is the fair cooperative solution to the game, and that their recognising this is common knowledge between them. This is not enough to guarantee that A sends and B returns, because the players may have motivations other than the achievement of fair cooperation. More specifically, if A sends, it

is in B's interest to keep; in B's motivational scales, self-interest might outweigh concern for fairness. If A believes that B will keep, he has a self-interested reason to hold. Thus, if the fair cooperative solution is to be reached, B must resist the temptation to act on self-interest, and A must have sufficient confidence that she will resist.

But suppose that A *does* send. This is not a signal of kindness on A's part, but is it is a signal *of trust*. The obvious interpretation is that A sees himself as participating in the joint action (*send*, *return*). In other words, he is playing his part in a fair cooperative practice, expecting B to play hers. Thus, for B to return is for her to play *her* part in a fair cooperative practice, knowing that A has already played his. In returning, she reciprocates A's intention of fair cooperation.

Notice that in this account, neither player is sacrificing his or her interests out of kindness for the other. Instead, each is playing his or her part in a joint action that benefits them both. In this sense, each intends that *they both benefit*. Each player's recognition that the other will benefit from the joint action is not corrosive of the motivation to participate, as it is in Rabin's account of emotional reciprocity. To the contrary, each player's belief that the other intends that they both benefit supports his or her own motivation to participate.

I suggest that this discussion of the Trust Game points the way towards a more general analysis of a motivation of mutual advantage. I now present a very rough sketch of what I have in mind. This sketch combines elements from a number of my previous papers (e.g. Sugden, 1984; 1993; 2003), but the combination is new. It should be thought of as a blueprint for theory that has yet to be built.

Consider a group of individuals $i = 1, ..., N$, who interact in some noncooperative game (which could be a one-shot game, or could be one that is played recurrently in some larger population). For each individual i, there is a set of alternative *strategies*. A *joint action* is a profile $s = (s_1, ..., s_N)$ of strategies, one for each individual. Suppose that one such profile s^0 can be interpreted as the noncooperative default. (Exactly how this default should be defined is one of the problems that needs to be resolved in constructing an adequate model.) Now consider some other profile s^*. Suppose that the following conditions hold for some individual i:

Mutual Gain. If s^* is played rather than s^0, every individual benefits.

Fairness. The benefits gained by playing s^* rather than s^0 are distributed reasonably fairly between the N individuals. (The definition of 'reasonably fair' is another problem to be resolved.)

Assurance. Individual i has reason to expect that every other individual j will play her component s_j^* of the joint action s^*.

Then if i is sufficiently motivated by *mutual advantage*, she will play s_i^* with the intention of participating in fair cooperation.

So far, this is just a definition of 'mutual advantage' as a motivation. But I want to claim that mutual advantage is a useful model of a motivation that activates a significant proportion of real people in real social interactions. This claim has a similar status to those that, in the existing literature of social preferences, are made on behalf of (for example) altruism, inequality aversion or emotional reciprocity.

To understand the implications of mutual advantage, it is important to recognise the role played by the Assurance condition. Because of this condition, an individual who is motivated by mutual advantage is motivated to participate in fair cooperative joint actions *if she has reason to expect that others will participate too*. This builds reciprocity into the structure of the motivation. The Assurance condition also has the effect of aligning the motivation of mutual advantage with ongoing cooperative practices. Notice that the Mutual Gain and Fairness conditions are not optimising principles which seek out the most efficient and fairest strategy profile in the game. Rather, they are satisficing principles which test whether a given strategy profile is 'good enough'.

How might the expectations required by the Assurance condition be generated? One possibility, described in some theories of team reasoning, is that each individual endorses and acts on certain *general* principles of game-theoretic reasoning which, when applied to particular types of game, induce coordination on particular strategy profiles which satisfy the Mutual Gain and Fairness conditions. Common knowledge that this is the case might be generated by repeated experience of successful coordination (Sugden, 2003). On one reading of Thomas Schelling's (1960) theory of focal points, each person's experience of games teaches her that other people tend to act on the principle of choosing the most salient Nash equilibrium. (For a defence of this reading of Schelling, see Sugden and Zamarrón, 2006.)

A more mundane possibility is exemplified by many everyday market transactions, such as those that Smith has in mind in his discussion of how we expect to get our dinners. If

I buy bread from Smith's baker, he and I participate in a joint action in which I transfer money to him and he transfers bread to me. Relative to the most obvious default, namely our not trading, this joint action benefits both of us. For the sake of the argument, let us take it as given that the baker and I both believe that the price he is asking is fair. Suppose the baker hands the bread to me before asking me to pay. His expectation that I will pay is grounded on a whole complex of inter-related considerations. If I don't pay, I will have to go somewhere else next time I want some bread. Not paying is contrary to social norms; and it is punishable by the criminal law. So, all things considered, it is almost certainly in my interest to pay. Further, there is an ongoing convention in our society according to which, if a person asks a shopkeeper for a product that is on sale, he thereby agrees to pay for it, and does in fact do so. By inductive inference from this general experience, the baker has good reason to expect me to behave as most other customers do. The upshot of all this is that the transaction does not *require* that the baker and I are motivated by mutual advantage – which, of course, fits with what Smith says about the relationship between the baker and the customer. Nevertheless, we *can* act as we do, motivated by mutual advantage. Each of us can construe the transactions we make in the market as fair cooperation, as being useful to one another.

In other cases, mutual expectations are generated by experience of an ongoing practice, unsupported by self-interest or by threats of punishment. For example, among British road-users there is a well-established custom by which, at busy times, drivers on main roads intermittently allow drivers from minor roads to enter the flow of traffic ahead of them. Because most British road junctions are regulated by fixed 'give way' signs rather than by traffic lights or roundabouts, the road network would seize up if there were not at least a sizable minority of drivers willing to behave in this way. But there is no legal requirement for this behaviour, and no firm rules about when a main-road vehicle should give way to a minor-road one. Considering the practice as a whole, and comparing it with the default situation in which main-road drivers always insist on their legal rights to priority, it seems reasonable to suppose that it works to the general benefit of all drivers. Since most people use both major and minor roads, and since each act of giving way induces only incremental costs and benefits, it seems equally reasonable to judge that the distribution of benefits is reasonably fair. (One might object that the costs fall unfairly on the more courteous drivers, but any such distributional effect is probably very small. When a main-road driver gives way, he usually bears only a small proportion of the total costs of his action; most of the costs

are borne by drivers behind him, whether willingly or unwillingly.) And each driver's experience of the roads tells him that the practice is generally followed. On my analysis, if an individual is sufficiently motivated by mutual advantage, this combination of circumstances will induce him to participate in the practice. When he gives way to another driver, he need not think of this as a gratuitous act of kindness. He can think of himself as playing a fair part in a practice that works to everyone's benefit, including his own.

Thus, the concept of mutual advantage allows us to perceive a moral alignment between ordinary market transactions and more obviously pro-social types of behaviour. Instead of understanding the former as self-interested or non-tuistic and the latter as altruistic, we can think of both as mutually advantageous cooperation.

4. Mutual advantage, conventions and team reasoning

On the account that I have offered, practices of mutual advantage are *conventional*. The root meaning of 'convention' is 'coming together', as in a political convention or in the idea of convening a meeting. By extension it has come to refer to the agreements that are made when people come together, and to practices that can be interpreted as tacit agreements, or as resting on common consent. Hume (1739–40 [1978]) seems to have had the idea of tacit agreement in mind when he argued that justice (in the sense of respect for property) was a convention:

> This convention is not of the nature of a *promise*: For even promises themselves, as we shall see afterwards, arise from human conventions. It is only a general sense of common interest; which sense all the members of the society express to one another, and which induces them to regulate their conduct by certain rules. I observe, that it will be for my interest to leave another in the possession of his goods, *provided* he will act in the same manner with regard to me. He is sensible of a like interest in the regulation of his conduct. When this common sense of interest is mutually express'd, and is known to both, it produces a suitable resolution and behaviour. And this may properly enough be call'd a convention or agreement betwixt us, tho' without the interposition of a promise; since the actions of each of us have a reference to those of the other, and are perform'd upon the supposition, that something is to be perform'd on the other part. (p. 490)

Notice how, for Hume, conventions are mutually beneficial practices which, in principle, *could have been* the result of reasonable agreement among individuals, even if in fact they have emerged spontaneously. The same idea is expressed by my Mutual Gain and Fairness conditions. Notice also that, on Hume's account, each individual plays his part in a convention on the supposition that others will play theirs, in the same way that each party to a

contract performs his part of the contract in the expectation that others will perform theirs. The same idea is expressed by my Assurance condition.

There is an affinity between this way of thinking about convention and the idea of team reasoning. Classical game theory can be thought of as an attempt to answer the question 'What it is rational *for me* to do?', asked by each player of a game, when each knows that the others are asking the same question. The starting point for the theory of team reasoning is the thought that the players might equally legitimately ask 'What is it rational *for us* to do?' (Sugden, 1993, 2003; Bacharach, 2007). When players engage in team reasoning, they look for the *profile* of strategies that leads to the best outcome for them collectively. Each player then chooses his component of that profile, construing that choice as his part of a joint action that is rational for the players as a collective or 'team'.

In theories of team reasoning, the idea of 'rationality for us' is usually modelled as the maximisation of some team objective, on the analogy of the conventional analysis of individual rationality. Clearly, that would not always be consistent with the concept of mutual advantage that I am now proposing. But the question 'What is it rational for us to do?' might be asked in something more like the sense of cooperative game theory. That is, the players who ask the question might be thought of as looking for the terms of an agreement that it would be rational for each to make, if all could be assured that the agreement would be kept. Correspondingly, the game theorist who imagines herself being called on to answer the question must picture herself, not as a coach instructing the members of a team about how best to achieve a common aim, but as the neutral chair of a negotiation session, trying to orchestrate an agreement that each negotiator can be satisfied with.

Despite the conceptual differences between these two interpretations of the question 'What is it rational for us to do?', most of the cases to which the theory of team reasoning has been applied are ones in which the two interpretations effectively coincide. When Bacharach (2006, pp. 58–64) discusses the problem of specifying the team objective of a group of players, he postulates that this is 'Paretian' with respect to individuals' private preferences. That is, for any pair of strategy profiles s' and s'', if every member of the group has a weak preference for the outcome of s' over that of s'' and if, for at least one member, this preference is strict, then the team's objective ranks s' above s''. Defending this assumption, Bacharach argues that in the (two-player) games in which people see team reasoning as most obviously rational, the most salient feature of the outcome it recommends is that it 'is *best for*

both players' (p. 62). Such an outcome would also be the obvious outcome of rational negotiation.

The two main games that Bacharach uses to explore the implications of team reasoning are Hi-Lo and the Prisoner's Dilemma. In the Hi-Lo game, there are two pure-strategy Nash equilibria, one of which strictly dominates the other; both equilibria strictly dominate every other outcome. In this game, the Paretian criterion identifies a unique team-optimal profile; this is also the unique core solution. In his analysis of the Prisoner's Dilemma, Bacharach assumes that the profile in which both players cooperate is team-optimal (pp. 132, 168–171). Although this profile is not singled out by the Paretian criterion, it is the unique core solution of the game. (Intuitively, if binding agreements are possible, mutual cooperation is the only profile on which the players could plausibly agree, given that each has the unilateral power to play the defection strategy.) I suggest that, even when Bacharach is using the concept of a team objective as a modelling device, some of the intuitions on which he is drawing are more naturally expressed in the language of cooperative game theory.

However the question 'What is it rational for us to do?' is interpreted, the mere idea that the question is relevant marks a fundamental departure from theories of social preferences. Unlike theories of social preferences, theories of mutual advantage and team reasoning do not explain socially-oriented behaviour in terms of individuals severally maximising utility functions whose arguments include 'social' variables. Instead, they explain each individual's behaviour as her part of a joint action. That action is the collective choice (however construed) of a group of which the individuals are members.

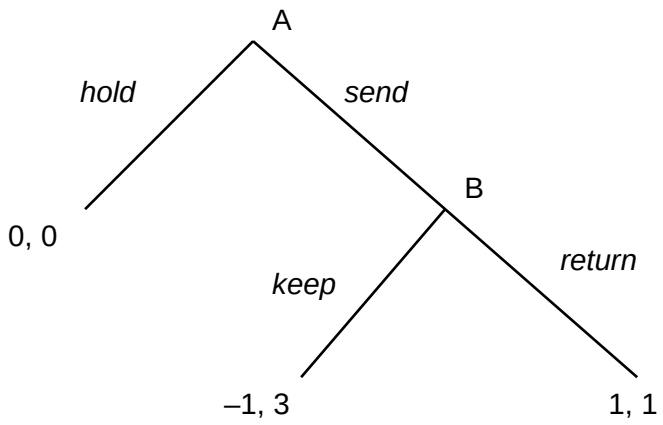
5. Conclusion

I have presented my analysis of mutual advantage as a model that may be useful in explaining cooperative behaviour in the real world. But I also want to declare my support for mutual advantage as a normative principle – whether or not it is the principle on which the majority of my fellow-citizens currently act. When I described emotional reciprocity as 'childish', what I meant was that the emotional repertoire modelled by Rabin's theory is one that is appropriate to a social environment in which the individual has strong affective relationships with others, but in which cooperation with others on equal terms is not a central activity. Conversely, the emotional repertoire of mutual advantage is appropriate to, and

likely to be fostered by, an environment in which individuals continually face the need to coordinate their actions with other people – often people with whom they have only weak affective ties. Mutual advantage is a motivation that coheres with a liberal ideal of society as a cooperative venture among free and equal individuals.

Acknowledgements This paper is based on a lecture that I gave at a 'Workshop in Honour of Robert Sugden' held at the Faculty of Political Sciences of the University of Naples 'Federico II' in May 2010. I am very grateful to Sergio Beraldo, Luigino Bruni and everyone else who contributed to the success of this workshop. I was touched that so many of my present and former research students, and so many people with whom I have worked over the years, wanted to participate. The theme of the lecture is a conception of economic and social life as a scheme of cooperation, rather than (as it is all too often seen as) a positional competition. In choosing this topic, I wanted to express my appreciation of the ethos of cooperation, friendship and commitment to the discovery of truth that characterises the community of researchers to which I am proud to belong.

Figure 1: The Trust Game



References

Bacharach, M (2006). *Beyond Individual Choice.* Natalie Gold and Robert Sugden (eds). Princeton University Press, Princeton.

Berg J, Dickhaut J, McCabe K (1995). Trust, reciprocity, and social history. *Games and Economic Behavior* 10: 122–142.

Bolton G, Ockenfels A (2000). ERC: A theory of equity, reciprocity and competition. *American Economic Review* 90: 166–193.

Bruni L, Sugden R (2008). Fraternity: why the market need not be a morally free zone. *Economics and Philosophy* 24: 35-64.

Dufwenberg M, Kirchsteiger G (2004). A theory of sequential reciprocity. *Games and Economic Behavior* 47: 268–98.

Fehr E, Schmidt K (1999) A theory of fairness, competition and cooperation. *Quarterly Journal of Economics* 114: 817–868.

Genovesi A (2005 [1765–67]) *Delle lezioni di commercio o sia di economia civile*. Instituto Italiano per gli Studi Filofící, Napoli.

Hobbes T (1961 [1651]) *Leviathan*. Macmillan, London.

Hume D (1978 [1739–40]) A treatise of human nature. Oxford University Press, Oxford.

Lewis D (1969) Convention: a philosophical study. Harvard University Press, Cambridge, MA.

Rabin M (1993) Incorporating fairness into game theory and economics. *American Economic Review* 83: 1281–1302.

Schelling T (1960) *The strategy of conflict*. Harvard University Press, Cambridge, MA.

Smith A (1976 [1759]) *The theory of moral sentiments*. Clarendon Press,Oxford.

Smith A (1976 [1776]) *An inquiry into the nature and causes of the wealth of nations*. Clarendon Press, Oxford.

Sugden R (1984) Reciprocity: the supply of public goods through voluntary contributions. *Economic Journal* 94: 772–787.

Sugden R (1993) Thinking as a team: toward an explanation of nonselfish behavior. *Social Philosophy and Policy* 10: 69–89.

Sugden R (2003) The logic of team reasoning. *Philosophical Explorations* 6: 165–181.

Sugden R (2004) *The economics of rights,cooperation and welfare*. Second edition. Palgrave Macmillan, Basingstoke. First edition 1986.

Sugden, R, Zamarrón I (2006) Finding the key: the riddle of focal points. *Journal of Economic Psychology* 27: 609–621.

Fischbacher U (2007) z-Tree: zurich toolbox for ready-made economic experiments. Exp Econ
322 10:171–178
323 Gold N, Sugden R (2007) Collective intentions and team agency. J Philos 104:109–137
324 Ledyard JO (1995) Public goods: a survey of experimental research. In: Roth A, Kagel J (eds) Handbook
325 of experimental economics. Princeton University Press, Princeton, pp 111–194
326 Lyubomirsky S, King L (2005) The benefits of frequent positive affect does happiness lead to success?
327 Psychol Bull 131:803–855