Economics, Experimental Methods in

Advanced article

Vernon L Smith, George Mason University, Arlington, Virginia, USA

CONTENTS

Experimental economics uses laboratory methodology to examine motivated human behavior and its interpretation in small group interactive games, and in bidding, auctioning, and market institutions. Subjects earn cash payments depending upon their joint interactive decisions, and the rules governing their interactions.

INTRODUCTION

It is useful to distinguish three complex self-ordering systems: the internal order of the mind (Hayek, 1952); the external order of social exchange between minds (McCabe and Smith, 2001); and the extended order of cooperation through markets and other cultural institutions (Hayek, 1988). Our concern here is with the first two.

We focus on social exchange because it was the cooperative behaviors registered in two-person anonymous interaction that first alerted experimental economists to a significant class of phenomena that violate certain static equilibrium concepts in game theory. Game theory is about strategic interdependent choice when the pay-off benefit to each of two or more people depends jointly on their decisions. These refutations generated alternative interpretations of that theory, and motivated questions directly concerned with the internal order of the mind. They are now leading to the study of the neural correlates of human decision-making in two-person strategic interactions.

Why do we study anonymous interactions? First, our theoretical model of single-play games assumes strangers without a history or a future, and anonymity provides the required control for testing this theory. Also, it is well documented that the effects of face-to-face interaction hide more subtle procedural effects in yielding cooperative outcomes (Hoffman and Spitzer, 1985). As will be illustrated in the experiments reported below, the anonymity condition provides great scope for exploring the natural human instinct for social exchange, and how it is affected by context, reward, and procedure.

Why Should Context Matter?

Context matters because all memory involves relationships and is associative. For example, priming experiments use cues to improve retrieval from memory because of associations between the cue and the stimulus. People perform better at completing word fragments (filling in missing letters) on words they have observed beforehand in lists, even if they are not told that the words appeared in the earlier list. Some priming effects are almost equally strong whether the interval between the original list and the test is a matter of hours or days. Furthermore, being able to state that a word was seen before does not correlate with improved completion performance. How one perceives a current task depends upon unconscious cues to past experience that are triggered by the context of the task. Two decision tasks with the same underlying logical structure may lead to different responses because they are embedded in different contexts and invoke different memory experiences (Gazzaniga *et al.*, 1998, pp. 258–261). This is because of the fundamental, though nonintuitive, nature of perception.

In the early 1950s Hayek articulated certain principles of perception, which are consistent with current neuroscientific understanding:

1. It is incorrect to suppose that experience is formed from the receipt of sensory impulses reflecting unchanging attributes of external objects in the physical environment. Rather, the process by which we learn about the external environment involves a relationship between current conditions and our past experience of similar conditions (Hayek, 1952, p. 165).
2. Categories are formed by the mind according to the relative frequency with which current perception and memory (past perceptions) concur (p. 64). What are

stored in memory are external stimuli modified by processing systems whose organization is conditioned by past experience of stimuli. All perception is produced by memory.

3. This leads to a 'constant dynamic interaction between perception and memory, which explains the … identity of processing and representational networks of the cortex that modern evidence indicates' (Fuster, 1999, p. 89). 'Although devoid of mathematical elaboration, Hayek's model clearly contains most of the elements of those later network models of associative memory … [and] comes closer, in some respects, to being neurophysiologically verifiable than those models developed 50 to 60 years after his.' (Fuster, 1999, pp. 88–89).

Hayek's model is incomplete, and did not influence the research it anticipated, but it captures the idea that perception is self-organized, created from abstract function combined with experience. This is relevant to the question of why context is important in the experiments reported below.

MENTAL MODULES AND EVOLUTIONARY PSYCHOLOGY

Evolutionary psychologists argue that the mind consists of circuitry, or interactive modules, that are specialized for vision, for language learning, for socialization, and for a host of other functions (Cosmides and Tooby, 1992). Language and socialization, which are of recent evolutionary origin, are hypothesized to have evolved in the two to three million years during which humans subsisted as hunter–gatherers. It is in this evolutionary environment of our ancestors (EEA) that humans developed mechanisms of social exchange in which assistance, meat, favors, information and other services and valuables were traded across time. This is evident in extant hunter–gatherer societies (e.g. the Ache of Paraguay) in which the product of the hunt is widely shared within the tribe as well as within the nuclear and extended family. In a world without refrigeration and only rare success in hunting, this made sense: if I am lucky in the hunt today, I share the meat with others; and tomorrow, when I fail to make a kill, you share your kill with me and with others. In contrast, the products of gathering – fruit, nuts, roots – depend more on effort than on luck, are much more predictable from day to day, and are shared only in the nuclear family where effort can be closely monitored. Traditions of sharing across time provide gains from exchange that support limited forms of specialization: women and children do the gathering; adult men do the hunting; older men make tools, advise in the hunt, and assist in gathering. Such patterns (subject to numerous variations) are common in tribal communities.

But delayed exchange across time based on reciprocity is hazardous. Favors cannot be retracted, and you might systematically fail to return mine. Without money – a recent invention not available in the EEA – it is adaptive to develop some skill in making judgments about who can or cannot be trusted. This puts a premium on 'mindreading', the ability to infer mental states from the words and actions of others. The minimal mental equipment required is a 'cheater-detector module' for social exchange. The results of experiments designed by Cosmides (1985) are consistent with the hypothesis that the human mind is attuned to detecting cheaters on perceived social contracts. With the development of language, our instincts for cheater detection were enhanced by gossip: comparing notes to determine those with good reputations for returning favors. Gossip, like language and reciprocity, is a human universal, an activity pursued in all human communities. None of this mental equipment was the product of our reason: rather, it was the unconscious product of the biological and cultural evolution that distinguished us from other primates.

Evolutionary psychologists see an inevitable tension between who we are (based on what we have inherited from the EEA) and the demands made on us by the world since the agricultural revolution 10,000 years ago. One account of this tension was articulated by Hayek: 'Part of our present difficulty is that we must constantly adjust our lives, our thoughts and our emotions, in order to live simultaneously within different kinds of orders according to different rules. If we were to apply the unmodified, uncurbed rules (a caring intervention to do visible good) … of the small band or troop, or of, say, our families, to … our wider civilization (the extended order of the market), as our instincts … often make us wish to do, *we would destroy it*. Yet if we were always to apply the rules of the extended order (action in the self-interest within competitive markets) to our more intimate groupings, *we would crush them*. So we must learn to live in two sorts of world at once.' (Hayek, 1988, p. 18).

This observation raises questions about game theory, which postulates that the players are strictly self-interested and that this condition is common knowledge to all the players. How is action in the self-interest affected by whether the anonymous players are in an *n*-person market or a two-person interactive game? How do the players come to have 'common knowledge'? Does the procedural and instructional context of a two-person game affect cooperation by influencing how the players perceive the game? It is most natural to investigate such questions in experimental environments where monetary pay-offs, context, and interaction procedures can be controlled.

EXPERIMENTAL PROCEDURES

The experiments reported below show that context is important in the decision behavior we observe. This is to be expected, given what is known about the autobiographical character of memory and the interaction between current and past experience in creating memory. Below are reported behavioral results in two-person sequential-move game trees in which each pair plays once and only once through the move sequence defined by the tree, and the game is completely known to the subjects. However, the instructions for the experiments do not (except in systematic treatments) use words like 'game', 'play', 'player', 'opponent', or 'partner'; rather, reference is made to the 'decision tree', 'decision maker' 1 (DM1) or 2 (DM2), 'your counterpart', and other terms designed to provide a baseline context.

Your experience as a subject in a typical experiment might be as follows. You have been recruited to participate in an economics experiment for which you will be paid $5 (or more, in some cases) for arriving on schedule, plus the amount in cash that you earn from your decisions, to be paid to you at the end. You arrive, sign in, receive $5, and are assigned to a computer terminal in a large room with 40 stations. There are 11 other people, well spaced throughout the room. Each station is a partially enclosed booth, making it very easy to maintain your privacy. After everyone has arrived you log in to the experiment as directed on your screen. You read through the instructions for the experiment at your own pace, respond to the questions, and learn that in this experiment you are matched anonymously with another person in the room, whose identity you will never know, and vice versa. This does not mean that you know nothing about that person: it may seem evident that he or she is another 'like' person – for example, an undergraduate – with whom you may feel more or less of an in-group identity. Obviously, you bring with you a host of past experiences and impressions that you are likely to apply to the experiment.

Ultimatum Game Experiments

Consider the following simple two-stage two-person game. A fixed sum of money m is provided by the experimenter (e.g. m might be 10 one-dollar bills, or 10 ten-dollar bills). Player 1 moves first, proposing that a portion $x \leq m$ of the money be offered to player 2, player 1 retaining $m - x$. The offer is a 'take it or leave it' ultimatum. Player 2 then responds by either accepting the offer, in which case the experimenter pays $m - x$ to player 1 and x to player 2, or rejecting the offer, in which case each player receives 0.

Now consider four different instructional–procedural contexts in which an ultimatum game with this underlying abstract structure is played. In each case, imagine that you are the first mover (player 1 in the above abstract form). (See Hoffman *et al.* (2000) for instructional details, and for references to the literature and origins of the ultimatum game.)

Context 1: 'divide $10'

In the first context, the instructions state that you and your anonymous counterpart have been 'provisionally allocated $10'. Your task is to 'divide' the $10 using the following procedure. You have been randomly assigned to the role of first mover. You (as person A) are asked to complete boxes (4) and (5) of the proposal form shown in Figure 1. The form then goes to your counterpart (person B) who checks 'Accept' or 'Reject'.

In this version, the $10 consists of 10 one-dollar bills. In another version, there is $100 (10 ten-dollar bills) to be divided.

Context 2: 'contest entitlement'

In the second context, each of the 12 people in the room takes the same general-knowledge quiz (10 questions). The results are used to determine the positions of persons A and B in each pairing. Your score is the number of questions answered

(1) Identification number..	#A
(2) Paired with..	#B
(3) Amount to divide...	$10
(4) Person B receives	
(5) Person A receives (3)–(4)	
(6) Accept	Reject

Figure 1. Proposal form for an ultimatum game experiment using the 'divide $10' context. You (as person A) are asked to complete boxes (4) and (5); the form then goes to your counterpart (person B) who checks 'Accept' or 'Reject'.

correctly, with ties broken in favor of the person who finished the quiz fastest. The scores are ranked from 1 (highest) to 12 (lowest). Those ranked from 1 to 6 will have 'earned' the right to be person A; the other six will be person B.

Context 3: 'exchange'

In the third context, person A is a 'seller' and B is a 'buyer'. A table lists the profit of the seller and of the buyer for each possible price ($0, $1, $2, …, $10) charged by the seller if the buyer chooses to buy. The profit of the seller is equal to the price chosen; the profit of the buyer is $10 minus the price. The profit of each is zero if the buyer refuses to buy at the price chosen by the seller.

Context 4: 'contest–exchange'

The fourth context combines the second and third: 'sellers' are selected by a general-knowledge quiz. In one version the total amount is 10 one-dollar bills; in another, it is 10 ten-dollar bills.

Results of Ultimatum Game Experiments

The game-theoretic concept of sub-game perfect equilibrium (SPE) yields the same prediction in all of these versions of the ultimatum game (Selten, 1975): player 1 offers the minimum positive unit of account ($1 if $m = \$10$, $10 if $m = \$100$), and player 2 accepts the offer. The analysis assumes that each player is self-interested in the sense of always choosing the largest of two pay-offs for himself or herself; that this condition is common knowledge for the two players; and that player 1 applies backward induction to the decision problem faced by player 2, conditional on player 1's offer. Thus player 1 should reason that any positive pay-off is better than zero for player 2, and therefore, player 1 need only offer the minimum positive amount.

But there are other models of decision for games like the ultimatum. A problem with the above analysis is that, perhaps depending on context, the ultimatum interaction may be interpreted as a social exchange between any two anonymously matched players who normally read intentions into the actions of others (Baron-Cohen, 1995). Suppose that the ultimatum game is perceived as a social contract in which player 2 has a ('fair claim') entitlement to more than the minimum unit of account; then an offer of less than the perceived entitlement (say, only $1, or perhaps even $2 or $3) may be rejected by some players 2. Player 1, reading this potential mental state of player 2 (e.g. by imagining what he or she would do in the same circumstance), might then offer substantially more than $1 to ensure acceptance.

Observe that in context 1, the original $10 is allocated imprecisely to both players, and does not clearly belong to either person A or B. Further, a common interpretation of the word 'divide' involves the separation of some divisible quantity into equal parts. Moreover, in western culture the use of a lottery or other random device is recognized as a standard mechanism for 'fair' or equal treatment. Hence, the instructions can be interpreted as suggesting that the experimenter is engaged in a 'fair' treatment of the subjects. This can serve as a strong, albeit unconscious, cue that the subjects ought to be 'fair' in their treatment of each other.

By contrast, context 2 deliberately introduces a contest procedure, before the game itself, in which those who score the highest earn the right to be person A, and those who score the lowest will be person B. In this treatment, nothing is said about who has been initially allocated the money, and the word 'divide' is not used. Rather, person A must choose how much person B is to receive, and person B must choose to accept or reject the proposal. Consequently, the instructions may cue some norm of 'just desert' based on test performance.

In context 3, the abstract ultimatum game is embedded in a transaction between a buyer and a seller. In such exchanges, buyers (in western culture) do not normally question the right of the seller to move first by quoting a price, nor that of the buyer to respond with a decision to buy or not to buy.

Context 4 combines the implicit 'property right' norm of a seller with an explicit mechanism whereby subjects 'earn' the privilege of being the seller in a contest whose outcome provides the same opportunity for all participants, depending on their general knowledge. This treatment introduces the 'equal opportunity' norm, as opposed to 'equality of outcome'.

Table 1 summarizes the results from two different studies of ultimatum game bargaining with stakes of either 10 one-dollar or 10 ten-dollar bills, where the number of pairs of players varies from 23 to 27. Note that 'divide' with random entitlement corresponds to context 1; 'divide' with earned entitlement to context 2; 'exchange' with random entitlement to context 3; and 'exchange' with earned entitlement to context 4.

Comparing 'divide $10' with 'divide $100' under random entitlement, we observe a trivial difference

Table 1. Mean percentage offered in ultimatum games, by context treatment. Data from Hoffman *et al*. (1996) and (1994)

		$10 stakes	*$100 stakes*		
		'Divide'	*'Exchange'*	*'Divide'*	*'Exchange'*
Random entitlement	Mean offer	43.7%	37.1%	44.4%	(n/a)
	N	24	24	27	(n/a)
	Rejection rate[a]	8.3%	8.3%	3.7%	(n/a)
Earned entitlement	Mean offer	36.2%	30.8%	(n/a)	27.8%
	N	24	24	(n/a)	23
	Rejection rate[a]	0	12.5%	(n/a)	21.7%

[a]Percentage of the *N* pairs in which the second player rejects the offer of the first.

in the amount offered between the low stakes (43.7%) and the high stakes (44.4%). There is no significant difference in the rate at which offers are rejected (8.3% and 3.7% respectively).

When 'exchange' is combined with an earned entitlement, the increase in stakes seems to lower the offer percentage from 30.8% for $10 stakes to 27.8% for $100 stakes, but this difference is within the normal range of sampling error using different groups of subjects and is not significant. Surprisingly, this minuscule decline in the mean offer correlates with an increase in the rejection rate from 12.5% to 21.7%. In the high-stake game, three out of four subject players 1 offering $10 are rejected, and one offer of $30 is rejected. As we shall see below in other games, this behavior is associated with a strong human propensity to incur personal cost to punish those who are perceived as cheaters, even under strict anonymity (as in Cosmides, 1985).

Comparing the 'divide' and 'exchange' conditions with random entitlement and $10 stakes, the offer percentage declines from 43.7% to 37.1%, and comparing the 'divide' conditions with random and earned entitlement and $10 stakes the offer percentage declines from 43.7% to 36.2%. Both reductions are statistically significant. Even more significant is the reduction from 43.7% to 30.8% with the 'exchange' condition and earned entitlement. Moreover, in all four of these contexts the rejection rate is null or modest (0 to 12.5%).

The small proportion of offers rejected (except when the stakes are $100 in the 'contest–exchange' context, where the mean offers decline to 27.8%) indicates that players 1 generally read their counterparts well and offer a sufficient amount to avoid being rejected. The one exception shows that trying to push back the boundary, even if it seems justified by the higher stakes, may provoke rejections.

One obvious conclusion from these data is that the effect of context on behavior cannot be ignored in the ultimatum game: the percentage offered varies by over a third between the highest (44%) to the lowest (28%) measured values. Studies of cross-cultural variation in ultimatum offers show a comparable variation. Thus, a comparison of two hunter–gatherer and five modern cultures reveals a variation from a maximum of 48% (Los Angeles subjects) to a minimum of 26% (Machiguenga subjects from Peru) (Heinrich, 2000). These comparisons attempted to control for instructional differences across different languages, but of course this is inherently problematic in that one cannot be sure that the translations, or the procedures for handling the subjects, completely control for context across cultures. Nor can it be assumed that the pay-offs are subjectively comparable across currencies.

The instructional comparisons also call into question the extent to which one can define what is meant by 'unbiased' instructions. Some results may be robust with respect to instructional changes, but this can only be established empirically, since we know little about the sources of behavioral variation due to context. Indeed, unless such robustness is established no claims can be made concerning the relative 'neutrality' of instructions, and the extent to which differences in behavior can be attributed to differences between cultures.

Because of the nature of perception and memory, we should expect context to be an important factor. In the ultimatum game, the variation of observed results with systematic instructional changes designed to alter context shows clearly that context can and does matter.

Trust Games

Ultimatum games have been studied extensively, but because of their simplicity they leave unanswered many questions about what underlies the behavior manifest in them. For example, one cannot vary independently the cost of player 2's rejection of player 1's offer. The game is constant-sum, and is inherently confrontational: neither player can take action that increases the total gains from the transaction, and therefore the interpretation of the game as an exchange is limited.

We turn therefore to a somewhat richer class of two-person extensive-form trust games in which the return to equilibrium play, cooperation, defection, and the prospect of costly punishment of defection can be studied in a richer parameter space than that of the ultimatum game.

Figure 2(a) shows a trust game tree. Play starts at the top, node x_1, with player 1. Player 1 can move right; this stops the game, yielding \$7 to player 1 and \$14 to player 2. Alternatively, player 1 can move down, in which case player 2 has to choose a move at node x_2. If player 2 moves right, each player gets \$8. If player 2 moves down, player 1 can then move right at node x_3, yielding \$10 for each, or down, yielding \$12 for player 1 and \$6 for player 2.

The SPE is \$8 for each player. This is because at node x_1 player 1 can look ahead (use 'backward induction') to see that if play reaches node x_3 player 1 will want to move down. But player 2, also using backward induction, will see that at node x_2 player 2 should move right. Since this yields a higher pay-off to player 1, at node x_1 player 1 should move down.

The SPE outcome would prevail by the logic of self-interested players who always choose dominant strategies. There are other behavioral possibilities, however, depending on whether other preferences or perceptions of the interaction are applied.

If player 1 has other-regarding preferences (altruism, or utility from the other's pay-off), and is willing to incur some cost to greatly increase the pay-off to player 2, player 1 may move right at x_1. That way, at a cost of \$1, player 1 can increase player 2's pay-off by \$6, compared with the SPE. Thus, player 1 need have only a modest preference for an increase in player 2's welfare in order to move right.

At x_2, player 2 may move down, signaling to player 1 that such a move enables both to profit, provided that at x_3 player 1 cooperates by reciprocating player 2's favor. Alternatively, at x_3 player 1 can defect, by choosing the dominant strategy and moving down.

Figure 2(b) shows the tree for a punishment version of the trust game shown in Figure 2(a). The trees are identical except that at node x_3, player 1 chooses between the cooperation pay-off and passing back to player 2 at node x_4. Now player 2 decides whether to accept the defection or, at a cost, punish player 1 for the defection. By backward induction, the SPE is the same in the punishment version. The cooperation outcome can be justified (as a Nash equilibrium) only if the threat of punishment by player 2 at node x_4 is credible. But under the anonymity conditions, with no capacity to communicate, such a threat is not credible.

The outcome frequencies for the trust game ($N = 26$ pairs), and for the trust–punishment game ($N = 29$ pairs) are summarized in Figure 3.

In neither game is there a single case of a player 1 choosing the altruism outcome: all choose to pass to player 2, seeking a higher pay-off for themselves, and being content to give player 2 a much smaller pay-off than would be achieved by altruism.

The sub-game perfect equilibrium is chosen by 54% of the pairs in the trust game, and 55% in the trust–punishment game. Thus, there is no significant difference in behavior between the two games in terms of the frequency with which players 2 offer to cooperate by passing to players 1 at x_2.

There is, however, a considerable difference in the response of players 1 to the offer to cooperate: only 50% cooperate in the trust game, while 85% cooperate when facing the prospect of punishment for defection.

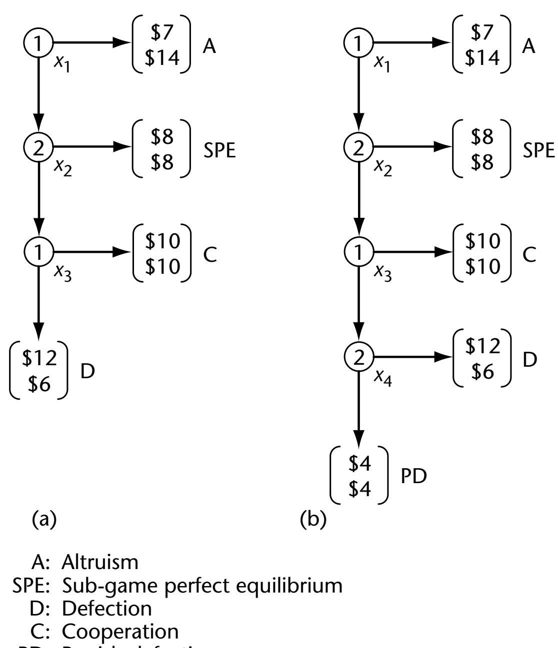

Figure 2. Trust game trees, (a) without punishment and (b) with punishment. At each terminal node the pay-off to player 1 is shown above the pay-off to player 2.

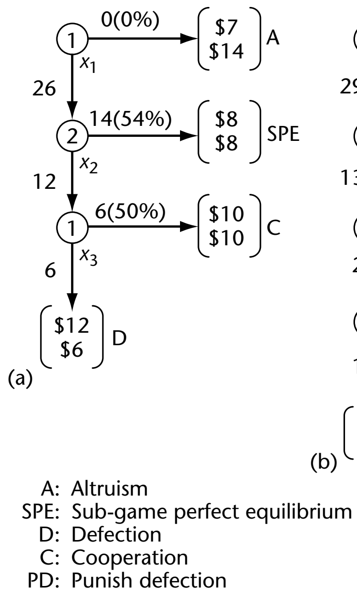

Figure 3. Experimental outcomes for the (a) trust and (b) trust–punishment games shown in Figure 2. A total of 26 subject pairs took part in the trust game and 29 in the trust–punishment game. The figures beside the arrows indicate the number of pairs following each route through the game tree and the percentage moving right at each decision node. The data are from McCabe *et al*. (2000). Source data for larger trees have been trimmed to eliminate rare outcomes, with commensurate reduction in sample size (from 30 to 29 in the punishment version).

Across the two games, why do nearly half of the players 1 forgo the sure SPE payoff in favor of the risky prospect of cooperation? McCabe and Smith (2001) argue that humans are eminently adapted for social exchange, or reciprocity among the individuals that constitute the small groups that form our primary networks of relationships. We constantly trade favors, services and assistance, with little conscious awareness of these trading relationships that are so much a part of our humanity. McCabe and Smith postulate an implicit mental accounting system for keeping track of trustworthy trading partners. This accounting system is part of the framework of our friendships and social connections.

Reciprocity is a human universal, characteristic of all cultures, as is the use of a spoken language. Like language, the form of reciprocity varies across cultures, but its common functionality is to produce gains from exchange. Smith (1998) argues that reciprocity in the family, extended family, and tribe is what ultimately led to the extended order of cooperation through market trade. He postulates that this proclivity for reciprocal social exchange is so natural and instinctive that it survives even in interactions between anonymously paired subjects in the two-person extensive-form games described above.

This interpretation has been reinforced by many other extensive-form game-tree experiments. Thus, in the game shown in Figure 4, player 1 chooses between the SPE, $10 for each, and passing to player 2 who chooses between pay-offs of $15 and $25 (for players 1 and 2 respectively) and pay-offs of $0 and $40. The move frequencies for 24 pairs of undergraduates are shown on the tree. Very similar outcomes prevail with a group of graduate students trained in economics and game theory (McCabe and Smith, 1999).

Effects of Context, Repetition, and Opportunity Costs in Trust Games

The sensitivity of cooperative behavior in trust games to the procedural, instructional, and opportunity cost context of the experiment has been demonstrated by several treatment manipulations.

'Partners' versus 'opponents'

Consider two treatment variations on the trust game of Figure 2(a): wherever the word 'counterpart' is used in the instructions to refer to the other decision maker, substitute the word 'partner' in one treatment and 'opponent' in the other (Burnham *et al.*, 2000). Subjects (156 pairs in total) were recruited in either 'small' groups of 12 in a session or 'large' groups of 24 in a session. In all sessions, half of the subjects (6 or 12) were randomly assigned to each of the two instructional conditions;

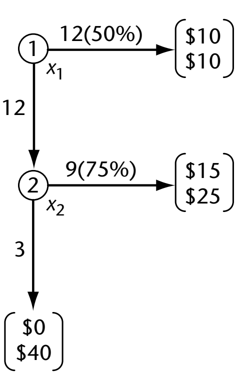

Figure 4. Experimental outcomes for a simple trust game (McCabe and Smith, 1999). A total of 24 subject pairs took part in the game. The figures beside the arrows indicate the number of pairs following each route though the game tree and the percentage moving right at each decision node.

each person was randomly paired with another and assigned to the player 1 or player 2 role; and the two experiments were run simultaneously in the same room. Neither group was informed that the other was reading slightly different instructions. Thus, the experimental design consisted of two group sizes, 12 or 24, and two instructional conditions, 'partner' and 'opponent'.

Each session began with a single play of the trust game. The subjects were then paid and informed that they would also participate in a second experiment. This second experiment used the same instructions except that the game was repeated for 10 periods of play. On each period of play, each person was matched with a new person, then randomly assigned the role of player 1 or 2. Each repetition was therefore between paired strangers. This is called 'repeat single' play. Repeat single play is like single play except that the subjects acquire experience under procedures that control for reputation formation across successive interactions.

It was found that 'partners' are more trusting (players 2 move down at node x_2) and more trustworthy (players 1 move right at x_3) than 'opponents'. (In the first single-play game, however, no difference was observed in the frequency of trust between the two treatments, but 68% of the 'partner' players 1 cooperated following an offer of cooperation, while only 33% of the 'opponent' players did.)

Over time (single play followed by 10 repeat single plays), with 'partners' trust increases through the first five plays then declines, while with 'opponents' trust steadily declines. Trustworthiness declines over time for 'partners', and remains low for 'opponents'. Hence, 'partners' learn to defect, but 'opponents' defect from the beginning.

Pairs who interact in groups of size 24 are less trusting than those in groups of size 12.

These results provide further support for the hypothesis, based on cortical memory theory, that context should matter. In this case, a simple two-level variation on the language used to describe the other person in each trial is sufficient to yield statistically significant differences in trust and trustworthiness.

Repeat single with and without punishment

The tendency for cooperation eventually to decline as play is repeated with distinct 'partners' is already suggested by Figure 3(a): of the 12 players 1 arriving at node x_3, half reciprocate and half defect. Hence, it is not profitable to offer cooperation in the trust game, and repetition with strangers is likely to cause a decline in cooperation, both offered and reciprocated, across time.

In the trust–punishment game in Figure 3(b), however, 85% reciprocate at node x_3, and only 15% defect, of which half are punished. Hence, it is profitable to offer cooperation, and it is not profitable to defect. This suggests that in repetition, using the repeat-single protocol, cooperation might not diminish. In fact, this is the case: when defection can be punished, the conditional probability of reciprocal cooperation by players 1 actually increases modestly across 15 periods of play (McCabe *et al*., 2000).

Opportunity cost

An important implication of reciprocity theory (the value of option 'y' given up by choosing 'x') is that when person A chooses to forgo the SPE outcome and offer the cooperative option to person B, the pay-off alternatives should be such that person B sees clearly that person A is incurring an 'opportunity cost' – forgoing a smaller pay-off in an attempt to allow both persons to achieve larger pay-offs. There should be a cost incurred in order to gain from exchange. Failing this condition, the basis for an exchange, or reciprocation, is compromised: person B would be less likely to read clearly the intentions of person A, and person A will anticipate that an unclear message would be being conveyed.

Thus, in Figure 2, if instead of \$8 the SPE is \$10 for each player – identical to the 'cooperative' outcome – the outcome frequency results should change dramatically. This has been tested for the trust–punishment game tree in Figure 2(b) (McCabe *et al*., 2002). The effect is to increase the frequency of the SPE outcome to about 95%. Thus, players have no difficulty concluding that the attempt to cooperate by player 2 at node x_2 is risky, and will not be chosen unless there is a compensating potential gain.

Another test of reciprocity is to contrast two versions of a game with the structure of Figure 4. Version 1 is like that in Figure 4, with different pay-offs but qualitatively the same outcomes. In version 2, player 1 has no option to move right. The prediction is that version 1 will yield more cooperative outcomes then version 2. In fact, defection is twice as frequent in version 2 as it is in version 1. The interpretation is that if nothing was given up by player 1 – the move did not deliberately forgo the pay-offs achievable at the SPE – then player 1's move does not constitute an 'offer'; so nothing need be reciprocated.

THEORY OF MIND AND ITS NEURAL CORRELATES

Experimental tests of non-cooperative equilibrium theory using anonymously paired subjects in two-person games consistently show that people do cooperate. Almost all subjects in the ultimatum game offer amounts in excess of the equilibrium predictions, and when they do offer equilibrium amounts their counterparts almost always reject the offer. Similarly, in trust games, up to half of subjects offer to cooperate at the risk of defection; and in varying degrees, depending on context, their counterparts cooperate at a cost to themselves. These data cannot be explained simply in terms of preferences – a utility for the other's payoff – nor can they be dismissed by the argument that the subjects are too unsophisticated or inadequately motivated.

A more satisfactory model is based on reciprocity and the human ability to communicate intentions through actions. This ability to invoke shared-attention, intention-detector, and 'mindreading' mechanisms in the brain is relevant to other observations of behavior in people impaired by frontal lobe damage and by autism.

Autism (whose genetic basis is indicated by its greater incidence in siblings and in identical twins) is characterized by 'mind blindness', a severe deficit in one's innate awareness of mental phenomena in other people. Children with autism fail developmentally to use pointing gestures to request objects or otherwise call the attention of others to items of joint interest. In contrast, blind children at age 3 are aware of what 'seeing' is in others, and will say 'see what I have'. At about age 3 or 4, normal children become aware of beliefs in others, and understand that others can hold false beliefs. Thus, shown a candy box, and asked what it contains, normal children will say that it contains candy. Upon opening the box, the child sees that pencils have replaced the candy. The child is then asked what the next child who comes in the room will think is in the box. Normal children will reply 'candy', whereas the majority of autistic children will say 'pencils' (Baron-Cohen, 1995).

Studies of autism, and of certain forms of brain damage from accidents or surgery, support hypotheses that particular regions of the brain have circuitry devoted to 'mindreading', an innate capacity for unconscious awareness of what others think or believe. Brain imaging studies of third-party false beliefs in story comprehension tasks have found activation in Broadman's area 8 (medial prefrontal cortex), and in other supporting regions such as the orbital frontal cortex (Fletcher *et al.*, 1995). This role of Broadman's area 8 has been specifically corroborated by functional magnetic resonance imaging of subjects playing trust and trust–punishment games like those presented above (McCabe *et al.*, 2001). These studies compare subjects' decision making when playing a human counterpart and when playing computer strategies with fixed known probabilities of moving 'left' or 'right'. Activation is significantly greater in the mindreading areas when playing a human than when playing a computer.

Thus, independent strands of research into the internal order of the mind and the external order of social exchange appear to be converging in support of the hypothesis that humans are so well adapted to personal exchange that reciprocity survives even in anonymous interactions.

References

Baron-Cohen S (1995) *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge, MA: MIT Press.

Burnham T, McCabe K and Smith VL (2000) Friend-or-foe intentionality priming in an extensive form trust game. *Journal of Economic Behavior and Organization* **43**: 57–73.

Cosmides L (1985) The logic of social exchange. *Cognition* **31**: 187–276.

Cosmides L and Tooby J (1992) Cognitive adaptations for social exchange. In: Cosmides L and Tooby J (eds) *The Adapted Mind*, pp. 163–228. New York, NY: Oxford University Press.

Fletcher P, Happe F, Frith U *et al.* (1995) Other minds in the brain: a functional imaging study of 'theory of mind' in story comprehension. *Cognition* **57**: 109–128.

Fuster J (1999) *Memory in the Cerebral Cortex*. Cambridge, MA: MIT Press.

Gazzaniga M, Ivry R and Mangun G (1998) *Cognitive Neuroscience*. New York, NY: Norton.

Hayek F (1952) *The Sensory Order*. Chicago, IL: University of Chicago Press.

Hayek F (1988) *The Fatal Conceit*. Chicago, IL: University of Chicago Press.

Heinrich J (2000) Does culture matter in economic behavior? *American Economic Review* **90**(4): 973–979.

Hoffman E (2000). In: Smith (2000), pp. 79–90.

Hoffman E and Spitzer M (1985) Entitlements, rights and fairness. *Journal of Legal Studies* **14**: 259–297.

Hoffman E, McCabe K and Smith VL (1996) On expectations and the monetary stakes in ultimatum games. *International Journal of Game Theory* **25**(3): 289–301. [Reprinted in Smith (2000).]

Hoffman E, McCabe K, Shachat K and Smith VL (1994) Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior* **7**: 346–380. [Reprinted in Smith (2000).]

McCabe K and Smith VL (1999) A comparison of naïve and sophisticated subject behavior with game theoretic predictions. *Proceedings of the National Academy of Sciences of the USA* **97**: 3777–3781.

McCabe K and Smith VL (2001) Goodwill accounting and the process of exchange. In: Gigerenzer G and Selten R (eds) *Bounded Rationality: the Adaptive Toolbox*, pp. 319–340. Cambridge, MA: MIT Press.

McCabe K, Houser D, Ran L, Smith VL and Trouard T (2001) A functional imaging study of cooperation in two-person reciprocal exchange in process. *Proceedings of the National Academy of Sciences of the USA* **98**: 11832–11835.

McCabe K, Rassenti SJ and Smith VL (1996) Game theory and reciprocity in some extensive form experimental games. *Proceedings of the National Academy of Sciences of the USA* **93**: 13421–13428. [Reprinted in Smith (2000).]

McCabe K, Rigdon M and Smith VL (2002) Positive reciprocity and intentions in trust games. *Journal of Economic Behaviour and Organization* (in press).

Selten R (1975) Re-examination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory* **4**: 25–55.

Smith VL (1998) The two faces of Adam Smith. *Southern Economic Journal* **65**: 1–19.

Smith VL (2000) *Bargaining and Market Behavior*. Cambridge, UK: Cambridge, University Press.

Education, Learning in

Introductory article

Stanton Wortham, University of Pennsylvania, Philadelphia, Pennsylvania, USA

CONTENTS

Learning takes place in many settings, but educational institutions foster both breadth and depth of learning. Different types of teaching make very different assumptions about what learning is.

INTRODUCTION

Theories of learning have been applied most often in educational institutions. The relationship between cognitive science and education has benefited both scientists and practitioners. Scientists have used educational settings to develop and test their theories, and practitioners have used new knowledge about learning to design more effective education.

Broadly conceived, education is the process of continuing the human species. All humans are born immature, without the knowledge and skills they will need to function – without language, without knowing how to use complex tools, and so on. The species continues because adults communicate knowledge and skills to the next generation. This intergenerational transfer allows future generations to build on prior accomplishments.

Thus all humans teach. Whether they realize it or not, all teachers act as if some theory of learning is true. Particular ways of teaching make assumptions about what learning is. Furthermore, theories of learning themselves rest on conceptions of human nature. Different accounts of how people learn assume different things about what people are essentially like.

This article describes three broad theories of learning – together with the conceptions of human nature underlying these theories – and the types of educational practice that have been built on these theories. The article has two purposes. First, it is important to recognize the theories of learning and conceptions of human nature that underlie various types of schooling. The article describes how typical teacher and student behavior makes assumptions about how learning happens. Second, as theories of learning have developed, we have learned that earlier theories were too simple. The article describes how more complex accounts of learning and human nature are needed to guide educational practice.

BEHAVIOR

Theories of learning that focus on behavior are called ‘behaviorist’. Behaviorists argue that humans should not consider themselves special. Copernicus showed that the earth was not the