ENCYCLOPEDIA OF COGNITIVE SCIENCE

Experimental Methods in (Neuro)Economics

Key words:

Smith, Vernon L.

University of Arizona, Tucson, Arizona, USA

> “. . . an economist who is nothing but an economist cannot be a good economist.”
>
> F. A. Hayek

This entry will be limited primarily to a discussion of the use of experimental methods in the study of decision behavior by humans in strategic interactions. This subfield within experimental economics also includes neuroeconomics (a term originated by Kevin McCabe). (For a discussion of experimental methods in market economics see the entry “Markets, Institutions and Experiments”).

For cognitive scientists it will be particularly useful to distinguish three complex self-ordering systems central to understanding the human enterprise.

1. The internal order of the mind (Hayek, 1952, calls it the Sensory Order which began as an essay “What is Mind?”) interpreted here as the entire neurobiological structure of the human organism.

2. The external order of social exchange between minds. (McCabe and Smith, 2001).

3. The extended order of cooperation through markets and other cultural institutions (Hayek, 1988).

Our concern here is with the first two, while the companion entry referenced above deals with the third.

Somewhat curiously, just as the economic writings of Malthus and that of the Scotsman Adam Smith had a profound influence on Darwin, so F. A. Hayek, the 1974 Nobel Laureate in economics, has had a similar influence on cortical memory theory in neuroscience and on some of the fundamental precepts in evolutionary psychology. This is less surprising to those who have penetrated the broadly informed, and deep insights contained in the immense diversity of Hayek's work. The Hayek problem – the formation of complex orders, or how does order create itself out of complexity – in economics, biology and mind was a lifetime pursuit beginning in his early 20s with his knowledge of Mach's Analysis of Sensations (and including what he learned from working for a few weeks in the laboratory of the Swiss brain anatomist, Constaintin von Monakow). From his theoretical contributions to understanding the mind, he went on to work out the implications of his understanding of knowledge and complexity for sociology, cultural evolution, psychology and law.

We focus on social exchange because it is the cooperative behaviors registered in two person anonymous interaction that first alerted experimental economists to the significant class of phenomena that violate the cannons of certain static equilibrium concepts in game theory. 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, the model of nonrepeated game theory is about strangers without a history or a future, and anonymity provides the required control for giving this theory its best shot. Also, it is well documented that face-to-face interaction swamps subtler procedural effects in yielding cooperative outcomes. (Hoffman and Spitzer, 1985). Finally, and most important, I believe it is this condition that provides the greatest scope for exploring the human instinct for social exchange, and how it is effected by contextual, reward and procedural conditions.

Why Should Context Matter?

Context matters because all memory involves relationships and is associative. This is indicated in priming, which uses cues to improve retrieval from memory because of associations between the cue and the stimulus. For example, in priming experiments people perform better at completing word fragments (filling in missing letters) on words they have observed in pretest lists, even if they were not told that the words appeared in the earlier list. Priming effects do not decay whether the test and pretest intervals are hours or days. Furthermore, being able to state that a word was seen before does not improve completion performance. Generally, context provides unconscious cues to past experience that embodies a process of how one perceives a current 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, Ivry and Mangun, 1998, pp. 258-261).

This is because of the fundamental, but nonintuitive, nature of perception. Hayek was a previsionary pioneer in the early articulation of certain principles of perception, which are consistent with contemporary 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 (Hayek, 1952, p. 165).

2. Rather the process by which we learn about the external environment involves a relationship between current conditions and our past experience of similar conditions (p. 165).
3. Categories are formed by the mind out of the relative frequency with which current perception and memory (past perceptions) concur (p. 64). That which is 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.
4. 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, p. 88-9).

Although Hayek's model is incomplete, it captures the idea that perception is self-organized, created from abstract function combined with experience to determine network connectivity and expansion. Loss follows from lack of either function or the stimulus of developmental experience. Block the sensory stimuli and functional development is impaired; impair function by brain lesions or a hereditary deficiency and development fails or is compromised.

Mental Modules and Evolutionary Psychology

Hayek's conception of mind is also consistent with that of recent developments in evolutionary psychology which argue that the mind consists of circuitry, or interactive modules, that are specialized for vision, for language learning, for socialization and a host of other functions. (See Cosmides and Tooby, 1992). Language and socialization, which are of recent evolutionary origin, are hypothesized to have evolved in the 2-3 million years in 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 gather societies (e.g. the Ache of Paraguay) in which the product of the hunt is widely shared within the tribe as well as the nuclear and extended family. In a world without refrigeration and only rare hunting luck this made sense. It I am lucky in the hunt today, I share the meat with others, and tomorrow, failing to make a kill, you share your kill with me and others. In contrast the products of gathering – fruits, nuts, roots – depend more on effort than luck, are far less volatile day-to-day and are shared only in the nuclear family where effort can be closely monitored. Sharing traditions across time provide gains from exchange that support limited forms of specialization: women and children under 18 do the gathering, adult men after 18 do the hunting, older men make tools, advise in the hunt and return to 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 "mind reading," the ability to infer mental states from the words and actions of others. The minimal equipment required is a cheater-defector "module" for social exchange. Experiments pioneered by Cosmides (1985) are consistent with the postulate that the human mind is attuned to detecting cheaters on perceived social contracts. With the development of language our instincts for cheater defection were leveraged by gossip – comparing notes to ferrite out those with good reputations for returning favors. Gossip, like language and reciprocity is a human universal, a favored pastime across all human communities. None of this equipment was the product of our reason; rather it was the unconscious product of co-evolution between the biological and cultural development 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 upon us created by the world since the agricultural revolution 10,000 years ago. This conclusion was reached independently 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).

Experimental Procedures

The experiments reported below will show that context is important in the social decision behavior we observe. As we have seen, 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. For the most part I will be reporting on behavior in two-person sequential-move game or decision trees in which each pair plays once and only once through the move sequence defined by the tree and this fact is known to the subjects. The instructions for each experiment, however, do not use words like "game," "play," "player," "opponent," "partner," etc. (except where they are used as systematic treatments), 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. I do not mean to imply that the baseline is "neutral," in fact I think that any such claims are indefensible.

In a typical experiment imagine that you have arrived at the laboratory at the designated appointment time for which you were 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 commonly 11 other people, well spaced throughout the room, in the experiments reported below. 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 monitor screen. You read through the instructions for the experiment at your own pace, responding 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. For example, it may appear evident that he or she is another "like" person, such as an undergraduate, with whom you may feel more-or-less an in-group identity. Obviously, you import into the experiment a host of past experiences and impressions that you are likely to apply to the current experiment.

The Context of Decision: The Ultimatum Game Example

Consider the simple two-stage two-person game described in abstract form as follows: 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 $x \leq m$ units of the money be offered to player 2, player 1 retaining $m - x$. 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 units of money.

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, pp. 79-90 in Smith, 2000 for instructional details, and for references to the literature and origins of the ultimatum game.

Context 1. "Divide \$10." In this version 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 first mover role. Your task is to fill out the following proposal form (in which you are person A):

(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 _____

The form goes to your paired number (B) who checks 'accept' or 'reject'. In this version the $10 consists of 10 one-dollar bills. In the second version there is $100 or 10 ten-dollar bills to be divided.

Context 2. "Contest entitlement." The 12 people in the room each take the same general knowledge quiz (10 questions) that is used to determine the positions of persons A and B in each pairing. Your score is the number of questions answered correctly, with ties broken in favor of the person who finished the quiz earliest in time. The scores will be ranked from 1 (highest) through 12 (lowest). Those ranked 1-6 will have <u>earned</u> the right to be persons A, the other six will be persons B.

Context 3. "Exchange" (Buy-Sell). Person A is now a seller, and B is a buyer. A table lists the profits of the seller and that of the buyer for each 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 – price). The profit of each is zero if the buyer refuses to buy at the price chosen by the seller.

Context 4. "Contest/Exchange." This treatment combines the conditions (2) and (3) above; i.e. the seller in (3) is selected by the context scoring procedure described in (2). In one version the total amount is 10 one-dollar bills, and in the second it is 10 ten-dollar bills.

Irrespective of context there is a game theoretic concept of (subgame perfect) equilibrium that yields the same prediction in all of these versions of the ultimatum game (Selten, 1975): player 1 offers the minimum unit of account, $1 ($10) if m = $10 ($100), and player 2 accepts the offer. Why? The analysis assumes that each player is self interested in the sense of always choosing the largest of two payoffs for himself; 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 payoff is better than zero for player 2 and therefore, player 1 need only offer x = $1 ($10).

But there are other models of decision for games like the ultimatum. A problem with this 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 an entitlement to more than the minimum unit of account, then an offer of less than the perceived entitlement (say, only $1, or even $2-$3) may be rejected by some players 2. Player 1, reading this potential mental state of player 2, e.g. by imagining what he would do in the same circumstance, might then offer substantially more than $1 to insure acceptance of his offer.

Our personal identity is defined by some combination of inherited mental characteristics and our developmental experience, the latter in turn being influenced by culture. As noted above, evolutionary psychologists have postulated that we possess specialized mental circuitry for our social behavior. What has co-evolved as part of our biological and cultural inheritance is an apparent strong propensity to punish cheaters on perceived social contracts. (Cosmides and Tooby, 1992). Hence, offering too little in the ultimatum game may violate a perceived social contract, that emerges in the form of small group solidarity, and be rejected by player 2. Since punishment is costly to the punisher, however, such behavior begs for an explanation of how it might have survived individually to improve biological fitness.

The point is simply that there are alternatives to the argument that subgame perfect equilibrium theory will rule behavior in the ultimatum game, and these alternatives leave wide latitude for the possibility of context effects on the decision behavior of both players.

Observe that in Context 1 the original $10 is allocated imprecisely to both players; it does not clearly belong to either person A or B. Further, a common definition of the word “divide” (see Webster) includes the separation of some divisible quantity into equal parts. Finally, in our 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 strongly suggesting that the experimenter is engaged in the “fair” treatment of the subjects. It follows that this can serve as a strong cue, albeit unconscious, that the subjects ought to be “fair” in their treatment of each other.

As a counterpoint to this interpretation of Context 1, Context 2 deliberately introduces a pregame contest procedure in which those who score the highest earn the right to be person A, and those who score the lowest will be persons B. Also, in this treatment, nothing is said about who has been initially allocated the money, and the word “divide” is not used. Rather A must choose how much B is to receive, and B is to choose to accept or reject the proposal. Consequently, the instructions may cue some insipient norm of just desert based on pregame test performance.

In Context 3 the abstract ultimatum game is imbedded in the gains from exchange from a transaction between a buyer and a seller. In such exchanges, both the buyer and the seller are made better off, and buyers in our culture do not normally question the right of the seller to move first by quoting a price, nor 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 upon their general knowledge.

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 for each of *N* pairs of players, where *N* varies from 23 to 27 subject pairs.

1. Comparing Divide $10 with Divide $100 under the random entitlement we observe a trivial difference in the amount offered between the low stakes (43.7%) and the 10-fold increase in the stakes (44.4%). Also, there is no significant difference in the percentage rate at which offers are rejected, 8.3% and 3.7% respectively.

2. When Exchange is combined with an Earned Entitlement the 10-fold increase in stakes lowers 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, however, this minuscule decline in the mean offer causes the rejection rate to go up from 12.5% to 21.7%. Three of four subject players 1 offering $10 are rejected, and one offer of $30 is rejected in the game with $100 stakes. 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.

3. We note that comparing the Divide $10/Random entitlement condition with the Exchange entitlement, the offer percentage declines from 43.7% to 37.1%, and comparing the former to the Earned entitlement the decline is from 43.7% to 36.2%, both reductions being statistically significant. Even more significant is the reduction form 43.7% to 30.8% when the

Earned and Exchange Entitlements are combined. Moreover, in all four of these comparisons the rejection rate is null or modest (0 to 12.5%).

4. The small proportion of the offers that were rejected, except when the stakes were $100 in the Earned/Exchange context and the mean offers declined to a low of 27.8%, indicates that players 1 read their counterparts well, and as the context is altered, normally offer a sufficient amount to avoid being rejected. The one exception shows clearly that pushing the edge, even if it seems justified by the higher stakes, may invite an escalation of 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 as we move from the highest (44%) to the lowest (28%) measured effect. Studies of cross-cultural variation in ultimatum offers show a comparable variation. Thus a comparison of two hunter-gather and five modern cultures reveals a variation from a high of 48% (Los Angeles subjects) to a low 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 payoffs 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 so 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 across cultures are due to variation in the cultures. The main lesson in all this is that, because of the nature of perception and memory, context should matter, and 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 With and Without the Option to Punish Defection

Ultimatum games have been studied extensively but are much too simple to allow an adequate understanding of the underlying 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, so that neither player can take action that increases the total gains from the transaction, and therefore the interpretation of the game as an exchange is strained.

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 the ultimatum game.

Figure 1(a) is a trust game tree. Play starts at the top, node x_1, with player 1 who can move right. This stops the game yielding the upper payoff to player 1, $7, and the lower payoff to player 2, $14. Alternatively, he can move down in which case player 2 chooses a move at node x_2. If the move is 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 subgame perfect equilibrium (SPE) is $8 for each player. This follows because at node 1, player 1 can apply backward induction by observing 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 he should move right. Since right at node x_2 yields a higher payoff to player 1, at node x_1 player 1 will conclude that he should move down. Hence, the SPE outcome would prevail by the logic of self-interested players who always choose dominant strategies. There are other behavioral possibilities, however, depending upon whether other preferences or perceptions of the interaction are applied.

If player 1 has other regarding preferences (altruism) and is willing to incur some cost to greatly increase the payoff to player 2, player 1 may move right at x_1. His payoff of $7 is only one-eighth smaller than his payoff at the SPE, and yields $14 for player 2. Hence, at a cost of $1 to himself, player 1 can increase his counterpart player 2's payoff by $6. Player 1 need have only a modest preference for an increase in player 2's welfare to induce him to move right because of the 6 to 1 return for the other player over the cost to player 1.

At x_2, player 2 may move down signaling to player 1 that such a move enables both to profit (gains from exchange), provided that at x_3 player 1 cooperates (C) by reciprocating (R) player 2's favor. Alternatively at x_3 player 1 can defect (D) on the offer to cooperate by choosing his dominate strategy, and move down.

Figure 1(b) displays the tree for a punishment version of the trust game shown in Figure 1(a). The trees are identical except that at node x_3, Player 1 chooses between the C(R) payoff, and passing back to Player 2 at node x_4. Now player 2 decides whether to accept the defection (D) or, at a cost to himself, punish player 1 for his move down at x_3 (PD). By backward induction the SPE is not changed in the punishment version. The outcome C(R) can be defended (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) is summarized in Table 2.

The first result – overwhelmingly decisively – is that in neither game is there a single case of a player 1 choosing the A (altruistic) outcome; all choose to pass to player 2 seeking a higher payoff for themselves, and being content to give player 2 a much smaller payoff than is achieved at A, depending upon the final outcome of the move sequence.

Secondly, we note the subgame perfect equilibrium is chosen by 56% (44% offer to cooperate) of the pairs in the trust game, and 55% (45% offer cooperation) in the trust punishment game. Hence, there is no difference in behavior between the two games in terms of the frequency with which players 2 offer to cooperate my 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% reciprocate (and 50% defect) in the trust game, while 85% reciprocate when facing the prospect of punishment for defection.

Why do nearly half of the 55 players 1 across the two games, forego the sure SPE payoff in favor of the risky prospect of cooperation? I and my coauthors have argued that humans are preeminently adapted for social exchange, or reciprocity among the individuals that constitute the small groups that form our primary network of relationships. We constantly trade favors, services and assistance with little conscious awareness of these trading relationship that are so much a part of our humanity. In McCabe and Smith (2001) we postulate an implicit mental accounting system for keeping track of trustworthy trading partners which is the bedrock of our friendships and social connectedness.

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

This interpretation has been further reinforced in many other extensive form game tree experiments. Thus in Figure 2, player 1 chooses between the SPE, $10 for each, and passing to player 2 who chooses between the payoff (Player 1, player 2) = ($15, $25) and the self-interested imputation ($0, $40). The move frequencies for 24 pairs of undergraduates is shown on the tree. Very similar outcomes prevail with a group of graduate students trained in economic/game theory. (McCabe and Smith, 1999).

Effect of Context, Repetition and Opportunity Costs in Trust Games

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

'Partners' vs 'Opponents' in the Trust Game

Consider one and only one change in the instructions used in each of two treatment variations on the trust game of Figure 1(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. Subjects (156 pairs in total) are recruited in either 'small' groups of 12 in a session or 'large' groups of 24 in a session. In all of the sessions half of the subjects (6 or 12) are randomly assigned to each of the two instructional conditions, each person is randomly paired with another and assigned to the player 1 or player 2 position, and the two experiments are run simultaneously in the same room; neither group is informed that the other is reading slightly different instructions. Hence, the experimental design consists of two group sizes, 12 or 24, and two instructional conditions, ‘partner’ and ‘opponent.’ Each session begins with a single play of the trust game. The subjects are then paid and informed that they will also participate in a second experiment. The second consists of the same instructions except that the game is repeated for 10 periods of play. On each period of play each person is matched with a new person, then randomly assigned the role of player 1 or 2. Each repetition is therefore between paired strangers, and represents what is called Repeat Single play. Repeat single is like Single play except that the subjects acquire experience under procedures that control for reputation formation. (Burnham, McCabe and Smith, 2000).

The principal results are as follows:

1. ‘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 there is no difference observed in the frequency of trust between the two treatments, but 68% of the ‘partner’ players 1 reciprocate with the cooperative outcome, while only 33% of the ‘opponent’ players reciprocate.

2. Over time (Single play followed by 10 Repeat Single plays) trust increases through the first five plays then declines with ‘partners,’ while trust steadily declines with ‘opponents.’ Similarly trustworthiness declines over time for ‘partners,’ but remains low for ‘opponents.’ Hence, ‘partners’ learn to defect, but ‘opponents’ defect from the beginning.

These results provide further support for the prior expectation, based on cortical memory theory, that context should and does 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/or trustworthiness. Yet both the word ‘partner’ and the word ‘opponent’ are natural ways of designating the other member of each pair; all other elements of the two treatments being the same.

Repeat Single With and Without Punishment

The tendency for cooperation eventually to decline as play is repeated with distinct ‘partners’ in the second result above is already suggested in the single play trust game results reported in Table 2: 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, however, 85% reciprocate at node x_3, and only 15% defect, of which half are punished. Therefore, it pays to offer cooperation, while defection does not pay. 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. See McCabe, Rassenti and Smith, in Smith (2000, p. 168).

Opportunity Cost Matters

An important implication of reciprocity theory is that when person A chooses to forego the SPE outcome and offer the cooperative option to her counterpart, B, the payoff alternatives should be such that B sees clearly that A is incurring an opportunity cost – giving up a smaller payoff in an attempt to allow both A and B to achieve larger payoffs; i.e. there should be a cost incurred in order to gain from exchange. In the absence of these conditions the basis for an exchange , or reciprocation, is compromised: B is less likely to read clearly the intentions of A, and A will anticipate that an unclear message is being conveyed.

Thus, in Figure 1 if instead of \$8 the SPE is \$10 for each player, identical to the C(R) outcome, the outcome frequency results should change dramatically. In fact this has been done for the trust punishment game tree in Figure 1(b). The effect is to increase the frequency of the SPE outcome to about 95%. This result tells you that unsophisticated players have no difficulty concluding that the attempt to cooperate by players 2 at node x_2 is perceived as 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 2. Version 1 is like that in Figure 1, with different payoffs but qualitatively the same structural outcomes, SPE, C(R) and D. In Version 2, player 1 has no options to move right – there is no stem allowing a move to the right and no SPE outcome; player 1 has no choice but to play down. (In Figure 2 imagine that there is no line segment to the right, and no payoff box, at x_1, only a down move is possible). The prediction is that Version 1 will exhibit more C(R) outcomes then Version 2. This is borne out as follows: the choice of defection, D, is twice as high in Version 2 as it is in Version 1. Conclusion: player 2 must see that player 1 gave up a smaller playoff to risk receiving the outcome C(R) in order to induce player 2 to cooperate. If nothing was given up by player 1 – the move was not a voluntary incurring of the opportunity cost of SPE – then player 2 cannot reciprocate player 1's offer; i.e. nothing was offered so nothing need be reciprocated.

Theory of Mind and Its Neural Correlates

Experimental tests of noncooperative equilibrium concepts using anonymously paired subject in two person games routinely fail to find significant support. An overwhelming number of subjects in the ultimatum game offer amounts in excess of the equilibrium predictions, and when they do offer such amounts their counterparts veto the offer. Similarly, in trust games, up to half the subjects offer to cooperate at risk of defection, and in varying degrees, depending on context, their counterparts cooperate at a cost to themselves. These data cannot be rationalized 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 hypothesized to be based on reciprocity and the peculiar human ability to communicate intentions through actions. This ability to invoke shared attention, intentionality detector and 'mind reading' mechanisms in the brain links up with other observations of behavior in people impaired by front lobe damage and by autism.

Autism, whose genetic basis is indicated by its greater incidence in siblings as well as 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 three are aware of what 'seeing' is in others, and will say "see what I have." Around age 3-4 normal children become aware of beliefs in others, and understand that others can hold a false belief. Thus, shown a candy box, and asked what it contains normal children will say that it contains candy. Upon opening, the child sees that the candy has been replaced by pencils. 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 others with certain forms of brain damage from accidents or surgery, show deficits in social skills commensurate with the hypotheses that particular regions of the brain have circuitry devoted to ‘mind reading,’ 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) in particular, 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 in recent new fMRI scanning results of subjects playing trust and trust punishment games like those presented here. (McCabe, et. al, 2001). These studies compare subject decision making when playing a human counterpart with decisions when playing computer strategies with fixed known probabilities of moving ‘left’ or ‘right’. Activation is significantly greater in the ‘mind reading’ areas when playing a human. Hence, independent strands of research into both 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 *The Adapted Mind*, pp. 163-228. New York: Oxford University Press.

Fletcher P, Happe F, Frith U, Baker S, Dolan R, Frakowiak R, and Frith C (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: Norton.

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

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

Heinrich J (2000) Does Culture Matter in Economic Behavior? *American Economic Review* September: 973-979.

Hoffman E and Spitzer M (1985) Entitlements, Rights and Fairness *Journal of Legal Studies* **14**: 259-97.

McCabe K and Smith VL (1999) A Comparison of Naïve and Sophisticated Subject Behavior with Game Theoretic Predictions *Proceedings National Academy of Science* **97**: 3777-3781.

McCabe K and Smith VL (2001) Goodwill Accounting and the Process of Exchange in *Bounded Rationality the Adaptive Toolbox*, edited by G. Gigerenzer and R. Selten, pp. 319-340. Cambridge, MA: MIT Press.

McCabe K, Houser D, Ryan L, Smith VL and Trouard T (2001) A Functional Imaging Study of Cooperation in Two-Person Reciprocal Exchange in process.

Selten R (1975) Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games *International Journal of Game Theory*, pp. 25-55.

Smith VL (1998) The Two Faces of Adam Smith Distinguished Guest Lecture, *Southern Economic Journal*, pp. 1-19.

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

Table 1

Mean Percentage Offered by Context Treatment
In Ultimatum Games[1]

		$10 Stakes		$100 Stakes	
		Divide $10	Exchange Entitlement	Divide $100	Exchange Entitlement
Random Entitlement	Mean	43.7%	37.1%	44.4%	NA
	N	24	24	27	NA
	(% Rejected)[2]	(8.3%)	(8.3%)	(3.7%)	NA
Earned Entitlement	Mean	36.2%	30.8%	NA	27.8%
	N	24	24	NA	23
	(% Rejected)[2]	(0)	(12.5%)	NA	(21.7%)

1. Source: Data from Hoffman, McCabe and Smith (1994/Smith 2000) and Hoffman, McCabe, Shachat and Smith (1994/Smith 2000).

2. Refers to percentage of the N pairs in which the second player rejects the offer of the first.

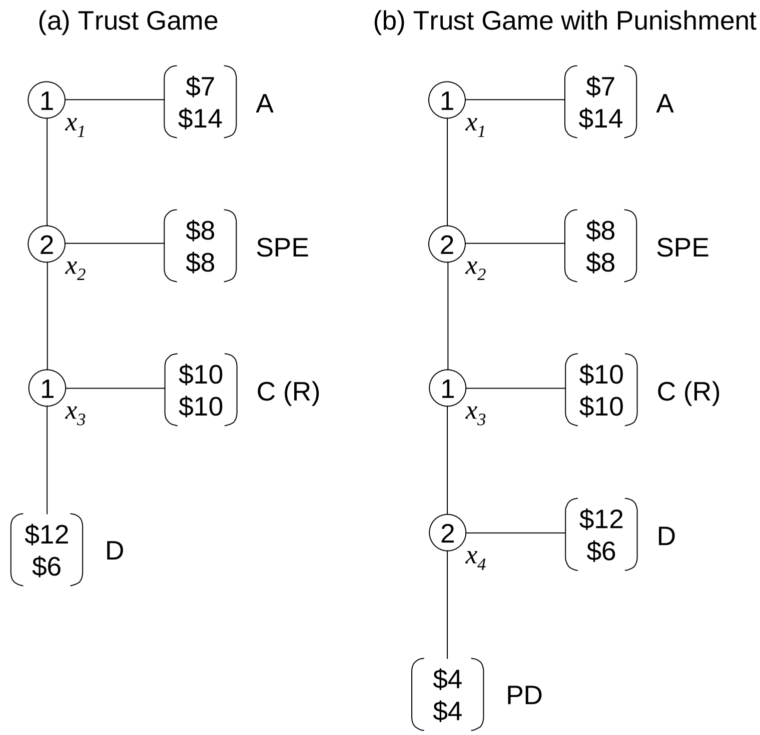
Table 2

Results: Trust and Trust/Punishment Game
Conditional Outcome Frequencies

Outcome	Trust	N	Node	Player	Trust/Punishment	N	Node	Player
A	0%	26	x_1	1	0%	29	x_1	1
SPE	56%	26	x_2	2	55%	29	x_2	2
C(R)	50%	12	x_3	1	85%	13	x_3	1
D	50%	6		1	50%	2	x_4	2
PD	NA	NA		NA	50%	1		2

Source: McCabe, Rassenti and Smith in Smith (2000, p. 163). For simplicity source data for larger trees (Game 1 and Game 2 in Smith, 2000) are trimmed to eliminate outcomes almost never reached, with commensurate reduction in sample size (N = 29 here, 30 in the source for the Punishment version). Each outcome frequency is computed conditional on the number of pairs reaching a particular node. Thus 26 pairs arrive at node x_2, of which 56% (14) choose SPE, leaving 12 to arrive at x_3, half of whom reciprocate by playing cooperatively.

Figure 1



A: Altruism
SPE: Subgame Perfect Equilibrium
D: Defection
C (R): Cooperation (Reciprocity)
PD: Punish Defection

Figure 2