# The Trust Paradox: A Survey of Economic Inquiries Into the Nature of Trust and Trustworthiness

**Harvey S. James, Jr.***
University of Missouri

April 23, 2001

Abstract: This paper examines the concepts of trust and trustworthiness in the context of a one-sided variation of the prisoner's dilemma, and it evaluates four different categories of solutions to the PD problem: changing player preferences, enforcing explicit contracts, establishing implicit contracts, and repeating the interaction of the players. Because these solutions rely on the creation of incentives to induce cooperation, this paper articulates a paradox of trust in that if one trusts another because there are incentives for the other to be trustworthy, then the vulnerability to exploitation is removed which gives trust its very meaning. The paper explores the implications of trust when understood to exist at two levels -- one in which there are incentives to trust, and the other in which appropriate incentives are absent.

Keywords: Trust, social norms, transaction costs, rationality, prisoner's dilemma

JEL Classifications: D23, D64, Z13

* Correspondence to: Harvey S. James, Jr., Department of Agricultural Economics, University of Missouri, 124 Mumford Hall, Columbia, MO 65211-6200, USA. Phone: 573-884-9682; Fax: 573-882-3958.

I am grateful for comments by Farhad Rassekh, two anonymous referees, and participants at a session of the Eastern Economic Association meetings and workshops at the University of Hartford and the University of Missouri.

# The Trust Paradox: A Survey of Economic Inquiries Into the Nature of Trust and Trustworthiness

## Introduction

What is trust? When should someone trust? Why should someone be trustworthy? How are trust and trustworthiness represented in economics? In the language of economics, trust can be viewed as an expectation, and it pertains to circumstances in which agents take risky actions in environments characterized by uncertainty or informational incompleteness. To say "*A* trusts *B*" means that *A* expects *B* will not exploit a vulnerability *A* has created for himself by taking the action. For instance, suppose *A*, in anticipation of realizing some potential gain, $g>0$, makes an investment, *c*. If the investment is specific to *B*, and if monitoring of *B* by *A* is costly, then *A* may trust that *B* is trustworthy. That is, *A* may expect (or hope) that *B* will not behave opportunistically by forcing a contractual renegotiation in an effort to appropriate quasi-rents generated by the investment (see Klein, Crawford, and Alchian, 1978).

The idea that some agents to a transaction may not behave honestly when environments are less-than-ideal is not a new one to economists. Indeed, an extensive literature describes the consequences of and solutions to environments in which agents have an incentive to exploit others. Examples include: *ex post* opportunism created by specific investments (Klein, Crawford, and Alchian); problems of principal-agent relationships created by moral hazard (Arrow, 1985; Holmstrom, 1979; Grossman and Hart, 1983) and adverse selection (Akerlof, 1970); free-riding resulting from team production externalities (Alchian and Demsetz, 1972); and general transacting costs associated with implicit contracting (Azariadis, 1975; Hart, 1983), incomplete contracting (Grossman and Hart, 1986; Hart and Moore, 1988), and self-enforcing agreements (Telser, 1980; Klein and Leffler, 1981).

When examining the question of whether or not an agent, *A*, should trust another agent, *B*, the economic approach is for *A* to assess the incentive *B* has either to honor or violate the trust offered by *A*. Then, if *B* has an incentive to be trustworthy will *A* trust *B*. This characterization of trust and trustworthiness reflects the economic assumptions that people are rational and strict utility maximizers, with the implication that people are honest only to the extent that honesty, or the appearance of honesty, pays more than dishonesty (see Sen, 1977; Telser, 1980). In economics, one is trustworthy if he does not have an incentive to exploit the trust of others. Or, conversely, "if the incentives are 'right,' even a trustworthy person can be relied upon to be untrustworthy" (Dasgupta, 1988, p. 54).

In order to illustrate the insight economics offers to issues of trust and trustworthiness, this paper examines the general class of problems based on the Prisoner's Dilemma (PD) game. In the PD, two players must decide whether or not to cooperate with (e.g., trust) each other, but ultimately find they have an incentive to doublecross that, ironically, results in a Pareto inferior outcome for the pair. Related to the PD are one-sided variations in which one player has an incentive to cooperate, but the other always has an incentive to doublecross. The purpose of this paper is to survey the variety of solutions to PD problems proposed by economists and to assess the nature of these solutions as they pertain to the issues of trust, honesty, cooperation, and trustworthiness. This analysis reveals that virtually all economic solutions to the PD and its one-sided variations require that the framework of the game be altered so as to create the *incentive* for individuals to cooperate. Of course, this is the natural solution for economists to offer. In models in which trustworthiness, honesty, or cooperation are strategic variables, economists argue that PD problems are "solved" only when it can be shown under what circumstances one agent will have an incentive to cooperate with or honor the trust offered by another. For example, according to transaction costs economics, the problem of *ex post* opportunism created by a specific investment is solved when *A* has secured for himself a contractual safeguard, perhaps enforced by a third party, to protect himself from the potentially opportunistic behavior of *B*. Other examples identified in the economic literature of incentives created to induce cooperation include, *inter alia*, the payment of fees to brokers by their clients in order to induce honest trading (Telser, 1972), the creation of costly devices to signal product quality

(Spence, 1973), investments in brand name or reputational capital to foster credible commitments in exchange relationships (Williamson, 1983), and the use of peer groups to monitor the repayment of microfinance loans to low income households (Morduch, 1999).

The problem with this approach is that, by changing the structure of the PD game so that agents have an incentive to cooperate, economists remove the vulnerability players face to exploitation by others, thus neutralizing the very nature of what it means to trust. That is, if I know my partner does not have an incentive to exploit my trust, does it make sense for me to say that I now trust her? The idea that trust is related to one's vulnerability is expressed by the following description: Trust "involves a recognition of one's vulnerability to the actions and choices of the trustee. It involves importantly, *retaining this vulnerability* by not attempting to erect barriers to protect one's interests" (Brien, 1998, p. 398; emphasis added). Indeed, the "purest trustee role serves principals in cases in which they [the principals] are unable to constrain their agent's performance. ... By definition, the principals of ... trust are vulnerable and impotent" (Shapiro, 1987, pp. 634-635).

This paper explores the meaning of trust and its use in economic models derived from or related to the PD. A principal objective is to advance the argument that there is a distinction between trust created through incentives and trust applied to situations in which players retain a vulnerability to the actions and choices of others. Such a distinction is important because economists generally have little understanding of trust and its role in economic exchanges (Williamson, 1993; Perelman, 1998), particularly in the absence of incentives for trustworthiness in others. While not the first paper to argue that there are different types of trust (see Zucker, 1986; Williamson, 1993; Lyons and Mehta, 1997; Lorenz, 1999), this paper surveys research in which trust and cooperation are modeled, describes the general categories by which economists create incentives for trust, and articulates the complications that arise by basing trust on either the presence or absence of appropriate incentives for trustworthiness.

## Trust and the Prisoner's Dilemma

Economists and social psychologists generally equate trust and trustworthiness with cooperation in a one-shot or repeated game of the Prisoner's Dilemma (PD) (see Deutsch, 1960; La Porta, Lopez-de-Silanes, Shleifer, and Vishny, 1997). Trust arises in "situations in which the risk one takes depends on the performance of another actor" (Coleman, 1990, p. 91). To trust means you rely on others not to take advantage of you. To be trustworthy means you do not take advantage of others when trusted. Generally, "when we say we trust someone or that someone is trustworthy, we implicitly mean that the probability that he will perform an action that is beneficial or at least not detrimental to us is high enough for us to consider engaging in some form of cooperation with him" (Gambetta, 1988, p. 217).

The PD is interesting to economists because the behavior of participants playing the game reflects the patterns of behavior observed under many "real-world" circumstances when individual and collective interests appear to conflict (Axelrod, 1984). Consequently, it provides an informative framework for examining the nature of ethical behavior (James, 1998), such as trust and cooperation. The PD is also interesting because experimental evidence indicates that people are often more willing to cooperate in PD environments than the "theory" predicts (Dawes, 1980; Axelrod, 1984; see also Ridley, 1997), although individuals trained in economics seem to be relatively less likely to cooperate than non-economists (Marwell and Ames, 1981; Frank, Gilovich, and Regan, 1993; Frank and Gunther, 2000). That people have a tendency to trust, even though there are incentives not to do so, makes the examination of trust of particular relevance to economists trained to appreciate the importance of incentives in affecting individual and group behavior.

As an example, consider the simultaneous-move, one-sided variation of the PD called the Trust-Honor Game (THG) illustrated in Figure 1. Agent I can either Trust or Distrust Agent II, who in turn can either Honor the trust or Exploit it. If both players cooperate by selecting the strategy pair [Trust, Honor], then they each receive a payout of $w>0$. If Agent I trusts but Agent II exploits that trust, then Agent I incurs a loss of $y>0$, while Agent II receives a premium of $x>0$ over $w$. If Agent I distrusts and Agent II is

willing to Honor, then Agent II incurs a cost of $z>0$; otherwise, the players receive nothing. Making the standard assumptions of player rationality, stable preferences for more rather than less, and common knowledge, it is clear that Agent II has a dominant strategy to exploit. But Agent I does not have a dominant strategy. Agent I would be willing to trust if Agent II is willing to honor that trust. However, since Agent I knows that Agent II has a dominant strategy to exploit, Agent I distrusts, resulting in the Nash equilibrium of [Distrust, Exploit].

| | | Agent II | |
|---|---|---|---|
| | | Honor | Exploit |
| Agent I | Trust | $w, w$ | $-y, w+x$ |
| | Distrust | 0, $-z$ | 0, 0 |

**Figure 1**. Trust-Honor Game

It is old news to state that in the THG and other variations of the PD, rational individuals will choose actions that result in the Pareto inferior outcome for the players. Indeed, barring a binding agreement between the agents or some other refinement to the game, economic logic tells us that rational agents will *have* to settle for the inferior outcome even though each would do better -- in fact, prefer -- the outcome in which the players cooperate (in that Agent II honors the trust offered by Agent I).

## Solutions to the Prisoner's Dilemma

There are a number of different types of "solutions" proposed by economists for the PD and its variations, each of which has the objective of motivating the players to cooperate (e.g., Trust and Honor). These solutions consist of refinements or alterations to the PD so as to induce the agents to take actions

that ultimately result in mutually beneficial outcomes. The necessity of implementing refinements derives from the fact that, although both players prefer outcomes characterized by Trust and Honor (in the THG) relative to outcomes resulting from Distrust and Exploit, they know that the incentives of the game are such that Agent II will be "rationally-compelled" to take advantage of cooperating play by the other agent. Simply stated, Agent I knows that Agent II has an incentive to be untrustworthy, so Agent I does not have an incentive to trust. The logic of the economic solutions is that if incentives are what drive the non-cooperative outcome, then incentives can be used to induce cooperation. Therefore, the objective of economic models is to alter the PD environment so as to create appropriate incentives for cooperative play. In the context of the THG, we must change the game so that Agent II has an incentive to honor the trust of Agent I, thus providing an incentive for Agent I to trust.

There are four general categories of solutions to PD problems proposed by economists: changing the preferences of players, writing an explicit contract, relying on an implicit social contract, and repeating the interaction. These will be surveyed and illustrated using the THG to show how economists incorporate the use of incentives to foster trust and trustworthiness.

### *1. Change the preferences of players*

One way to solve the dilemma illustrated in the THG is to introduce either exogenous or endogenous changes to the preferences of the players so that they "prefer" to cooperate rather than pursue myopic self-maximizing strategies. In the case of exogenous changes, we can assume that Agent II has preferences for honoring trust. For example, this may involve adjusting the perceptions by Agent II about the value of $x$ (e.g., $x<0$ rather than $x>0$) or assuming that the marginal utility of $x$ is negative so that Agent II no longer has a dominant strategy to Exploit the trust offered by Agent I. In the case of endogenous changes, we can introduce emotional predispositions for cooperation that generate feelings of guilt for non-cooperation (Frank, 1987) or improve genetic fitness in an evolutionary environment (Güth and Kliemt, 1994), thus resulting in the internalization of cultural norms which place taboos on the

realization of benefits achieved at the expense of others (Hodgson, 1996; see also Axelrod, 1997, ch. 3). Indeed, one view of culture is that it "changes the tastes of the individuals directly, so that the resulting behavior is the desirable one, even in the absence of any explicit penalties" (Lazear, 1995, p. 98).

An example of a model that incorporates exogenous changes to player preferences is Kandel and Lazear (1992), who incorporate variables into player utility functions to reflect feelings of shame or guilt for dishonest or exploitative behavior. Specifically, Kandel and Lazear introduce a "peer pressure" function, $P(e_i; e_j, ..., e_N, a_i, a_j, ..., a_N)$, in their analysis of the free-rider problem. From this function "the pressure that worker *i* feels depends on his own effort, $e_i$; on the effort of his peers, $e_j, ..., e_N$; and on other actions that he and his peers may take, $a_i, ..., a_N$" (p. 804). Kandel and Lazear use this function to introduce feelings of guilt and shame in order to induce higher levels of equilibrium effort by workers. Bøhren (1998) offers a similar proposal by suggesting that an agent's utility function be modified to include not only terms for wealth, *W*, and effort, *e*, but also a variable that captures preferences for honest and dishonest uses of private information, *I*. Such a utility function would be of the form U=f(*W*, *e*, *I*), where *W* and *e* positively and negatively affect utility, respectively, and *I* reduces utility as the consumer utilizes his private information for dishonest uses. In this way agents may derive discomfort from "improper" uses of information or from unethical behavior, thus inducing them to behave cooperatively.

An example of an endogenous formulation is proposed by Lazear, in which preferences evolve so as to promote cooperation in a genetic model of corporate culture. In this model, workers have one of two preferences (i.e., genetic endowments) for work as follows: Type *A* workers, whose offspring survive with probability *f*, are more productive in structured environments, and Type *B* workers, whose offspring survive with probability $f(1-s)$, are more productive in unstructured environments. Lazear shows that if managers can manipulate the environment by making investments in *s*, they can, over time, increase the proportion of workers with preferences for hard work and cooperation within a structured (i.e., firm) environment.

Frank develops an evolutionary model to show that "it will sometimes be in a selfish person's interest to have a utility function that predisposes him not to cheat, even when he is *certain* he would not

be caught" (p. 593; emphasis in original). In his model, a population consists of two types of members -- those who are honest (H) and those who are dishonest (D), where honesty means that one refrains from cheating another, even when cheating cannot be detected and punished. Members can choose to pair with someone in a joint venture or to work alone. The nature of the incentives are such that two H-types working together do better collectively than if the players work alone or when one or both are of type D. However, D-type players can improve their rewards by exploiting the honesty of the H-types, thus creating PD incentives in the joint venture. Each individual in the population has an observable characteristic, *S*, that is based on the person's true type (H or D) and a random variable (thus the signal is imperfect). Frank shows that in a population with a relatively low proportion of type H persons, only those with high values of *S* will be willing to bear a cost of becoming sensitized to inspect the *S* values of others, thus increasing the chance that honest types will interact and improving the equilibrium numbers of honest persons in the population over time. Güth and Kliemt develop a similar model to show how the cost of detection technology affects the composition of a population with both cooperative and non-cooperative members.

Huang and Wu (1994), Huck (1998), and Bohnet, Frey, and Huck (2000) present models of preference change that incorporates both exogenous and endogenous features. In the model by Huang and Wu, agents face payoffs that include a psychological variable reflecting the remorse they feel when they betray the trust of a principal. The degree of remorse agents feel is based on their expectations regarding how many other agents betray their principals. Huang and Wu show that the less prevalent corruption is, the stronger is the remorse agents feel for betraying others. They conclude that principals can affect the propensity of others to engage in unethical conduct by highlighting their (the principal's) exemplary behavior and by cultivating expectations that corruption will not be tolerated. Huck and Bohnet *et al* develop models that include psychological variables to reflect player remorse, as well as evolutionary features that endogenize preference formation over time. Huck shows that the choice of legal rules, rather than the behavior of leaders as in Huang and Wu, affects the viability of certain types of preferences, and that appropriately chosen legal rules can promote the development of preferences for truth-telling and

honest behavior. Bohnet *et al* show that preferences for trustworthiness can be fostered with weak enforcement rules, but "crowded out" with medium enforcement policies.

The upshot of this analysis is that trust and trustworthiness can be achieved by fostering preferences for cooperative behavior. Of course, changing player preferences so that they prefer honor and trust over exploitation and distrust, as in the exogenous models, creates the potential for a tautological argument: Just assume the preferences necessary to get the outcome you want. While endogenous models do not suffer from this weakness, they are generally long-term solutions that require time and repeated exposure to environmental stimuli to affect the requisite changes in player preferences, although the length of time for appropriately developed preferences might be relatively short (see Bohnet *et al*). By their nature, endogenous models do not produce cooperation in single-period exchanges.

### *2. Write an explicit contract*

Another way to solve the dilemma inherent in the THG is to write an explicit and binding contract, enforced by a third party, requiring participants to select strategy pair [Trust, Honor]. These contracts take one of two forms: (1) monitoring with punishment, and (2) monitoring with incentives. In the first type of contract, the parties can agree that Agent I will monitor the choice of Agent II, perhaps at some cost, $c$, and that Agent I will call on a third party to punish Agent II for exploitation or to compel cooperation. For instance, consider the revision of the THG depicted in Figure 2(a). In this version, $c$ is the monitoring cost to Agent I for observing and verifying to third parties the choice of Agent II, and $d$ is the punishment Agent II incurs by the third party if caught exploiting by Agent I. So long as the cost of observing and verifying exploitation is not too large (i.e., if $c<w$) and the punishment Agent II incurs for exploitation is significant (i.e., if $d>x$), then [Trust, Honor] becomes a Nash equilibrium.

In the second type of contract, Agent I can offer an incentive, $a$, to Agent II to honor his trust, as illustrated in Figure 2(b). In this type of contract, Agent I has an incentive to trust as long as the cost of

monitoring and rewarding performance is not too large (i.e., $c+a<w$), and Agent II has an incentive to Honor Agent I's trust as long as the reward is substantial (i.e., $a>x$).

| | | Agent II: Honor | Agent II: Exploit |
|---|---|---|---|
| Agent I | Trust | $w-c, w$ | $-y-c, w+x-d$ |
| | Distrust | $0, -z$ | $0, 0$ |

**Figure 2(a)**. Trust-Honor Game with monitoring and punishment in an explicit contract

| | | Agent II: Honor | Agent II: Exploit |
|---|---|---|---|
| Agent I | Trust | $w-c-a, w+a$ | $-y-c, w+x$ |
| | Distrust | $0, -z$ | $0, 0$ |

**Figure 2(b)**. Trust-Honor Game with monitoring and incentives in an explicit contract

This approach of relying on explicit contracts to solve the THG is characteristic of solutions derived from transaction cost economics, the basic insight of which is that when (some) individuals are assumed to be opportunistic -- defined by Williamson as "self-interest seeking with guile" in that they "mislead, distort, disguise, obfuscate, or otherwise confuse" (1985, p. 47) -- efforts to screen *ex ante* and to create safeguards *ex post* are required. Typically, though not always, the *ex post* safeguards are developed within the context of explicit contracts, which, according to Williamson, "would be ubiquitous in the face of nonopportunism" (1985, p. 66).

For example, in the classic principal-agent relationship (see Ross, 1973, for an early treatment), an agent is recruited to supply an effort, *e*, which generates a benefit, *b*(*e*), for a principal. Because effort is (assumed to be) costly to the agent, the agent may have an incentive to supply less than the optimal effort level, or, more generally, to act in his rather than his principal's interest. Additional problems arise for the principal when the agent's effort is costly to observe or verify to third parties, the agent supplies more than one type of task, or the benefit is a function not only of the worker's effort but also of unobservable random variables and efforts of other workers. In these circumstance agents have an incentive to behave "opportunistically." Consequently, the principal's task is to design a contract such that the agent has an incentive to take the action, *e**, that maximizes the principal's benefit, net of the wage, *w*, paid to the agent. Such contracts will likely entail varying degrees of supervisory monitoring, rewards, and punishments to provide the requisite incentive to the agent to act in the principal's interest (see Hart and Holmstrom, 1987; Levinthal, 1988; Gibbons, 1998; Predergast, 1999; and James, 2000).

The point of this approach is that trust and trustworthiness can be promoted through the use of explicit contracts, which involve monitoring and either the payment of incentives or provisions for third-party enforcement (or, more generally, hybrids of these two elements). The problem with explicit contracting, however, is that it is a "second best" solution to the PD problem, in part because monitoring is costly, incentives can be distorted, and third party enforcement requires that agent's action be observable and verifiable.

Furthermore, the explicit contracts creating the [Trust, Honor] Nash equilibrium as a solution of the THG do not negate the fact that [Distrust, Exploit] remains a Nash equilibrium, creating both commitment and coordination problems for the players. In terms of the commitment problem, if Agent II chooses to exploit the trust of Agent I, then Agent I will have an incentive not to enforce the contract, a fact known and thus potentially exploitable by Agent II (see Miller, 1992, for an extensive treatment). Moreover, with two pure strategy Nash equilibria, the players will also face the existence of mixed strategy equilibria. This would be manifested as a coordination problem in that each player, before committing to cooperate, would have to develop and rely on probabilistic beliefs that the other player will

indeed cooperate. For instance, if $p$ is the probability that Agent I will Trust, and if $q$ is the probability that Agent II will honor his trust, then the agents will have an incentive to cooperate in contracts with monitoring and third-party punishment when Agent II believes $p \geq \frac{z}{z+d-x}$ and Agent I believes $q \geq \frac{y+c}{w+y}$. Similarly, the agents will cooperate under incentive contracts when Agents II and I have beliefs of $p \geq \frac{z}{z+a-x}$ and $q \geq \frac{y+c}{w-a+y}$, respectively. Thus, other considerations may be necessary to ensure that players have appropriate expectations from explicit contracting in order to provide the requisite incentives for them to cooperate.

### *3. Rely on an implicit social contract*

Related to the solutions based on changing agent preferences and utilizing explicit contracting is the reliance on social rewards and punishments to achieve mutual trust and cooperation. The basic distinction between explicit and implicit contracts is that implicit contracts cannot be enforced by third parties, such as courts. Only the parties to the contract, or those affected by its non-compliance, can determine whether the agreement has been violated, and only they can enforce the contract. Social enforcement typically involves either the threat to nullify or discontinue the interaction if the other player does not cooperate, or the leveling of social sanctions (e.g., ostracism) against the non-cooperator by the player.

An example of an implicit social contract in the context of the THG is given in Figure 3. In this refinement, $c$ is the cost to Agent I of monitoring the choice of Agent II, and $e$ is the effort Agent I expends to inflict the social sanction of $f$ on Agent II for exploitation. As long as $w$-$c$>0 and $f$>$x$, then [Trust, Honor] becomes a Nash equilibrium strategy for the players. However, as with the case of explicit contracting, implicit social contracts may also have multiple equilibria. Mutual trust and cooperation [Trust, Honor], as well as non-cooperation [Distrust, Exploit], are potential outcomes of the THG when

implicit contracts are utilized. Consequently, players will also have to form appropriate beliefs regarding the probability that the other player will choose the cooperative strategy. As an illustration, suppose $p$ is the probability that Agent I chooses Trust, and $q$ is the probability that Agent II chooses Honor. Agent I will trust Agent II when he believes that Agent II will honor the trust with a probability of $q \geq \frac{y+c+e}{w+y+e}$. Similarly, Agent II will honor when he expects Agent I to trust with a probability that is $p \geq \frac{z}{z+f-x}$.

| | | Agent II: Honor | Agent II: Exploit |
|---|---|---|---|
| Agent I | Trust | *w-c*, *w* | *-y-c-e*, *w+x-f* |
| | Distrust | 0, *-z* | 0, 0 |

**Figure 3**. Trust-Honor Game with implicit social contracting

There are many models that rely on implicit, social relations as a solution to opportunistic behavior of agents. Consider, for instance, Telser's (1980) model of self-enforcing agreements. In this model, a buyer and a seller interact in some time period, $t$. The buyer offers to pay a price, $x_t^*$, for a good that provides a benefit, $b_t^*$, and that imposes a production cost of $a_t^*$ on the seller. Telser shows that if the probability is sufficiently low that time $t$ will be the last period of interaction, then there exists a price, $x_t^*$, such that the buyer and seller have an incentive to honor their contractual agreement. This is because the transactors know that if either party violates the current agreement (e.g., if the buyer pays $x_t<x_t^*$ or the seller supplies $b_t<b_t^*$), then trade will never occur in subsequent periods. Thus, the potential loss of future benefits provides an incentive for players to cooperate in the current period.

Kandori (1992) presents a similar model in which participants choose either to cooperate or defect in a pair-wise matching of a PD game. He shows that when a player can observe the past actions of

his partner, results similar to those of Telser (1980) emerge. However, when players cannot observe the history of play, then the only equilibrium in which cooperation is sustained is one that he calls the "contagious equilibrium." In this equilibrium, defection by a single player breaks down the trust of the entire community so that no player cooperates after observing defection in another player. Such a stark solution to the PD can be avoided, however, if the community of players is able to label members according to their play in previous rounds of the PD. Thus, "when two players are matched, they observe each other's labels first and then take some action. After that, their labels are updated depending only on *their* original labels and actions by a given rule," which rule embodies a "social norm" (Kandori, p. 71; emphasis in original). Thus, an uncooperative label defines a "social sanction," and such "informal sanctions by community members can induce desirable behavior in infrequent trades" (p. 76).

In Spagnolo (1999), a group of identical workers can choose either to cooperate or not cooperate (defect) with each other in pair-wise social and production interactions characterized by PD incentives. Spagnolo shows that workers have an incentive to cooperate in both the social and production interactions if they trust each other -- that is, if the probability agents assign to the event that the other worker cooperates is large enough. If the social and production interactions are linked, in that the same two agents face each other in both the social and production PD games, then agents can rely on trust derived from social cooperation to promote cooperation during production interactions. Thus, firm owners can facilitate cooperation in the production relation by linking the social and production relations and by encouraging cooperation in social contexts.

Because players must make personally costly efforts to enforce cooperative agreements, a weakness of socially based contracts is that there exists the potential for free-riding, particularly when play involves a community of members. Indeed, enforcement has characteristics of a public good so that each player prefers that other members take the costly action to sanction. Furthermore, these models require that the interactions among participants be repeated or long term. Agents participating in a one-shot play of the PD typically will not be able to rely on the use of social, implicit solutions to foster trust and trustworthiness.

### *4. Repeat the interaction*

A final solution to PD problems is to repeat the interaction between the agents. That is, Agents I and II can continue to play the stage game from Figure 1, knowing (perhaps with some probability) they will repeat the interaction another time. Such a repetition allows application of the "folk theorem" for infinitely repeated games, which asserts that virtually any outcome can be enforced by an equilibrium, so long as the probability of repeating the game is high enough (or, as long as players are sufficiently patient). Thus, cooperative outcomes can be achieved with an appropriately defined strategy of play, which generally involves one player punishing, in subsequent rounds of play, exploitation occurring in the current round. In other words, repetition as a solution to the PD problem generally also requires the implementation of either implicit or explicit contractual arrangements to facilitate the punishing of agents who abuse the cooperation and trust of others.

Two well-known strategies for punishing non-cooperators in repeated PD games are *tit-for-tat* and the *trigger strategy*. According to *tit-for-tat*, a player cooperates in the first period and thereafter mimics the other player's choice from the previous round of play (see Axelrod, 1984, 1997). According to the *trigger strategy*, a player cooperates until the other player exploits, resulting in the first player refusing to cooperate in all subsequent rounds of play (see Green and Porter, 1984, for an application). For both of these strategies, players have an incentive to cooperate as long as they expect to interact in a subsequent round of play, the reward to cooperation is substantial, and the gain to exploitation is not too large.

As an illustration in the context of the THG, if we assume the game will be repeated with some probability, $\beta$, where $0<\beta<1$, then Agent II will have an incentive to honor the trust offered by Agent I if the expected value to Agent II of Honor exceeds the expected value of Exploit. That is, [Trust, Honor] becomes a Nash equilibrium strategy when

$$\mathrm{EV}_{\mathrm{II}}(\text{Honor} \mid \text{Trust}) = \sum_{t=0}^{\infty} \beta^t w \geq w + x = \mathrm{EV}_{\mathrm{II}}(\text{Exploit} \mid \text{Trust}),$$

or when $\beta \geq \frac{x}{w+x}$. Notice that $\lim_{x \to \infty} \frac{x}{w+x} = 1$. That is, as the premium, *x*, to Agent II of choosing Exploit increases, the probability of repeating the interaction must also increase, until at the limit the participants must be perfectly certain (i.e., ß=1) that the interaction will occur again.

There are a plethora of models based on the idea that repeated interactions have the potential of overcoming PD problems in exchange relationships, in addition to the models described above. Klein and Leffler develop a model in which trading partners participate in a market with the expectation that sellers will experience repeat purchases. Sellers produce goods of either high ($q_h$) or low ($q_{min}$) quality, with the minimum average cost of $q_h$ exceeding the minimum average cost of $q_{min}$. In order to give sellers an incentive to honor contracts to supply $q_h$, buyers offer to pay a price in excess of the marginal cost of producing $q_h$, whereas the market price equals the marginal cost of $q_{min}$, thus providing sellers with a perpetual stream of quasi rents. Sellers risk losing the quasi rents if they ever cheat by supplying $q_{min}$ when they agree to supply $q_h$ at the higher market price. Moreover, "a firm that cheats will become known as a 'notorious' cheater, and consumers will not purchase from the firm any product the quality of which cannot be determined prepurchase" (p. 621).

Kreps' (1990) discussion of corporate culture provides a similar argument for the role of reputation and repetition in exchanges in order to overcome the (short run) incentives agents have to deal opportunistically with others. He states that as long as a firm's opportunities in later rounds are tied to their behavior in earlier rounds, then businesses that have a reputation for honest dealing will have decided advantages over those that do not have such a reputation. Corporate culture, according to Kreps, not only acts as a "reputation bearer," but it also communicates to corporate citizens the rules that they must follow in their interactions with potential future trading partners so that the corporation's reputation can be perpetuated. In a related discussion, Dasgupta argues that individuals can make appropriate investments in reputations that foster formations of trust over time.

Neilson (1999) develops a model in which two players interact in an infinitely repeated PD game. In each period, each player has an opportunity to take a personally costly action that provides a benefit to the other player. Neilson shows that players are only willing to make the costly "favors" when they expect to receive favors in return at some future date. That is, Neilson assumes that "individual rationality requires that a player only performs a favor if the performance of that favor makes him better off than if he does not perform that favor" (p. 390). Consequently, players will assess the discounted future benefits derived from making a currently costly favor in order to determine whether or not to grant the favor. Thus, some favors may not be performed if they are too costly, and some socially inefficient favors may be performed because they are personally beneficial to the individual. Nevertheless, the expectation that interactions will occur in the future drives the results that participants have an incentive to provide favors.

The common element of these and similar models is that repetition in the exchange relationship provides in part a means of overcoming the short run incentive for players to exploit the cooperative play of others. The reason is that exploitation, when detected, generally results in a punishment such that offenders are denied future opportunities of entering into exchange relationships with those they have cheated. The drawback, however, is that there must be an expectation that the interaction will be repeated for cooperation to result. Random or one-shot exchanges will not produce cooperation in these models.

## Discussion

Trust and trustworthiness are important elements in economic exchanges. Without trust, "no market could function" (Arrow, 1973, p. 24), which explains, for example, why Fukuyama (1995) concludes that social capital or "trust" is just as important as physical capital in facilitating the creation of large-scale business organizations necessary for economic growth and development. But, what exactly does it mean to trust, and how are trust and trustworthiness fostered? According to the economic literature, trust can be achieved in a roundabout way through institutional and other devices that give individuals an incentive to be trustworthy. These solutions generally involve the changing of player

preferences, the use of either explicit or implicit contractual arrangements, or repetition in the exchange relationship. Often, solutions involve combinations of these particular types (see Zucker; Shapiro).

The common element of each of the general types of solutions is that the assumptions or framework of the PD game are altered in order to create the incentive for Agents to cooperate (by choosing Trust and Honor in the THG, for instance). This is a standard economic "solution." As Hausman and McPherson state: "What economists would like to show is that norm-following increases profits of firms or the consumption or leisure of individuals…. Individuals comply with norms because of the individual benefits attached to doing so and because of the costs of not doing so" (1996, pp. 58, 59). The problem with this approach, however, is that if agents require an incentive to cooperate, then as long as the standard assumptions of rationality and self-interest hold it is trivially obvious why they are willing to cooperate with or trust each other. If Agent II has an incentive to honor trust, and if Agent I knows this, then it is "prudent" for Agent I to trust Agent II. But, would Agent I ever trust if it is not prudent? If people are willing to trust, even when they know others have an incentive to behave opportunistically, then trust and trustworthiness can be viewed from two different perspectives. On the one hand, I may trust because it is prudent for me to do so, if I believe (rationally) that my partner has an incentive to be trustworthy. On the other hand, if I believe that my trading partner retains an incentive to exploit my trust, then I may still choose to trust. While prudence may suggest otherwise, if I choose to trust I do so out of the "hope" that my partner will not exploit my trust. Trusting out of hope, rather than prudence, cannot be rationally justified (see Lahno, 1995).

The distinction between "trust as prudence" and "trust as hope" is important, and it mirrors, though not identically, the distinction Williamson (1993) makes between "calculative trust" and "personal trust." Calculative trust -- which Williamson believes is a contradiction in terms -- is based on a rational assessment of the costs and benefits of trusting. Personal trust, on the other hand, involves no conscious mental calculation but "is warranted only for very special personal relations that would be seriously degraded if a calculative orientation were 'permitted'" (Williamson, 1993, p. 486). This distinction is also similar to the one Sen (1977) makes between "sympathy" and "commitment." A person is sympathetic

when his concern for another's welfare directly affects his own utility. By contrast, a person is committed if she is willing to undertake an activity that clearly conflicts with her self-interest and sympathetic preferences (i.e., if the activity does not benefit her).

The distinction between "trust as prudence" and "trust as hope" is also important because if Agent II no longer has an incentive to honor the trust offered by Agent I, then we may be forced to conclude that Agent I will no longer trust or that trusting is not "rational." Though it may not be prudent for Agent I to trust Agent II when incentives for doing so are absent, what if he did? What if the economic incentives for trust are not present, but agents trusted nonetheless? For example, in solutions based on a repetition of the PD, the probability of a repeated interaction must be large enough in order for the threat of lost future gains from cooperation to become a strong enough deterrent to non-cooperation. That is, in the THG described above, Agent I has an incentive to trust Agent II, who has an incentive to honor that trust, when $\beta \geq \frac{x}{w+x}$. But, what if $\beta < \frac{x}{w+x}$? What if the probability that the interaction will repeat is too low? Will Agent I never trust Agent II? Or, will Agent I decide to "take a chance" and hope that Agent II will not choose to exploit? In spite of the theoretical results that players in a finitely-repeated game of the PD will not cooperate in or near the final round of play (Kreps, Milgrom, Roberts, and Wilson, 1982), experimental evidence confirms that when players know the interactions between them are finite, many still cooperate, suggesting that people do in fact trust when vulnerable to exploitation. For instance, Tullock (1999) finds that in a 10 round play of a prisoner's dilemma, in which players are allowed to choose their opponents and communicate, he observed "heavily cooperative play" (p. 456). Similar findings are presented from experiments in which participants cannot communicate with their trading partners (see Berg, Dickhaut, and McCabe, 1995; Cohen and James, 2001).

In describing the theoretical research on finitely repeated games, Sen (1988) remarks as follows:

> In the formal literature, there have been various attempts to explain the emergence of cooperation by introducing some kind of 'defect' in either the knowledge, or the reasoning of the players. The players could be ignorant as to how many times the game would be played, thereby making the backward induction impossible. Or, the players may not know fully what the other players' objective or knowledge happens to be, and they may

> entertain the belief -- falsely as it happens -- that others would actually enjoy cooperation and respond accordingly. Or the possible behavior patterns that may be considered may be arbitrarily limited in some special way without all variational possibilities being examined.
>
> These 'defects' may indeed be present, but it is also possible that cooperative behaviour has quite a different explanation. Indeed, such cooperation is often found even in *non-repeated* games of this type, in one-off real-life situations. It could, of course, be the case that the real goals of a person are not ones that they believe they are trying to maximize. But it is also possible that people clearly understand their goals, wish to maximize them, but nevertheless take note of other people's goals, due to a recognition of the nature of mutual interdependence of the achievements of different people in these situations. (pp. 84-85; emphasis in original).

In short, the possibility could exist that, in the absence of incentives for trustworthiness by others, people have a predisposition to *just* trust each other (see Ridley). This idea is supported by additional experimental research indicating that people with a high degree of trust are more willing to cooperate than individuals with low trust levels, but, interestingly, *only* when there is a possibility that cooperators will not receive a payoff from their actions (i.e., if they expect to be exploited). Without the vulnerability or fear of loss, high-trust and low-trust individuals cooperate at the same rates (Parks and Hulbert, 1995).

## Conclusions

This paper has examined the nature of trust in a PD environment as characterized in the economics literature. From the economic point of view, trust is an expectation that one will not be exploited by another, which exists when there are no strong incentives for players to behave opportunistically. Though we often think of rationality in terms of the consequences of actions, this is not sufficient for explanations of trust, in part because people often trust when they retain a vulnerability to exploitation. Unlike classical game theoretic analyses, in which behavior is determined by the payoffs of the game (see Sugden, 1989), trust can exist even in the absence of explicit incentives. Is it really "trust" when you know your partner has an incentive to be trustworthy? Trust when vulnerability is retained is a much stronger concept than trust derived from incentives.

Are there solutions out of the trust paradox? One possibility is to stop calling it "trust" when cooperation is induced by incentives for trustworthiness in others. That is, "trust as prudence" may not be "trust" at all (see Williamson, 1993; Craswell, 1993). The problem with this approach is that it begs the question of why people "trust" when doing so does not involve incentives for trustworthiness. An alternative approach is to continue to accept the notion that individuals take actions to satisfy their interests, however those interests are defined (see Coleman, ch. 5; Becker, 1993), but then to examine why some still find it in their interest to trust when they retain a vulnerability to exploitation. If people *just* trust each other, even when there is a possibility of being exploited, they may still feel it is in their interest to do so, but why that is true is an interesting and unresolved avenue of study for economists. Perhaps, as Ridley claims, people are *just* genetically programmed to trust others -- an interesting puzzle for economists.

Clearly, more effort needs to be directed at determining how trust and trustworthiness are developed and utilized in economic contexts. At a minimum we know that trust can be fostered by the development of shared common experiences and group identity (see Dawes and Thaler, 1988). Communication is one element of a shared common experience, as illustrated by experimental evidence showing how communication and the development of a group identity improves the likelihood that players in a PD environment cooperate (Valley, Moag, and Bazerman, 1998; Tullock; see also Dawes and Thaler). But other questions remain. First, will the development of incentives to foster trust erode trust? There is some evidence that extrinsic incentives "crowd out" intrinsic incentives (Frey and Oberholzer-Gee, 1997; Frey and Jegen, 1999), suggesting that attempts to foster trust via incentive mechanisms may ultimately undermine trust. Second, if trust without incentives erodes, so that institutional or contractual solutions are required to restore trust through the use of economic incentives, can trust without such incentives be restored? Finally, if economic agents are willing to trust, even in the absence of incentives, what determines when agents are willing to do so? These are important questions that need to be examined in order to expand the economic understanding of trust.

## References


Akerlof, George A., 1970, The Market for 'Lemons': Quality Uncertainty and the Market Mechanism, Quarterly Journal of Economics, 84(3), pp. 488-500.

Alchian, Armen, and Harold Demsetz, 1972, Production, Information Costs, and Economic Organization, American Economic Review, 62, pp. 777-795.

Arrow, Kenneth J., 1973 Information and Economic Behavior, (Federation of Swedish Industries, Stockholm, Sweden).

Arrow, Kenneth J., 1985, The Economics of Agency, in J. Pratt and R. Zeckhauser (eds.) Principals and Agents: The Structure of Business, (Harvard University Press, Boston), pp. 37-51.

Axelrod, Robert, 1984, The Evolution of Cooperation, (Basic Books, New York).

Axelrod, Robert, 1997, The Complexity of Cooperation, (Princeton University Press, Princeton).

Azariadis, C., 1975, Implicit Contracts and Underemployment Equilibria, Journal of Political Economy, 83, pp. 1183-1203.

Becker, Gary S., 1993, Nobel Lecture: The Economic Way of Looking at Behavior, Journal of Political Economy, 101(3), pp. 385-409.

Berg, Joyce, John Dickhaut, and Kevin McCabe, 1995, Trust, Reciprocity, and Social History, Games and Economic Behavior, 10(1), pp. 122-142.

Bøhren, Øyvind, 1998, The Agent's Ethics in the Principal-Agent Model, Journal of Business Ethics, 17, pp. 745-755.

Bohnet, Iris, Bruno S. Frey, and Steffen Huck, 2000, More order with Less Law: On Contract Enforcement, Trust and Crowding, working paper no. 00-009, Kennedy School of Government, Harvard University.

Brien, Andrew, 1998, Professional Ethics and the Culture of Trust, Journal of Business Ethics, 17, pp. 391-409.

Cohen, Jeffrey P., and Harvey S. James, Jr., 2001, If Teaching Economics Discourages Cooperation, Can the Damage Be Undone? working paper, Department of Economics, University of Hartford.

Coleman, James, 1990, Foundations of Social Theory, (Harvard University Press, Cambridge, MA).

Craswell, Richard, ,1993, On the Uses of 'Trust': Comment on Williamson, 'Calculativeness, Trust, and Economic Organization,' Journal of Law and Economics, 36(April), pp. 487-500.

Dasgupta, Partha, 1988, Trust as a Commodity, in D. Gambetta (ed.), Trust: Making and Breaking Cooperative Relations, (Blackwell, New York), pp. 49-72.

Dawes, Robyn M., 1980, Social Dilemmas, Annual Review of Psychology, 31, pp. 169-193.

Dawes, Robyn M., and Richard H. Thaler, 1988, Cooperation, Journal of Economic Perspectives, 2(3), pp. 187-197.

Deutsch, M., 1960, Trust, Trustworthiness, and the F-scale, Journal of Abnormal and Social Psychology, 61, pp. 138-140.

Frank, Bjorn, and Gunther G. Schulze, 2000, Does Economics Make Citizens Corrupt? Journal of Economic Behavior and Organization, 43(1), pp. 101-113.

Frank, Robert H., 1987, If Homo Economicus Could Choose His Own Utility Function, Would He Choose One With a Conscience? American Economic Review, 77(4), pp. 593-604.

Frank, Robert H., Thomas Gilovich, and Dennis T. Regan, 1993, Does Studying Economics Inhibit Cooperation? Journal of Economic Perspectives, 7(2), pp. 159-171.

Frey, Bruno S., and Reto Jegen, 1999, Motivation Crowding Theory: A Survey of Empirical Evidence, working paper no. 26, Institute for Empirical Research in Economics, University of Zurich.

Frey, Bruno S., and Felix Oberholzer-Gee, 1997, The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out, American Economic Review, 87(4), pp. 746-755.

Fukuyama, Francis, 1995, Trust: The Social Virtues and the Creation of Prosperity, (Free Press, New York).

Gambetta, Diego, 1988, Can We Trust in Trust? in D. Gambetta (ed.) Trust: Making and Breaking Cooperative Relations, (Blackwell, New York), pp. 213-237.

Gibbons, Robert, 1998, Incentives in Organizations, Journal of Economic Perspectives, 12(4), pp. 115-132.

Green, E., and R. Porter, 1984, Noncooperative collusion under imperfect price information, Econometrica, 52, pp. 87-100.

Grossman, Sanford J., and Oliver D. Hart, 1983, An Analysis of the Principal-Agent Problem, Econometrica, 51, pp. 7-46.

Grossman, Sanford J., and Oliver D. Hart, 1986, The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration, Journal of Political Economy, 94(4), pp. 691-719.

Güth, Werner, and Hartmut Kliemt, 1994, Competition or Co-operation: On the Evolutionary Economics of Trust, Exploitation and Moral Attitudes, Metroeconomica, 45(2), pp. 155-187.

Hart, Oliver D., 1983, Optimal Labor Contracts Under Asymmetric Information: An Introduction, Review of Economic Studies, 50, pp. 3-35.

Hart, Oliver D., and Bengt Holmstrom, 1987, The Theory of Contracts, in Truman F. Bewley (ed.) Advances in Economic Theory: 5th World Congress, (Cambridge University Press, New York), pp. 71-155.

Hart, Oliver D., and John Moore, 1988, Incomplete Contracts and Renegotiation, Econometrica, 56, pp. 755-785.

Hausman, Daniel M., and Michael S. McPherson, 1996, Economic Analysis and Moral Philosophy, (Cambridge University Press, New York).

Hodgson, Geoffrey M., 1996, Corporate Culture and the Nature of the Firm, in John Groenewegen (ed.), Transaction Cost Economics and Beyond, (Kluwer Academic Press, Boston, MA).

Holmstrom, Bengt, 1979, Moral Hazard and Observability, Bell Journal of Economics, 10, pp. 74-91.

Huck, Steffen, 1998, Trust, Treason, and Trials: An Example of How the Evolution of Preferences Can Be Driven by Legal Institutions, Journal of Law, Economics, and Organization, 14(1), pp. 44-60.

Huang, Peter H., and Ho-Mou Wu, 1994, More Order without More Law: A Theory of Social Norms and Organizational Cultures, Journal of Law, Economics, and Organization, 10(2), pp. 390-406.

James, Harvey S., Jr., 1998, Using the Prisoner's Dilemma to Teach Business Ethics When Personal and Group Interests Conflict, Teaching Business Ethics, 2, pp. 211-222.

James, Harvey S., Jr., 2000, Separating Contract From Governance, Managerial and Decision Economics, 21, pp. 47-61.

Kandel, Eugene, and Edward P. Lazear, 1992, Peer Pressure and Partnerships, Journal of Political Economy, 100(4), pp. 801-817.

Kandori, Michihiro, 1992, Social Norms and Community Enforcement, Review of Economic Studies, 59, pp. 63-80.

Klein, Benjamin, Robert G. Crawford, and Armen A. Alchian, 1978, Vertical Integration, Appropriable Rents and the Competitive Contracting Process, Journal of Law and Economics, 21(2), pp. 297-326.

Klein, Benjamin, and Keith B. Leffler, 1981, The Role of Market Forces in Assuring Contractual Performance, Journal of Political Economy, 89, pp. 615-641.

Kreps, David M., 1990, Corporate Culture and Economic Theory, in J.E. Alt and K.A. Shepsle (eds.) Perspectives on Positive Political Economy, (Cambridge University Press, New York), pp. 90-143.

Kreps, David M., Paul Milgrom, John Roberts, and Robert Wilson, 1982, Rational Cooperation in the Finitely Repeated Prisoner's Dilemma, Journal of Economic Theory, 27, pp. 245-252.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andreai Shleifer, and Robert W. Vishny, 1997, Trust in Large Organizations, American Economic Review Papers and Proceedings, 87(2), pp. 333-338.

Lahno, Bernd Lahno, 1995, Trust and Strategic Rationality, Rationality and Society, 7(4), pp. 442-464.

Lazear, Edward P., 1995, Corporate Culture and the Diffusion of Values, in Horst Siebert (ed.), Trends in Business Organization: Do Participation and Cooperation Increase Competitiveness?: (J.C.B. Mohr, Tübingen, Germany).

Levinthal, Daniel, 1988, A Survey of Agency Models of Organization, Journal of Economic Behavior and Organization, 9(2), pp. 153-186.

Lorenz, Edward, 1999, Trust, Contract and Economic Cooperation, Cambridge Journal of Economics, 23(3), pp. 301-315.

Lyons, Bruce R., and Judith Mehta, 1997, Contracts, Opportunism and Trust: Self-Interest and Social Orientation, Cambridge Journal of Economics, 21(2), pp. 239-257.

Marwell, Gerald, and Ruth Ames, 1981, Economists Free Ride, Does Anyone Else? Journal of Public Economics, 15, pp. 295-310.

Miller, Gary, 1992, Managerial Dilemmas: The Political Economy of Hierarchy, (Cambridge University Press, New York).

Morduch, Jonathan, 1999, The Microfinance Problem, Journal of Economic Literature, 37(December), pp. 1569-1614.

Neilson, William S., 1999, The Economics of Favors, Journal of Economic Behavior and Organization, 39, pp. 387-397.

Parks, Craig D., and Lorne G. Hulbert, 1995, High and Low Trusters' Responses to Fear in a Payoff Matrix, Journal of Conflict Resolution, 39(4), pp. 718-731.

Perelman, Michael, 1998, The Neglected Economics of Trust: The Bentham Paradox and Its Implications, American Journal of Economics and Sociology, 57(4), pp. 381-389.

Predergast, Candice, 1999, The Provision of Incentives in Firms, Journal of Economic Literature, 37(March), pp. 7-63.

Ridley, Matt, 1997, The Origins of Virtue: Human Instincts and the Evolution of Cooperation, (Viking Penguin, New York).

Ross, Steven, 1973, The Economic Theory of Agency: The Principal's Problem, American Economic Review, 63(2), pp. 134-139.

Sen, Amartya, 1977, Rational Fools: A Critique of the Behavioural Foundations of Economic Theory, Philosophy and Public Affairs, 6, pp. 317-344.

Sen, Amartya, 1988, On Ethics & Economics, (Blackwell Publishers, Malden, MA).

Shapiro, Susan P., 1987, The Social Control of Impersonal Trust, American Journal of Sociology, 93(3), pp. 623-658.

Spagnolo, Giancarlo, 1999, Social Relations and Cooperation in Organizations, Journal of Economic Behavior and Organization, 38, pp. 1-25.

Spence, A.M., 1973, Job Market Signaling, Quarterly Journal of Economics, 87, pp. 355-374.

Sugden, Robert, 1989, Spontaneous Order, Journal of Economic Perspectives, 3(4), pp. 85-97.

Telser, Lester G., 1972, Competition, Collusion, and Game Theory, (Aldine-Atherton, Chicago, IL).

Telser, Lester G., 1980, A Theory of Self-enforcing Agreements, Journal of Business, 53, pp. 27-44.

Tullock, Gordon, 1999, Non-prisoner's Dilemma, Journal of Economic Behavior and Organization, 39, pp. 455-458.

Valley, Kathleen L., Joseph Moag, and Max H. Bazerman, 1998, 'A Matter of Trust': Effects of Communication on the Efficiency and Distribution of Outcomes, Journal of Economic Behavior and Organization, 34, pp. 211-238.

Williamson, Oliver E., 1983, Credible Commitments: Using Hostages to Support Exchange, American Economic Review, 73, pp. 519-540.

Williamson, Oliver E., 1985, The Economic Institutions of Capitalism, (The Free Press, New York).

Williamson, Oliver E., 1993, Calculativeness, Trust, and Economic Organization, Journal of Law and Economics, 36(April), pp. 453-486.

Zucker, Lynne G., 1986, Production of Trust: Institutional Sources of Economic Structure, 1840-1920, Research in Organizational Behavior, 8, pp. 53-111.