

Trust Building Among Strangers

Teck-Hua Ho

Haas School of Business, University of California, Berkeley, California 94720, hoteck@haas.berkeley.edu

Keith Weigelt

The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104, weigelt@wharton.upenn.edu

The trust-building process is basic to social science. We investigate it in a laboratory setting using a novel multistage trust game where social gains are achieved if players trust each other in each stage. In each stage, also, players have an opportunity to appropriate these gains or be trustworthy by sharing them. Players are strangers because they do not know the identity of others and they will not play them again. Thus, there is no prospect of future interaction to induce trusting behavior, and we study the trust-building process where there is little scope for social relations and networks.

Standard game theory, which assumes all players are opportunistic and untrustworthy and thus should have zero trust for others, is used to construct a null hypothesis. We test whether people are trusting or trustworthy and examine how inferring the intentions of those who trust affects trustworthiness. We also investigate the effect of stake on trust, and study the evolution of trust.

Results show subjects exhibit some degree of trusting behavior, although a majority of them are not trustworthy and claim the entire social gain. Players are more reluctant to trust in later stages than in earlier ones and are more trustworthy if they are certain of the trustee's intention. Surprisingly, subjects are more trusting and trustworthy when the stake size increases. Finally, we find the subpopulation that invests in initiating the trust-building process modifies its trusting behavior based on the relative fitness of trust.

Key words: trust building; backward induction; behavioral game theory

History: Accepted by Detlof von Winterfeldt, decision analysis; received January 9, 2004. This paper was with the authors 4 months for 1 revision.

1. Introduction

Graham Greene aptly captured the centrality of trust to social relations when he wrote, "It is impossible to go through life without trust: that is to be imprisoned in the worst cell of all, oneself" (1943). Trust is one of a handful of concepts to transcend territorial barriers of social disciplines; sociologists, anthropologists, economists, and psychologists study trust. The social sciences view trust as an emergent process fundamental in social interaction and market efficiency (Creed and Miles 1994). Without trust, opportunism rules and social exchange is costly. Despite the costs of trusting when others are untrustworthy, establishing trust can greatly benefit those involved (Axelrod 1984). Some psychologists claim trust is the hallmark of social adjustment (Gurtman 1992) and that without it neuroses prevail. Sociologists generally believe that social relations and obligation are more important than morality, contracts, or institutions in producing trust (Granovetter 1985, Shapiro 1987). Most economists view trust rationally and posit people only trust when it pays them to do so (Camerer and Weigelt 1988, Berg et al. 1995).

Arguably, trust delivers its greatest social value as a linchpin to low-cost cooperative behavior. It holds

immense strategic value for a group or organization of any size. Monitoring and incentives can motivate cooperative behavior, but at a higher cost. Organizational researchers theorize and empirically show that trust reduces managerial monitoring, decreases coordination costs, and increases individual effort (Ring and Van de Ven 1992, McAllister 1995, Chiles and McMackin 1996). Trust often results in social gains, prompting some to suggest it leads to prosperity in society (Fukuyama 1995).

Similar to religion, trust requires a leap of faith. As Bradach and Eccles (1989) note, when there is faith where opportunism might be rationally expected, trust prevails. Some social biologists believe humans are genetically programmed not to trust or to be trustworthy except with regard to kin (Wright 1995). This is echoed by a dominant assumption in game theoretic equilibrium analysis of self-interested and not-trustworthy individuals (Camerer 2003). Clearly, if some people trust and others are not trustworthy, the latter exact a heavy toll from the former. Consequently, the existence of trusting and trustworthy behaviors is problematic.

Though problematic, we observe trusting and trustworthy behavior in others: a colleague who borrows a book and returns it, or a pedestrian who crosses the

street in the face of oncoming traffic because of a stop sign. Neoclassical economists use institutional mechanisms (Alchian and Demsetz 1973) to help explain trusting and trustworthy behavior by self-interested people within the classical framework of utility maximization. They rationalize that mechanisms such as explicit contracts, incentive schemes, and hierarchical authority can induce individual behavior consistent with trust and trustworthiness. In this world, trust is used not to explain behavior, but to label it (Craswell 1993).¹ To trust or not to trust is purely calculative; institutional mechanisms help individuals act as if they do trust, usually because these mechanisms increase the costs of opportunistic behavior. As Granovetter (1985) aptly notes, such mechanisms do not produce trust but are functional substitutes for it.

Granovetter (1985) posits, however, that concrete personal relations are an important determinant of trusting behavior. This so-called embeddedness argument emphasizes the role of social networks and reputation arising from interpersonal contacts. The higher the embeddedness of the social relation, the more likely (on average) trust is extended and returned. Granovetter suggests embeddedness is a necessary (though not sufficient) condition for generating trusting behavior. General population-level reputation is significantly less important than individualized reputations generated by direct prior contacts, because direct contacts provide more reliable information of others' intentions. Granovetter (1985) states (p. 506) that the inclusion of such relations in the trust decision does not call for the abandonment of a rational decision-making process. Instead, the degree of embeddedness should be factored into such models—for example, as with a matching protocol.

In this paper, we create an experimental setting where there is minimal scope for social embeddedness. In our experiments, subjects interact anonymously and only play each other once. Contracting is not possible, and in fact subjects never find out with whom they play the trust game.² The novel game has additional desirable features. It allows us to untangle trusting and trustworthy behavior; it has multiple stages, which allow subjects to detect the intention of others; and mutual trusting behavior in the game leads to sizable social gains. Delineating trusting from trustworthy behavior is important because they are fundamentally different; most previous research on the prisoners' dilemma (PD) entangles the two.

¹ Malhotra and Murnighan (2000) report an interesting study to show that contracts make development of trust difficult or block preexisting trust.

² We use student subjects. It is possible that some students know each other prior to the experiment. Thus, they are not strictly strangers. We define strangers here to mean people who engage in anonymous and one-shot interactions.

The ability to detect intention of others allows us to study an important construct: pseudotrust, which has not been investigated much in prior studies. Pseudotrust occurs when one trusts and gives up property rights in early stages of the game in order to regain larger ones in the later stages. Pseudotrust is differentiated from the trusting behavior in later stages when there is no chance to regain property rights.

This paper aims to answer five basic research questions. First, can anonymous interactions (such as exchanges between participants of eBay auctions) generate trust and trustworthy behavior? Our experimental design allows us to quantify the extent to which these behaviors exist in our population. Second, how does the potential of future social gains influence current trusting behavior? This potential varies across the stages of our trust game. Third, does better knowledge of others' trusting behavior make people more trustworthy? Within our game, the uncertainty of intent of trusting players is high in early stages and low in later stages. Fourth, is trusting behavior dependent on stake size? In three sessions, we increase the stake size tenfold to see whether this increase affects trusting behavior. The dramatic increase in stake size allows us to remove any concerns on whether subjects are sufficiently motivated and to test whether stake size increases the extent of rational behavior. Fifth, do evolutionary paths of trust building differ across subpopulations? Specifically, do the trust dynamics of the subpopulation that invests and initiates the trust-building process differ from the one with the first decision rights to terminate the process? We answer the above five questions by collecting data in three countries: the United States, Singapore, and China. This cross-cultural data collection effort helps to check for robustness of the basic results and helps to increase external validity.

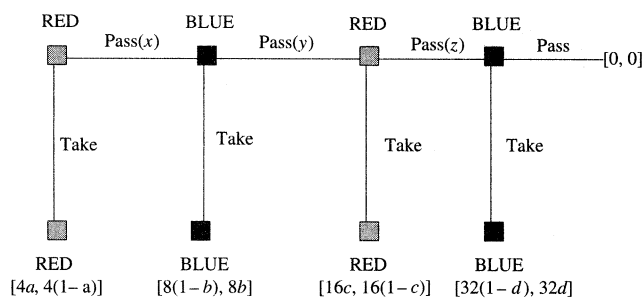
The paper is organized as follows. In §2, we describe our trust-building game. Section 3 develops hypotheses using theories of trusting and trustworthy behavior from sociology, economics, and psychology. Experimental method is presented in §4, and results are presented in §5. Section 6 discusses the methodological issues, explores theoretical implications of results, and suggests future research directions.

2. A Trust-Building Game

Our trust-building game is shown in Figure 1.³ The game consists of four decision stages, and is played between two players, one designated as RED,

³ Although our trust-building game looks like the centipede game (see Rosenthal 1982, McKelvey and Palfrey 1992), there is an important difference. In our game, the player can choose to be either trustworthy or not, whereas in the centipede game players are forced to be not trustworthy.

Figure 1 The Trust-Building Game



the other as BLUE. Akin to the real world, individual property rights are voluntarily given up in exchange for potential social gain.

RED moves first. She can either trust BLUE and pass, or choose not to trust and select down. If down is chosen, RED exercises her property rights and unilaterally decides how to split the payoff of 4 (rewarding herself with $4a$ ($0 \leq a \leq 1$), whereas BLUE receives $4(1 - a)$), and the game is over. If RED trusts, she passes, thereby giving up her property rights in hopes of sharing future social gains. Social gains occur because the payoff doubles, although BLUE now owns the property rights. That is, BLUE must now make a similar decision of whether to trust. If BLUE trusts in Decision Stage 2, he passes. If BLUE does not trust, he takes and splits the payoff so he receives $8b$ ($0 \leq b \leq 1$) and RED receives a payoff of $8(1 - b)$. Similar to Glaeser et al. (2000), we measure BLUE's level of trustworthiness by the amount returned to RED. The higher the returned amount, the more trustworthy is BLUE. The game is over if BLUE does not trust. If BLUE trusts and passes, we again double social gain to 16 in Stage 3.

Then RED again must decide whether to trust. If RED does not trust, she takes, and must decide whether to be trustworthy. Again, RED's level of trustworthiness is measured by the amount she returns to BLUE. If RED trusts, she passes to BLUE and the payoff doubles to 32. In Stage 4, BLUE can only choose his level of trustworthiness.

In sum, we measure the degree of trusting behavior by the proportion of passes in Stages 1–3. We denote these proportions by x , y , and z , respectively. The degree of trustworthiness at each stage is represented by the proportions a , b , c , and d , respectively.⁴

⁴ We implicitly assume the measure of trustworthiness is stage independent. That is, $b = 0.6$ and $c = 0.6$ represent equivalent trustworthiness behaviors in Stages 2 and 3. An alternative view could allow trustworthiness measures to be stage dependent. For example, RED in Stage 3 may feel she is partly responsible for the larger payoff of 16 by passing to BLUE in Stage 1, and therefore is entitled to a larger proportion of the payoff. Hence, the level of trustworthiness in Stages 2 and 3 could be equivalent even if $c > b$. This latter view suggests that players will take more for themselves in later stages

The game has several desirable features for examining the trust-building process. Unlike some previous trust games (e.g., Yamagishi et al. 1998), our game allows an individual to exhibit both trusting and trustworthy behavior. Subject decisions in each stage consist of two components: strategy choice (to trust or not)⁵ and payoff choice (to be trustworthy or not). Strategy choice determines whether a subject trusts (passes or takes, respectively). Payoff choice determines whether the subject is trustworthy.⁶ Hence, one can be trusting without being trustworthy.

Second, our trust game allows subjects to reveal their trustworthiness. Previous games force subjects to choose between two payoff choices, which might not coincide with their most preferred choice (for example, in McKelvey and Palfrey 1992, subjects are forced to claim 80% of the property rights. The same thing can be said about most studies of repeated PD), so subjects may choose the lesser of two evils. Because subjects in our game can divide the social gain in any way, they choose their most preferred level of trustworthiness.

With the multistage design, we can study how future social gains might influence the trust-building process. For example, if higher future social gains induce more trusting behavior, we will see more passes in earlier decision stages. The design also allows us to test process-based trust, a widely identified trust process in sociological work (Zucker 1986), while controlling for institutional and person-based trust.

Our game shares similarities with other trust games—it is a two-person game, and trusting results in social gains. It also differs from other games in several important ways. For example, our trust game is different from a PD game with an option to exit (PDO). In our game, trusting at each stage means

for the same degree of trustworthiness. Thus, our proportions c and d provide a lower bound for the degree of trustworthiness in Stage 3 and 4. We find no evidence to support a stage-dependent view because players do not seem sensitive to history within a game round (to be discussed in §5).

⁵ Mayer et al. (1995) suggest one can cooperate but not trust, because the use of external mechanisms can compel others to cooperate. For example, a team leader can punish those who will not cooperate. Members may not trust each other, but they prefer cooperation to the punishment for noncooperation. No external mechanisms or authorities enforce punishment in our design. To trust or not is solely an individual's decision.

⁶ Strictly speaking, RED's payoff choice in the first stage is not an indication of her trustworthy behavior because she is endowed with the payoff, and it is not a consequence of social gain from trust. Also, BLUE's strategy choice in Stage 4 is not an indication of trust because property rights are last exchanged in Stage 3. We discuss this further in §3.2. Because RED knows BLUE will not give back control of property rights, RED is clearly hoping to get a fair split of the social gains. She must believe that BLUE is trustworthy.

giving up the property rights—the truster has no control over the payoff realization if the trustee chooses to end the game at the next stage. In the PDO game, choosing to play does not mean the trustee has no influence over the expected payoffs. A player who chooses to play the game or trust can still guarantee herself a nonnegative payoff by choosing defection (as in Orbell and Dawes 1991, 1993). Also, our game gives those who are trustworthy complete freedom to divide the property rights as he wishes. The outcomes of the PDO game are determined by the experimenter, not by the subjects.⁷

Similarly, there are other trust games closely related to ours (e.g., Berg et al. 1995, Glaeser et al. 2000). These dictator-type trust games also give trustees complete control in division of social welfare. However, they are single staged. The trustee (the dictator) is never asked to relinquish control once it has been given to her. Consequently, these dictator-type trust games can only study initial stages of the trust-building process.

3. Hypotheses

3.1. Rationality and Social Uncertainty

A standard assumption in game-theoretic equilibrium analysis is that individuals are rational (they are self-interested and monetary maximizing). Consequently, they cannot be entrusted with property rights if there is no chance for penalizing untrustworthy behavior (either through formal institutional mechanisms or informal sanctions of social relations). This game-theoretic reasoning is as follows: If Stage 4 is reached in our game, a rational BLUE player will keep the property entrusted to him for himself and take 32 (i.e., $d = 1$). A rational RED player, knowing this, will neither trust BLUE nor be trustworthy. Hence, she will take the entire payoff of 16 in Stage 3. Anticipating this action of RED in Stage 3, a rational BLUE player exhibits this behavior in Stage 2, takes 8, and leaves nothing for RED. Completing this backward induction argument, the rational RED player takes 4 in the first stage. Hence all players are neither trusting nor trustworthy.

Sociologists also believe little trust will develop where social uncertainty is high, there is little commitment to relations, and social embeddedness is insignificant. Kollock (1994) and Yamagishi et al.

(1998) indicate social uncertainty exists when (1) the trustee has an incentive to impose harm or cost on the truster, and (2) the truster is uncertain about whether the trustee will actually impose this harm or cost. Our design imposes high levels of social uncertainty on subjects by decreasing the probability of committed relations. Committed relations reduce social uncertainty in two ways: (1) committed partners accumulate information about each other (i.e., mutual monitoring) and (2) reciprocity is possible (i.e., mutual hostages or tit-for-tat) (Yamagishi and Yamagishi 1994). Our experimental design makes any degree of commitment in a relation almost impossible. In fact, it provides an estimate of the degree of trusting behavior in a population of strangers where social embeddedness is insignificant. Subjects play against each other at most once (i.e., they are strangers), each subject only assumes the role of either a RED or BLUE player, RED and BLUE players are in different rooms, and players are not told with whom they are matched. Consequently, we expect subjects to have a high level of social uncertainty. In our experiments, there is neither a way to reveal the identity nor a disincentive to penalize untrustworthy individuals. The act of being untrustworthy is known only to the untrustworthy subject and the one whose trust was betrayed. In addition, the two will never meet in the future, so there are no formal or informal mechanisms available to punish an untrustworthy player. The lack of monitoring and penalty increases the probability of opportunism. Consequently, we expect less trustworthy behavior.

The standard theories in economics and sociology lead to the predictions that people will exhibit neither trusting nor trustworthy behavior. Consequently, we have the following hypotheses.

HYPOTHESIS 1A. *Proportions $b = c = d = 1$. That is, individuals are not trustworthy (i.e., they are opportunistic).*

HYPOTHESIS 1B. *Proportions $x = y = z = 0$. That is, individuals exhibit zero trust; they do not give up property rights.*

Because there is increasing evidence to suggest that humans are boundedly rational (Camerer et al. 2004) and care about fairness (see Camerer 2003 for an excellent review), the above hypotheses serve as our baseline hypotheses. They allow us to compare our results with those reported in the literature and quantify the degree of deviation from the standard rational hypotheses.

3.2. Informed Reciprocity

There is a subtle difference between trusting behaviors in the first two stages and the third stage. A BLUE player who receives the property rights from a RED

⁷ This selective play paradigm literature shows that social welfare increases when subjects have a freedom to choose between playing or not playing a particular PD game. These increases occur because intending cooperators are more likely to enter such games relative to intending defectors, resulting in a better chance of cooperative outcomes. Consequently, the welfare of intending cooperators is higher than that of intending defectors (Orbell and Dawes 1991, 1993; Hayashi and Yamagishi 1998).

player in Stage 3 can infer unambiguously the intention of RED. By passing, RED clearly indicates her belief of BLUE's trustworthiness, because the only decision facing BLUE is one of trustworthiness. The same thing cannot be said for trusting behavior in Stage 1 or 2. We cannot unambiguously infer players who relinquish their property rights in these earlier stages believe others to be trustworthy. Players who give up their property rights in Stages 1 and 2 have the opportunity to regain future larger property rights. Consequently, there is greater social uncertainty in the earlier stages for the trustee. Do subjects give up property rights because they believe others are trustworthy, or for cultivating pseudotrust? Pseudotrust occurs when one trusts to increase property rights for the possibility of claiming them in the later stages. Because players are better informed about intentions of trusting subjects in Stage 3 relative to Stages 1 and 2, we expect a greater level of trustworthiness in Stage 4. Because BLUE can unambiguously detect and interpret the intention of RED in Stage 3, he can selectively be more trustworthy to those who trust not for own, but for social gain. This informed reciprocity is characterized as *translucent* (Orbell and Dawes 1991), which is the degree to which one can recognize another's intention. Consequently, individuals who are more certain of the intentions of others' trusting behaviors are more trustworthy.

HYPOTHESIS 2. *Proportion $d < (b + c)/2$.*

3.3. The Potential Value of Trust

In each stage of the trust-building game, those who trust must give up property rights in exchange for potential social gain. We define trust's potential value as the difference between the expected future return from trusting and the property rights currently held. The expected future return at any stage is computed given the percentages of passing and taking and the mean percentage of property rights claimed in future stages. Necessarily, the expected return for earlier stages is higher than for later ones. Hence, we expect to see less-trusting behavior in later stages of the game than in earlier ones. For example, a RED player in Stage 1 can take and receive a payoff of 4, but is losing the opportunity to share a potential future payoff of 32. In Stage 3, this same player can take and realize a gain of 16, versus the opportunity to share a future payoff of 32. Consequently, individuals are less trusting as the potential value of trust decreases.

HYPOTHESIS 3. *Proportion $x > y > z$.*

3.4. High-Stake Trust Building

Although relatively little research shows how the magnitude of payoffs affects trusting behavior—despite stories that of soldiers who sacrifice themselves to save comrades are as old as war itself—the

effect of financial incentives on experimental subject behavior has attracted substantial attention among social scientists. For some reviews, see Smith and Walker (1993), Hertwig and Ortmann (2001), Jenkins et al. (1998), and Camerer and Hogarth (1999). These authors examine hundreds of experiments, but no definitive answer arises from them, though there is consensus on general trends. The majority of reviews finds increased financial incentives have no significant impact on mean subject behavior, though they do reduce the variance across choices (Smith and Walker 1993, Camerer and Hogarth 1999). Hertwig and Ortmann (2001) find incentives slightly affect behavior (though most often when increased from zero). Jenkins et al. (1998) find the strongest link between incentives and performance, though most of the studies they review are one-person output tasks (e.g., coding) that measure performance on a purely quantitative scale.

Another consensual finding is that incentives matter more in nonstrategic than in strategic tasks. A strategic task is one where a subject's payoff depends on her actions and those of others. Nonstrategic ones examine one-person decisions such as recalling items, predicting future states, or output per time period. For example, Camerer and Hogarth (1999) report the following. They categorize 58 studies with regards to whether an increase in financial incentives significantly changed mean performance. In 31 (53.4%) studies, they did. Of these 31 studies, 7 (22.5%) were strategic experiments and 24 were nonstrategic. In contrast, of the 27 studies where mean performance did not change, 21 (74.1%) were strategic. Furthermore, if we categorize the 34 strategic studies into bargaining trust-building or nonbargaining (e.g., signaling, auctions, tournaments) games, all 9 bargaining games fall into the category of increased incentives having no effect on mean performance. Camerer and Hogarth (1999) state, "results in Table 1 [summarizing the above results we discuss] suggest little reason to think the effects of very large incentives will be substantial" (p. 10). They conclude this finding is applicable to strategic decisions, especially in bargaining and auctions. Consequently, we hypothesize the level of trust is not dependent on stake size.

HYPOTHESIS 4. *The level of trust does not depend on stake size. Proportions x , y , and z are not functions of stake size.*

3.5. Population Learning and the Emergence of Trust

The logic behind Hypothesis 1 is grounded in full rationality of players. Recent research in economics increasingly suggests players are boundedly rational with adaptive expectations (Ho and Weigelt 1996, Camerer and Ho 1999, Camerer et al. 2002).

These models assume subjects' actions are path dependent and self-correcting. Subjects stick to strategies that worked well in the past, or experiment with previously unchosen strategies that are likely to improve performance. Alternatively, one can view trust building in an evolutionary perspective. For instance, Hoffman et al. (1998) suggest that their subjects exhibited reciprocity even in single-shot games (counter to rational hypothesis) because humans have evolved mental algorithms for identifying and punishing cheaters in social exchange. In evolutionary dynamics terms, subjects adjust their behavior by mimicking the person that is the most successful, or fittest. Thus, future trusting behavior depends on its current success or relative fitness. The higher the number of trusting and trustworthy individuals in the population, the more likely trusting and trustworthy behavior increases both the payoffs of the collective and individuals. We expect trusting behavior in the population increases as its expected value (or fitness) increases. Although our focus is on the evolution dynamics of group behavior, other studies suggest group dynamics can be derived from individual learning models in which subjects modify their propensity of choosing a strategy based on its performance (see, for example, Friedman 1991).

HYPOTHESIS 5. The emergence of trust in a population is governed by evolution dynamics. There is more trust in the population if trust has a higher relative fitness.

4. Experimental Method

A total of 386 subjects participated in the experiment. Subjects were recruited from undergraduate courses at National University of Singapore in Singapore ($n = 106$), University of Pennsylvania in the United States ($n = 100$), and from Beijing University in the People's Republic of China (China) ($n = 180$).

A typical experimental session consisted of 20 subjects who met in a room at a specified time. As they entered the room, they chose a bingo ball from a cage containing 20 balls numbered 1 to 20 (a ball chosen was not replaced).⁸ The number on the ball represented his subject number. Then the instructions (for details see Ho and Weigelt 2004) were publicly read. After questions, BLUE subjects (those with odd subject numbers) were asked to go to another room. RED subjects remained in the room. Subjects in each room were positioned so no subject could see the worksheet of another.

Our multistage trust-building game consisted of four stages. RED subjects made choices in Stages 1 and 3 and BLUE subjects in Stages 2 and 4. Although

the game could end in any stage (when subjects chose Take), subjects recorded their choices in both stages. We asked subjects to record their contingent strategy for two reasons. First, we reduced the data collection time and hence the potential for boredom. Second, contingent strategies help us understand subject behavior in all stages, even though the actual game might end earlier.⁹

Each session consisted of 10 rounds and lasted for about one hour. In each round, a BLUE subject was paired with a RED subject (this was her pair member). Subject choices were privately made and recorded by an administrator in the room. After all subjects in both rooms made their decisions, the two administrators met outside the rooms and recorded all choices. They then went back to their respective rooms, privately informed each subject of the choices of the pair member and the generated payoff. The administrator then announced the end of that round and the beginning of the next round. Each BLUE subject was paired with a RED subject only once, so each subject had a new partner in each round. This procedure of matching subjects rules out any possibility of a shadow of the future across rounds.¹⁰ A RED player clearly cannot influence the action of another RED player because the RED players never interact. In addition a RED player cannot influence a BLUE player because she only plays with BLUE once and BLUE players never play against each other. A similar line of reasoning applies to BLUE players. This matching method, however, does not rule out a shadow of the future across stages within a game round.¹¹ After 10 rounds, subjects add up their points from each round, and multiply this by a conversion rate that transforms the points into a monetary payoff. The subjects were explicitly told they would be paid in cash after the experiment, so they understood earnings were a function of decision choices.

Subjects seemed to understand the instructions and the payoff associated with each of their actions. This

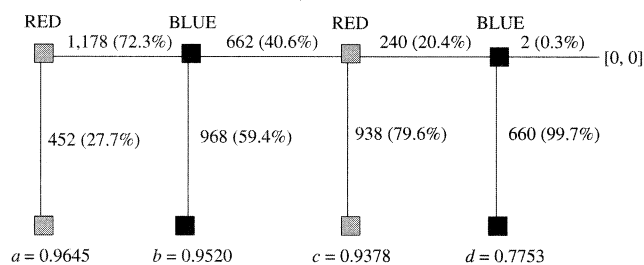
⁹ Consequently, in Figure 2, the total choice count in Stage 3 is 1,178, and in Stage 4, 662.

¹⁰ This matching protocol rules out even the triadic effects. That is, player A may want to be nice to player B because player B will be playing player C with whom player A will play in the future. Clearly, this is not possible in our design because players of the same color never play each other.

¹¹ We believe our simultaneous social exchange protocol deters subjects from perceiving a shadow of the future within a game round because subjects must submit a full contingent strategy without knowing the strategies of others. As discussed in §5, we test this belief by running additional experiments using a sequential social exchange protocol. In the sequential protocol, subjects choose after knowing the strategies of others in prior stages within a game round. Because the results from the two exchange protocols are almost identical, we conclude that subjects did not perceive a shadow of the future within a game round.

⁸ Two sessions in Singapore used 52 and 54 subjects and required a larger number of bingo balls.

Figure 2 Descriptive Statistics of Players' Choices



Note. Pooled across 13 experimental sessions and 10 rounds.

was evidenced in two ways. First, after the experiment and during the debriefing, subjects' comments suggested that they did not have trouble writing down contingent strategies (e.g., "Many of my choices in Stage 3 or 4 did not matter") and the matching protocol (e.g., "I could not build relationship with a person because we did not match with the same person more than once"). Second, almost 100% of the choices in Stage 4 were Take (see Figure 2), which is a dominant choice. This result indirectly suggests subjects did understand the game.

We conducted over 16 experimental sessions, 5 in the United States, 9 in China, and 2 large sessions in Singapore, to check the robustness of our findings. Six China sessions were low stake, and 3 were high stake.¹² High-stake sessions had a tenfold increase in the conversion rate. Because we collected data in different countries, we took the following precautionary controls: Instructions in a country were written in its primary language—English in Singapore and the United States and Chinese in China. We used the reverse translation method to ensure Chinese instructions were equivalent to those in English. That is, an independent translator translated the instructions from English into Chinese and another independent translator from Chinese back into English. Then, one of the authors who is fluent in both Chinese and English checked both translations. The conversion rate was chosen to normalize average payoff so that subjects were paid about twice the minimal wage for the country. On average, U.S. subjects earned US\$11, Singapore subjects earned about S\$10, and Chinese subjects about 24 and 320 yuan Renminbi in the low-stake and high-stake games, respectively. This latter earning is approximately equivalent to two months' wages for a blue-collar worker.

5. Results

Altogether 386 subjects participated in the experiment (193 RED and 193 BLUE subjects). Each subject made choices over 10 rounds. Figure 2 summarizes their choices in each of the four decision stages. As you can

see, trusting behavior monotonically declines across decision stages while the degree of trustworthiness monotonically increases. Recall we examine trusting behavior in Stages 1–3 and trustworthiness in Stages 2–4. In each stage, there is a substantial decline in the proportion of trusting choices. The degree of trustworthiness increases marginally for Stages 2–3, however, with a substantial increase in Stage 4.

Seventy-two percent of the RED choices initiate trust building in Stage 1. Correspondingly, 41% of the BLUE's choices and 20% of the RED's choices are of the trusting nature in Stage 2 and Stage 3, respectively. In Stages 2 and 3, those who did not trust were also not trustworthy; they kept 95% and 94% of the property rights. Interestingly, in Stage 4 only 78% of the property rights were claimed.¹³ We formally test our hypotheses below.

We test whether subjects are trustworthy by examining the percentages of property rights claimed by BLUE players in Stage 2. We focus on the trustworthiness of BLUE subjects who choose not to trust and immediately terminate the trust-building process. A more trustworthy BLUE should return a higher entrusted property rights to RED. A fully opportunistic BLUE takes 100% of the property rights. Table 1 reports the average proportion of property rights claimed along with the 95% confidence interval. It clearly shows BLUE subjects are largely opportunistic in Stage 2. The average proportions claimed exceed 95% in all three countries. Overall, these results support Hypothesis 1a.

In spite of this lack of trustworthy behavior, we find some support for trusting behavior in Table 1. We test Hypothesis 1b by the percentage of RED subjects who initiate the trust-building process by choosing Pass in Stage 1. About 70% of RED subjects trust by giving up their property rights in Stage 1. For example, 73% of the Singaporean RED subjects chose Pass in the first stage. Clearly, a significant percentage of subjects initiate the trust-building process. Thus, we find no support for Hypothesis 1b.

Hypothesis 2 examines the effect of behavioral translucency on trustworthy behavior. It states trustworthiness is more likely when the intention of the trustee is less ambiguous. We test this hypothesis by comparing trustworthy behaviors in Stage 4 relative to earlier stages. A BLUE subject in Stage 4 is certain about the trusting intention of RED because there is no opportunity for RED to regain property rights. The same cannot be said for earlier stages. For example, a BLUE subject in Stage 2 cannot be sure why a RED subject passed her property rights in Stage 1. She may

¹² Our research budget did not allow us to conduct high-stake sessions in Singapore and the United States.

¹³ Obviously, these percentages are generated only by those who trusted in Stage 2.

Table 1 Percentages of Subjects Showing Trust in Stage 1 and Trustworthy Behavior in Stage 2

	Country		
	China	Singapore	United States
Mean proportion of trust in Stage 1	(73.3%)	(73.0%)	(69.8%)
95% confidence interval	(71.7%, 74.9%)	(71.3%, 74.7%)	(68.0%, 71.7%)
Mean proportion of trustworthiness in Stage 2	(96.04%)	(96.44%)	(95.06%)
95% confidence interval	(93.47%, 96.67%)	(95.16%, 97.71%)	(93.47%, 96.66%)

pass selfishly to increase property rights for the possibility of claiming them in Stage 3, or she may pass for the collective social gain. In Stage 4, there is no such uncertainty.

Table 2a shows the average proportions claimed are 95.85%, 93.96%, and 77.4% in Stages 2, 3, and 4, respectively. Across the three countries, the 95% confidence intervals for Stage 4 do not overlap with those in Stages 2 or 3. In fact, in the first five rounds, the modal proportion claimed in Stage 4 is 50% (131 out of the 354 subjects who take in Stage 4).¹⁴ Clearly, subjects are more trustworthy in Stage 4. Thus, we find some support for Hypothesis 2.

Hypothesis 3 states trust increases with the difference in the expected return from trusting and the size of property rights currently held. Because this difference decreases over stages, we expect less-trusting behavior in later stages. Table 2b clearly supports this hypothesis. There is a significant decrease in trusting behavior from Stage 1 to Stage 2, and from Stage 2 to Stage 3. This is true across all countries. The average percentages of subjects trusting are 72%, 40%, and 20% in Stages 1, 2, and 3, respectively.

Hypothesis 4 predicts that stake size will not affect trusting behavior. Figure 3 shows subject behavior in the three high-stake trust games. Comparing Figures 2 and 3, it is obvious that subjects exhibit significant more trust in the high-stake games. For example, in Stage 3 RED subjects pass 80.3% compared with 20.4% in low-stake games (a simple proportion test indicates this difference is highly significant ($p < 0.001$)). This higher level of trust in Stage 3 is also matched by a higher level of trustworthiness in Stage 4. For example, 167 out of 270 choices in the high-stake games were $d = 0.5$, whereas only 197 out of 660 choices in the low-stake games were $d = 0.5$ (i.e., divide the property rights evenly).

We examine Hypothesis 5 by testing whether changes in subjects' behaviors are consistent with evolution dynamics. For each RED player in every round, there are three possible trusting behaviors: (1) Take in Stage 1 (T), (2) Pass in Stage 1, and Take in Stage 3

(PT), or (3) Pass in Stage 1 and Pass in Stage 3 (PP). The fitness (or payoff) for T is 4. The fitness for PT is the sum of two terms: (1) the percentage of Take in Stage 2 times 8 ($1 - b$) and (2) the percentage of Pass in Stage 2 times 16. Similarly, the fitness for PP is a sum of two terms: (1) the percentage of Take in Stage 2 times 8 ($1 - b$), and (2) the percentage of Pass in Stage 2 times the percentage of Take in Stage 4 times 32 ($1 - d$). Hypothesis 5 suggests the expected use of a strategy in the next round depends on its use in the current round and increases with its current fitness relative to the fitness of other strategies. Mathematically, we model the proportion of strategy j in round t to be

$$P_j(t) = \alpha P_j(t-1) + (1-\alpha) \frac{e^{\beta EV_j(t-1)}}{\sum_K e^{\beta EV_K(t-1)}}$$

The model suggests the proportion of subjects choosing strategy j in a round is a function of the proportion of subjects choosing the same strategy in the previous round and the fitness of strategy j in the previous round relative to other strategies. $\alpha > 0$ implies subjects show a tendency to repeat their previous choices regardless of the fitness from their previously chosen strategies. This could be due to inertia or insufficient experimentation. $\beta > 0$ suggests subjects are more likely to pick a strategy j if they receive a higher fitness from the strategy relative to others. That is, trusting behavior is driven by fitness (i.e., subjects modify their behaviors in response to fitness changes if $\beta > 0$). Such behavior can be interpreted as subjects either imitating a more successful other or as merely selecting a more successful action with a higher probability regardless of whether or not it was chosen in the current round (Camerer and Ho 1999).

We test Hypothesis 5 by examining whether α and β are significantly greater than zero. We run a nonlinear regression and report the results in Table 3. (See Ho and Weigelt 2004 for a detailed example of how relative fitness is computed.) As shown, both RED and BLUE players tend to repeat their previously chosen strategies ($\alpha = 0.562$ (RED), $p < 0.0001$; $\alpha = 0.8209$ (BLUE), $p < 0.0001$). In addition, results suggest that RED players are sensitive to the relative fitness of trust ($\beta = 0.2794$, $p < 0.0001$). They

¹⁴ The number of subjects who chose a 50:50 split in Stages 1, 2, 3, and 4 across all 10 rounds were 11 (out of 453), 23 (out of 968), 38 (out of 938), and 197 (out of 660), respectively.

Table 2a A Test of Trustworthiness (Percentages of Property Rights Claimed)

Country	Decision stages			
	1	2	3	4
China				
Mean proportion	—	96.04%	91.77%	79.17%
95% confidence interval	—	(93.47%, 96.67%)	(89.29%, 93.06%)	(76.44%, 81.91%)
Singapore				
Mean proportion	—	96.44%	96.39%	79.30%
95% confidence interval	—	(95.16%, 97.71%)	(95.18%, 97.62%)	(76.07%, 82.52%)
United States				
Mean proportion	—	95.06%	93.71%	73.73%
95% confidence interval	—	(93.47%, 96.66%)	(92.10%, 95.33%)	(70.45%, 77.01%)

Table 2b A Test of Trusting Behavior (Percentages of Subjects Who Pass at Each Stage)

Country	Decision stages			
	1	2	3	4
China				
Mean percentages	73.3%	48.3%	25.8%	—
95% confidence interval	(71.7%, 74.9%)	(46.3%, 50.3%)	(24.0%, 27.6%)	—
Singapore				
Mean percentages	73.0%	34.2%	20.2%	—
95% confidence interval	(71.3%, 74.7%)	(30.9%, 37.5%)	(18.6%, 21.8%)	—
United States				
Mean percentages	69.8%	38.2%	13.8%	—
95% confidence interval	(68.0%, 71.7%)	(36.1%, 40.3%)	(12.6%, 15.0%)	—

trust if rewarded for the trust. However, we do not observe similar calculative behavior for BLUE players. Overall, these results suggest that both RED and BLUE exhibit status quo bias and only RED seems to be sensitive to payoffs and engage in imitative behavior.

The behavioral difference in RED and BLUE players is interesting. We note, on the one hand, that RED players are the initiators of the trust-building process and own the original property rights of 4. BLUE players, on the other hand, do not explicitly invest in the trust-building process. Hence, RED players may expect a higher payoff from the trust-building process than do the BLUE players. Consequently, they care more about the expected payoff than do BLUE players.

Our results suggest the two subpopulations follow a different evolutionary trust-building path. The evolutionary path of RED is governed by the relative fitness of trust, whereas that of BLUE is not. We conjecture that this is due to RED being endowed with an initial payoff of 4 before the trust-building process begins; this is, hence, a natural reference point with which to compare her actual payoff. Consequently, they are likely to be more conscious of the actual payoff. Conversely, BLUE does not have this initial endowment, and BLUE appears to have a commitment to reward a trusting RED when he is certain of her trusting intention. Alternatively, one may interpret that RED has more power than BLUE in the sense that she can choose not to initiate the trust pro-

Figure 3 Descriptive Statistics of Players' Choices in the High-Stake Games

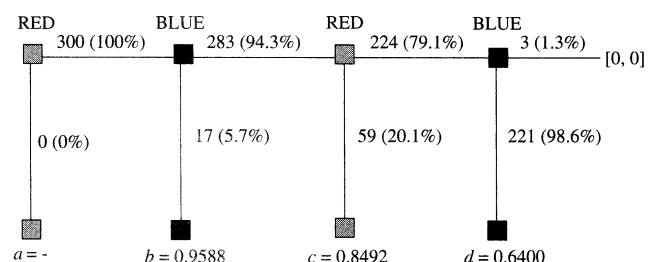


Table 3 A Test of Evolutionary Trust-Building Paths Across Subpopulations

	Estimate	t-statistic	p-value
RED			
α	0.5261	5.83	0.0001
β	0.2794	8.01	0.0001
Adj. R-Sq	0.8156		
BLUE			
α	0.8208	6.07	0.0001
β	-0.0591	-0.32	0.7489
Adj. R-Sq	0.2136		

cess without any inputs from BLUE. This difference in perceived power can cause her to be more sensitive to actual payoffs.

6. Discussion

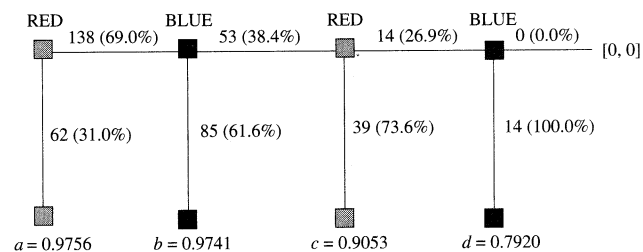
This research is designed to examine trust building in communities with anonymous exchanges. We define an anonymous exchange as one in which individuals interact with each other only once and do not know the identity of their partners. We disentangle trusting and trustworthiness behaviors using a novel experimental design. We find players in the population are not trustworthy when they are uncertain about others' trusting intentions. Interestingly, even with a total lack of formal mechanisms and little (if any) informal sanctions due to the minimum scope of social relations, we find evidence of trusting behavior. Stake size appears to increase both trusting and trustworthy behaviors. Subpopulations, however, may follow different evolutionary paths of trusting behaviors. Some subpopulations may trust based on its relative fitness and others may not. Results suggest subpopulations with the decision rights to initiate the trust building and those who hold the first rights to continue the process follow different evolutionary paths.

6.1. Methodological Issues

The experimental design allows us to differentiate and examine trust and trustworthiness through the choice set of subjects. Choices consist of two components—strategy (trust or not) and division of social gains (trustworthiness). Many previous studies fail to distinguish between these components of the trust process because they test trust using a sequential PD design (Heckathorn 1988). Results suggest that there is value in separating these components because we do see divergence in behavior. For example, in the initial stage of our game there is significantly more trusting behavior than trustworthiness. We find a decrease in trusting behavior as the potential value of trust drops.

Our design uses a social exchange protocol in which players must decide what to do without knowing the choices of others (i.e., contingent strategy). Game theory predicts that using this design should not alter the strategy choices of subjects. However, previous research in PD games suggests that subjects are more likely to cooperate if they know the other subjects are willing to cooperate (Kiyonari et al. 2000, McCabe et al. 2000). If this finding generalizes to our trust game, then we should see more trust-building behavior when subjects are provided with the choices of others in previous stages.

Figure 4 Descriptive Statistics of Players' Choices (Sequential Exchange Protocol)



We test this empirically.¹⁵ We ran two additional 20-subject experiments in the United States using a sequential exchange protocol to determine whether social exchange protocol has an impact on trust building. That is, subjects were not asked for contingent strategies. In each stage, they were told the choice of their pair members in the previous stage. For example, if RED passed in Stage 1, BLUE was told this and then asked what he wanted to do in Stage 2. If BLUE passed in Stage 2, we went back to RED and asked what she wanted to do in Stage 3, and so on. Figure 4 reports the results of these experiments. If we compare Figure 4 with Figure 2, we see that subject behaviors are alike. At each of the four stages, the degree of exhibited trust and trustworthiness is very similar. It is as if subjects can predict the behavior of others without knowing their action choices.

6.2. Theoretical Implications

Prior explanations for trust building generally revolve around the promise of future interactions. Axelrod (1984) claims this shadow of the future helps to develop trust. Our experiment design allows us to test this formally. There is no shadow of the future; subjects are told they will not interact for more than one round.¹⁶ Verbal or nonverbal cues are not possible because subjects do not play face to face and play is anonymous, yet we see a significant level of trust and a nonnegligible level of trustworthiness develop.

We conjecture the following: In our experimental setup, trusting may lead to higher payoffs, so it is not surprising to see people trust. In fact, people seem to trust more the higher the likelihood of a greater payoff. For example, RED subjects who trust in Stage 1

¹⁵ In addition to testing behaviors in different exchange protocols, this manipulation provides an indirect test of whether or not subjects can predict others' intentions. Our results suggest they can, in this setting.

¹⁶ We take steps to ensure that subjects believe in the instructions. For example, they are explicitly told that they will only interact with one another once. Subjects are also told that there are 10 RED subjects and 10 BLUE subjects in each room so that we have the required number of subjects to match them with each other for each round. We also have shown subjects our scoring sheet before the experiment, which clearly shows the different pairings in each round.

have a higher probability of receiving a payoff greater than the nontrusting payoff of 4, relative to that of the trusting RED subjects in Stage 3 receiving a payoff greater than the nontrusting payoff of 16. Hence, it is the promise of social gain that drives this trust-building process.

It is, however, more difficult to rationalize the significant level of trustworthiness behavior in the last stage. As discussed before, subjects at Stage 4 exhibit the greatest level of trustworthiness because they are certain about the trusting intention of others. Some authors have suggested that our trust game can be interpreted as a multistage dictator game. Clearly, our game at Stage 4 is a single-stage dictator game (Kahneman et al. 1990; Forsythe et al. 1994; Hoffman et al. 1994, 1996; Malhotra and Murnighan 2000).¹⁷ Thus it might be interesting to compare our results with those from single-stage dictator games. Payments to the respondents (or in our case the RED subjects) are often interpreted as altruistic behavior (Camerer 2003). At Stage 4 of our game, this interpretation may not be appropriate, because our Stage 4 also contains trusting history from prior stages and subjects may want to reward this past trusting behavior.

We compute the mean proportion of payment to respondents in three similar studies that allow proposers to divide the property rights continuously from 0% to 100% (Forsythe et al. 1994; Hoffman et al. 1994, 1996). The mean payment to respondents is 17% and the modal response is 0% across seven of the nine experimental conditions.

Our mean proportion is 22.6%. This is higher than that from the above three dictator game studies. We speculate that some BLUE subjects initially reward RED's trusting behavior because they are certain about their trusting intention. However, this effect wears off gradually. This conjecture is supported by the fact that our modal response in Rounds 1–5 was 50%, whereas in Rounds 6–10 it was 0%.¹⁸

¹⁷ Kiyonari and Yamagishi (1997) devised an interesting variant of the dictator game in which the respondent has an option whether or not to trust the proposer for a division. In their studies, the respondent could opt out for a sure payoff, which is less than half of the proposer's property rights. They found more than 80% of their subjects trusting; the average division of the proposer was 42%. Compared to our results from Stages 3 and 4, these were significantly higher trusting and trustworthiness behaviors. One reviewer suggests that this difference could be due to the fact that a truster's payoff is totally dependent on a trustee's action in Kiyonari and Yamagishi's game, whereas the same action only accounts for one-tenth of a truster's payoff in our design.

¹⁸ In Rounds 1–5, of the 354 observations of BLUE subjects passing in Stage 2 and taking in Stage 4, 131 times (37.0%) they took 50% and 110 times (31.1%) they took 100%. In Rounds 6–10 they took 50% 66 times (21.6%) and 100% 177 times (58.0%) out of 306 times.

Overall, our results suggest a significant number of subjects exhibit pseudotrust when interacting anonymously. This pseudotrust is cultivated in order to harness a return in exchange for the risk taken in giving up property rights. As long as everyone takes this risk, it benefits the population, because this pseudotrust increases social gain. This expectation becomes a self-fulfilling prophecy or a population-level behavioral norm. For such a behavioral norm to develop, it is crucial that each player has the opportunity to regain the increased property rights. This is the case in the first three stages of the game. Consequently, players who want to stop trusting are not trustworthy. Hence, it is possible to have a population that exhibits pseudotrust, but is not trustworthy.

To sustain the pseudotrust behavior of BLUE in Stage 2, there must be a nonnegligible proportion of RED subjects who trust in Stage 3. These RED subjects clearly trust for social gains because they will not regain the property rights. Figure 2 shows about 20% of RED subjects exhibit this trusting behavior. The result that 20% of our population trust for social gain without any formal enforcement mechanism or informal sanctions mirrors results of earlier studies. For example, similar to the dictator-type games discussed above, Camerer and Weigelt (1988) find about 17% of their subjects exhibit this type of trust. These findings suggest that about 15%–20% of the subject population are willing to trust even in anonymous exchanges.

Acknowledgments

The authors thank seminar participants at the ESA meeting and the University of Pennsylvania for helpful suggestions, Colin Camerer for his insightful comments on an earlier draft, and Christophe Van den Bulte for his suggestions and advice on data analysis. They are especially grateful to Karen Jehn for her early collaboration and ideas. This research is partially supported by NSF Grant SBR 9730187 and the Reginald Jones Center of Management at the Wharton School.

References

- Alchian, Armen, Harold Demsetz. 1973. The property rights paradigm. *J. Econom. History* 33 16–27.
- Axelrod, R. 1984. *The Evolution of Cooperation*. Basic Books, New York.
- Berg, Joyce, John Dickhaut, Kevin McCabe. 1995. Trust, reciprocity, and social history. *Games Econom. Behavior* 10 122–142.
- Bradach, J., R. Eccles. 1989. Price, authority, trust: From ideal types to plural forms. *Ann. Rev. Sociology* 15 97–118.
- Camerer, Colin. 2003. *Behavioral Game Theory: Experiments on Strategic Interaction*. Princeton University Press, Princeton, NJ.
- Camerer, Colin, Teck-Hua Ho. 1999. Experience-weighted attraction learning in normal-form games. *Econometrica* 67(4) 827–874.
- Camerer, Colin, Robin Hogarth. 1999. The effects of financial incentives in economic experiments: A review and capital-labor production framework. *J. Risk Uncertainty* 19 7–42.

- Camerer, Colin, Keith Weigelt. 1988. Experimental tests of a sequential equilibrium reputation model. *Econometrica* 56(1) 1–37.
- Camerer, C., Teck-Hua Ho, Juin-Kuan Chong. 2002. Sophisticated learning and strategic teaching. *J. Econom. Theory* 104 137–188.
- Camerer, C. F., T-H. Ho, J-K. Chong. 2004. A cognitive hierarchy model of one-shot games. *Quart. J. Econom.* 119(3) 861–898.
- Chiles, T. H., J. F. McMackin. 1996. Integrating variable risk preferences, trust, and transaction cost economics. *Acad. Management Rev.* 21 73–99.
- Craswell, Richard. 1993. On the uses of “trust”: Comment on Williamson, “Calculativeness, trust, and economic organization.” *J. Law Econom.* 36 487–500.
- Creed, W. E. D., R. E. Miles. 1994. Trust in organizations: A conceptual framework linking organizational forms, managerial philosophies, and the opportunity costs of controls. R. M. Kramer, T. R. Tyler, eds. *Trust in Organizations: Frontiers of Theory and Research*. Sage, Thousand Oaks, CA, 16–38.
- Forsythe, Robert, Joel Horowitz, N. Z. Savin, Martin Sefton. 1994. Fairness in simple bargaining experiments. *Games Econom. Behavior* 6 347–369.
- Friedman, D. 1991. Evolutionary games in economics. *Econometrica* 58 637–666.
- Fukuyama, F. 1995. *Trust: The Social Virtues and the Creation of Prosperity*. The Free Press, New York.
- Glaeser, Edward, David Laibson, Jose Scheinkman, Christine Soutter. 2000. Measuring trust. *Quart. J. Econom.* 115(3) 811–846.
- Granovetter, M. 1985. Economic action and social structure: The problem of embeddedness. *Amer. J. Sociology* 91 481–510.
- Greene, Graham. 1943. *The Ministry of Fear*. Penguin Books, New York.
- Gurtman, M. B. 1992. Trust, distrust, and interpersonal problems: A circumplex analysis. *J. Personality Soc. Psych.* 62 989–1002.
- Hayashi, Nahoko, Toshio Yamagishi. 1998. Selective play: Choosing partners in an uncertain world. *Personality Soc. Psych. Rev.* 2(4) 276–289.
- Heckathorn, Douglas. 1988. Collective sanctions and the emergence of prisoner’s dilemma norms. *Amer. Sociology Rev.* 55 366–384.
- Hertwig, Ralph, Andreas Ortmann. 2001. Experimental practices in economics: A challenge for psychologists. *Behavioral Brain Sci.* 24 383–403.
- Ho, Teck-Hua, Keith Weigelt. 1996. Task complexity, equilibrium selection, and learning: An experimental study. *Management Sci.* 42 659–679.
- Ho, Teck-Hua, Keith Weigelt. 2004. Trust building among strangers. Working paper, University of California, Berkeley, Berkeley, CA.
- Hoffman, Elizabeth, Kevin McCabe, Vernon Smith. 1996. Social distance and other regarding behavior in dictator games. *Amer. Econom. Rev.* 86(3) 653–660.
- Hoffman, Elizabeth, Kevin McCabe, Vernon Smith. 1998. Behavioral foundations of reciprocity: Experimental economics and evolutionary psychology. *Econom. Inquiry* 36(3) 335–352.
- Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, Vernon Smith. 1994. Preferences, property rights and anonymity in bargaining games. *Games Econom. Behavior* 7 346–380.
- Jenkins, Douglas Jr., Atul Mitra, Nina Gupta, Jason Shaw. 1998. Are financial incentives related to performance? A meta-analytic review of empirical research. *J. Appl. Psych.* 83 777–787.
- Kahneman, Daniel, Jack L. Knetsch, Richard Thaler. 1990. Experimental tests of the endowment effect and the Coase theorem. *J. Political Econom.* 98(6) 1325–1348.
- Kiyonari, Tohko, T. Yamagishi. 1997. A new game of trust: Letting the partner be the dictator. *7th Internat. Conf. Soc. Dilemmas*, Cairns, Australia.
- Kiyonari, Tohko, Shigehito Tanida, Toshio Yamagishi. 2000. Social exchange and reciprocity: Confusion or a heuristic? *Evolution Human Behavior* 21(6) 411–427.
- Kollock, P. 1994. The emergence of exchange structures: An experimental study of uncertainty, commitment, and trust. *Amer. J. Sociology* 100 313–345.
- Malhotra, Deepak, J. Keith Murnighan. 2000. Self versus social impression management in dictatorship games. Unpublished manuscript, Northwestern University, Evanston, IL.
- Mayer, R. C., J. H. Davis, F. D. Schoorman. 1995. An integrative model of organizational trust. *Acad. Management Rev.* 20 709–734.
- McAllister, D. J. 1995. Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Acad. Management J.* 38 24–59.
- McCabe, K., V. Smith, M. LePore. 2000. Intentionality detection and “mindreading”: Why does game form matter? *Proc. National Acad. Sci. United States of America*, Vol. 97. National Academies of Science Press, Washington, D.C., 4404–4409.
- McKelvey, R., T. Palfrey. 1992. An experimental study of the centipede game. *Econometrica* 60 803–836.
- Orbell, John, Robyn Dawes. 1991. A “cognitive miser” theory of cooperators’ advantage. *Amer. Political Sci. Rev.* 85 515–528.
- Orbell, John, Robyn Dawes. 1993. Social welfare, cooperators’ advantage, and the option of not playing the game. *Amer. Sociology Rev.* 58 787–800.
- Ring, P. S., A. H. Van de Ven. 1992. Structuring cooperative relationships between organizations. *Strategic Management J.* 13 483–498.
- Rosenthal, R. 1982. Games of perfect information, predatory pricing, and the chain-store paradox. *J. Econom. Theory* 25 92–100.
- Shapiro, S. P. 1987. The social control of impersonal trust. *Amer. J. Sociology* 93 623–658.
- Smith, Vernon, James M. Walker. 1993. Rewards, experience, and decision costs in first price auctions. *Econom. Inquiry* 31 237–244.
- Wright, Robert. 1995. *The Moral Animal*. Vintage Books, New York.
- Yamagishi, T., M. Yamagishi. 1994. Trust and commitment in the United States and Japan. *Motivation Emotion* 18 9–66.
- Yamagishi, T., K. A. Cook, W. M. Watabe. 1998. Uncertainty, trust, and commitment formation in the United States and Japan. *Amer. J. Sociology* 104 165–194.
- Zucker, L. 1986. Production of trust: Institutional sources of economic structure, 1840–1920. *Research in Organizational Behavior*, Vol. 8. JAI Press, Stamford, CT, 53–111.