Trust Building via Risk Taking:
A Cross-Societal Experiment*

KAREN S. COOK
Stanford University

TOSHIO YAMAGISHI
Hokkaido University

COYE CHESHIRE
Stanford University

ROBIN COOPER
Stanford University

MASAFUMI MATSUDA
Hokkaido University

RIE MASHIMA
Hokkaido University

The role of risk taking in building trust relations has largely been overlooked in the burgeoning literature on trust in the social sciences; yet it is central to understanding how trust develops. We argue that a series of risk-taking behaviors is indispensable to building a trust relation. We conducted experiments in Japan and the United States to examine the independent and cross-cultural effects of risk taking on trust building. The results of these experiments indicate that the American participants took more risks than did the Japanese, supporting the general claim that Americans are inclined toward risk taking and trust building. Even so, the Americans were no better than the Japanese at improving the level of cooperation. The cumulative results of these experiments imply that risk taking is a critical element in trust building for Americans, but less so for the Japanese. Our results show clearly that it is important to distinguish trusting behavior from cooperation and to measure them separately if we are to study trust and trust building in relation to social cooperation.

The role of risk taking in building trust relations has largely been overlooked in the burgeoning literature on trust in the social sciences (Cook 2001; Hardin 2002); yet it is central to understanding how trust develops. In the absence of monitoring and the sanctioning of opportunistic behavior, trusting always involves some risk.[1] We define trust building as the process through which social interaction opportunities involving risk are transformed into trust relations in which the people involved come to trust each other and honor that trust. How does this happen?

We argue that a series of risk-taking behaviors is indispensable to building a trust relation (Blau 1964; Holmes and Rempel 1989). In a typical trust-building scenario, two people realize that they can potentially gain from engaging in social exchange. The exchange may take the form of a date, a pleasant conversation, or even a business transaction. Each party knows that he or she

[1] The risk involved can be either the standard risks in economic exchanges or social risk—that is, the risk of failing to locate a partner for exchange or of losing an exchange partner. Our experimental setting operationalizes the risk involved primarily in economic terms, though social risk is involved because an actor must engage with another actor in order to profit. In some settings the risk can simply be the risk of loss from an impersonal investment decision (as in Hsee and Weber 1999).

will gain from an exchange with a potential partner if that partner turns out to be trustworthy. Yet there is always the risk of potential loss if the partner is not.

Typically there is no basis for expecting trustworthy behavior from a new potential exchange partner. In addition, in many situations no regulatory agency exists to control behavior, and there is little initial information about the reputations of possible exchange partners. Moreover, monitoring and sanctioning are usually too costly, even in specific organizational settings. Under these conditions, few rational people would engage in exchange; thus the benefits that could be obtained from mutual exchange are not realized.

In situations containing the clear risk of opportunism, a trust relation in which one person can expect trustworthy behavior from another becomes a highly valuable commodity. Such a relation is too valuable to jeopardize by engaging in the kind of myopic opportunistic behavior that generates short-term gain but blocks the possibility of building a long-term relationship. If you have a trustworthy friend or a business partner who reliably acts in a trustworthy manner, you would not want to lose that person as an exchange partner. This implies that a trust relation is self-sustaining in the sense that each person has an incentive to maintain the relationship once it develops (Blau 1964; Hardin 2002). Unilaterally untrustworthy behavior in an exchange situation destroys the opportunity to gain benefits that can only be generated through mutual trust.

A major obstacle to trust building, however, is the *initial* lack of trust in one's potential exchange partner. One way to encourage another to act in a trustworthy manner is to increase one's own value to that person. Yet to make oneself valuable to a potential partner, one somehow must invite that person to be trusting. This situation is a classic "Catch-22": each party must induce her or his partner to be trusting before actually proving her own trustworthiness. Therefore, a unilateral act of trust by one partner, involving risk taking,[2] is

required to break the deadlock of a mutual lack of trust.

The key to success in breaking this deadlock lies in the use of a GRIT (graduated reciprocation in tension reduction) strategy as it has been called in the conflict literature. Osgood (1962) proposed GRIT as an effective methodology for mutual disarmament during the cold war, as well as for other conflict situations involving risk. He proposed that the series of moves aimed at eliminating tensions must be both *graduated* and *reciprocated*. A player starts with a unilateral yet low-risk move to alleviate tension, and then waits until his partner reciprocates that move. When it is reciprocated, the next initiative for alleviating tension involves a little more risk. According to Etzioni (1967, 1969), this strategy led to success in the "Kennedy Experiment," which took place between June 10 and November 22, 1963, in reducing tension between the United States and the Soviet Union during the Cuban missile crisis.

Without actual knowledge of the GRIT strategy, the most successful participants (those who earned the most) in Matsuda and Yamagishi's (2001) recent experiment used basically a GRIT-type strategy in building trust with their partners. Initially they cooperated unilaterally, entrusting only a small amount to the partner; then they increased the amount over time as the partner demonstrated her or his trustworthiness. A relatively high level of mutual cooperation was achieved in this study, in which participants used a new variant of the prisoner's dilemma setting called a PD/D. In this setting, in contrast to the standard, one-shot prisoner's dilemma or other, repeated PD experiments, the participants can lower the stakes involved and minimize their initial risk (the amount they entrust) without compromising their own willingness to cooperate until a trust relationship has formed.

In the ordinary repeated prisoner's dilemma (PD), this strategy is not possible because the choice of cooperation or defection may reflect either trust (or lack of it) or willingness (or unwillingness) to cooperate. Those who are afraid that their partner will not cooperate often refuse to cooperate even when they might be willing to cooperate with someone they assume might be generally

---

[2] This is the same risk as is involved in the first move in reciprocal exchange (see Molm and Cook 1995; Molm, Petersen, and Takahashi 2001).

cooperative (Arneson 1982; Pruitt and Kimmel 1977). Those who do not trust their partners and want to avoid being exploited by their partners have no choice other than defection; typically this generates a spiral of mutual noncooperation. In the new experimental setting (PD/D), however, one can continue to cooperate, as did many of the participants in Matsuda and Yamagishi's (2001) experiment, even when they did not trust their partner on a specific trial. Instead they reduced the amount at stake, signaling a reduction in trust. In short, the major obstacle to trust building—the deadlock involved in a mutual lack of trust—can be overcome by engaging in a series of *graduated* and *reciprocated* risk-taking opportunities.

Trust Building Among Americans and Japanese: Uncertainty Avoidance

The argument that trust building requires risk taking implies that those who are risk-averse or bothered by high levels of uncertainty may find it difficult to enter trust relationships. They may prefer more formal commitment mechanisms (what Yamagishi calls "assurance structures") as a means of avoiding the risk of exploitation or the uncertainty of not finding an exchange partner. Our decision to examine trust building in the United States and Japan is based on our knowledge that Americans and Japanese are reported to differ greatly in their orientations to risk taking. Japanese generally are more concerned about avoiding uncertainty and risk than are Americans; some Americans even prefer to seek risks. If this is so, we would expect to observe a *significant difference between Americans and Japanese in the process of trust building*. That is, the large cultural difference in their tendencies to avoid uncertainty should be reflected in their behavior in exchange situations.

For example, Americans have been shown in survey research to have a higher level of general trust[3] than the Japanese

(Hayashi et al. 1982; Yamagishi and Yamagishi 1994). Furthermore, Americans are more cooperative than Japanese in *N*-person PDs in the absence of opportunities to sanction defectors (Yamagishi 1988). More generally, the current findings on the role of risk taking in building trust among American and Japanese participants are consistent with Yamagishi and his colleagues' characterization of Japanese social relations primarily as assurance relations rather than as trust relations, in contrast to the results for Americans (Yamagishi 1998; Yamagishi, Cook, and Watabe 1998; Yamagishi and Yamagishi 1994).

In a cross-cultural examination of work-related values, Hofstede (1980; also see Triandis 1993) shows that the United States and Japan differ on their levels of uncertainty avoidance and on their place on the continuum from individualism to collectivism. In a list of 40 modern countries, the United States registers the highest individualism score (91), whereas Japan (46), is near the middle. The situation is reversed, however, for uncertainty avoidance: with controls for the age of those surveyed, Japan emerges near the top of the list (112), while the United States is closer to the bottom (36). Although the United States and Japan differ on both scales, they are differentiated especially on their levels of uncertainty avoidance. This difference in orientations to risk suggests that the Japanese are more likely than Americans to prefer a social interaction situation in which the level of social uncertainty involved is lower (but see Buchan, Croson, and Dawes 2002).[4]

---

[3] General trust is measured using survey items such as "Generally speaking, would you say that most people can be trusted, or that you can't be too careful dealing with people?" Some authors (e.g., Hardin 2002) argue that these items measure lack of cynicism or a general sense of optimism in dealing with others rather than anything we would call trust in a relational sense. Although we agree in general with this criticism, we use the term *general trust* here instead of *optimism* because it has been employed so widely in the trust literature to refer to generalized orientations toward trusting, especially in cross-national studies. Levi and Stoker (2000), among others, offer additional criticisms of the standard survey items on trust.

[4] Recent research comparing risk preferences in the United States and China (Hsee and Weber 1999; Weber, Hsee, and Solowska 1998) also indicates cultural differences in risk seeking related to individualism and collectivism. Although researchers both in the United States and in China predicted that Americans would be more inclined to seek risks, the results suggest that the Chinese actually are more risk-seeking than Americans, but only in regard to

One strategy for reducing social uncertainty in exchange situations is to form committed relations with particular partners. In this context, commitment is primarily a mechanism for avoiding uncertainty (see Kollock 1994): One continues to engage in exchange with one partner to the exclusion of others, even in the face of more profitable alternatives. Although Americans may form commitments when uncertainty and risk are high, they try to avoid doing so if they can. Instead they may be more willing, early in exchange relationships, to take the type of low-level risks that enable trust building. The differences in Japanese and American psychological orientations may be reflected in exchange behavior at varying levels of uncertainty and risk. We investigate these behavioral differences in our experiment.

Kollock (1994) provides a clear account of commitment formation as uncertainty avoidance. Those high in uncertainty avoidance will be more in need of assurance mechanisms, and thus more likely to form committed relations in high uncertainty (Yamagishi et al. 1998). Kollock compares two commodity markets in southeast Asia, the rubber market and the rice market, and argues that the level of social uncertainty or risk of default involved in the trade of rice is quite different from that for raw rubber. The quality of rice is immediately apparent upon simple inspection; thus the rice buyer faces little risk of being cheated on quality. In contrast, the quality of raw rubber can be known only after it has been processed. Thus cheating on quality in the trade of raw rubber is easier, and the consequences of being cheated are extremely damaging to the buyer. If the rubber is bad, it will be worth less than the price paid for it. The difference in social uncertainty and in the risk involved in the trades of these commodities, Kollock argues, explains the observed difference in the dominant form of trade. Rice is usually traded at open markets between strangers, whereas rubber is usually traded between particular producers and brokers who form long-term relationships; often these extend over several generations within families (also see Farrell 2004).

This general argument implies that trust building via risk taking as a means of dealing with social uncertainty is preferred more strongly by Americans than by Japanese, whereas commitment formation as a mechanism for uncertainty avoidance[5] is preferred more strongly by the Japanese. Several lines of research support this claim. First, we find evidence in a cross-societal experiment of elementary and instrumental cooperation in social dilemmas (Yamagishi 1988). Both American and Japanese participants in this experiment played a repeated social dilemma game. In one condition, they had an additional choice: establishing a sanctioning system that punished noncooperators. Such an opportunity was not provided in the control condition. The results demonstrated that Americans were more cooperative in the absence of a sanctioning system that disciplines members' behavior, whereas the Japanese were more willing to contribute to the establishment of such a system. Furthermore, the opportunity to establish a sanctioning system improved the cooperation levels of the Japanese participants more strongly than that of the Americans.

An implication of this experiment relevant to the current study is that the Japanese are less willing than the Americans to engage in cooperative behavior in the absence of an assurance system that reduces social uncertainty, and are more willing to engage in behavior that reduces such uncertainty. In another cross-societal study, researchers compared how levels of general trust affect American and Japanese subjects' tendency to form commitment relations (Yamagishi et al. 1998). The results of two experiments provide strong evidence for the argument that distinguishes between com-

---

investments, not in the social domain. Hsee and Weber speculate that this may be the case because the Chinese live in a collectivist culture, where family and friends would provide a "cushion" in the event of a need for financial resources. Further research is needed to fully explore the factors that contribute to such cross-cultural differences.

[5] We examine the behavioral consequences of risk taking for trust building, but lack a direct measure of uncertainty avoidance. In future research we plan to develop such a measure to directly assess the assumption that uncertainty avoidance mediates the effects we observe in our experiment.

mitment formation as uncertainty avoidance and trust as risk taking. High trusters—those who showed a high level of general trust in their responses on a trust scale—are less likely than low trusters to form commitments when faced with a socially uncertain situation (Yamagishi et al. 1998).

The results of these experiments provide evidence that in an uncertain situation, those who prefer to form a commitment relation with a particular partner and thus reduce the risks within such a relation are less trustful of other people in general. The same finding concerning the effect of general trust on commitment formation was obtained with both American and Japanese participants: those low in general trust of others in both societies are more likely to form commitment relations with trustworthy persons despite the opportunity cost involved. The predicted difference in our experiment between the Japanese and the Americans reflects not only relative levels of uncertainty avoidance but also different levels of general trust: typically the latter are higher in the United States than in Japan.

In the experiment reported below, we provide an empirical test of our argument that trust building requires risk taking. To test our argument, we first compare two types of experimental games: the standard prisoner's dilemma (PD) game and the newer prisoner's dilemma game with variable levels of dependence (PD/D). In the remainder of this paper we will refer to the PD/D game as the PD/R game (prisoner's dilemma game with risk). The PD/R game is an exact replica of what was called the PD/D[6] game in previous research (e.g., Matsuda and Yamagishi 2001). Yet because we focus here on *risk taking*, we refer to this game as the PD/R game. In addition, we test our argument that trust building requires risk taking by comparing the levels of trust and cooperation exhibited by American and Japanese participants.

Prisoner's Dilemma With Risk: A New Experimental Paradigm

To study trust, researchers first used a standard prisoner's dilemma (PD) paradigm initially designed for experiments on cooperation. These researchers simply used cooperative behavior in the PD game as an *indicator of trust* (e.g., Deutsch 1973; Lindskold 1978; Meeker 1983; Pilisuk and Skolnick 1968; Solomon 1960). Because the PD game was designed to study cooperation, not trust, this move confounded measures of trust and cooperation. If the role of trust is to ease the way to cooperation, treating cooperative behavior as a measure of trust would have made the experimental evidence circular. Does trust lead people to cooperate, or does cooperation lead people to trust one another? We cannot determine the answer to this question from much of the existing experimental evidence. Furthermore, factors other than trust are known to affect rates of cooperation in the PD; thus the interpretation of cooperation as a direct expression of trust is dubious.

Modifications of the standard PD protocol subsequently were designed to generate distinct behavioral measures of trust and cooperation.[7] One modification is now known as the "trust game" (TG) or the "trust-honor game" (Dasgupta 1988; Kreps 1990; Snijders 1996). The trust game is similar to a PD in that individually rational choices by two players lead them to a Pareto deficient outcome. The TG is different from a PD, however, in that trusting behavior is clearly distinct from cooperative behavior. This significant difference between the PD and the TG is important for studying trust and the development of trust relations.

The element critically lacking in a standard PD game, as a means of studying trust, is what constitutes the *core* of an act of trust or

---

[6] For an earlier variant see Kakiuchi and Yamagishi (1997) and Yamagishi and Kakiuchi (2000).

[7] One of the first efforts to disentangle measures of trust and cooperation was an experimental study by Orbell and Dawes (1993). In their revision to the standard PD game, subjects were allowed an exit option: they could choose not to play at all. Those who remained in the game were described as displaying trust. The shortcoming of this design is that it allowed only a dichotomous measure of trust, not a continuous measure.

trusting behavior: taking risks and thus making oneself vulnerable. In a trust game, player 1's choice to behave in a trusting manner is an act of putting her or his fate in the hands of another player to achieve an outcome better than the status quo. In a standard trust game, for example, player 1 chooses whether or not to trust player 2, and player 2 chooses whether to honor player 1's trust. When player 1 chooses not to trust player 2, both players receive a small benefit ($10). To achieve the greater benefit of $20 each, however, player 1 *must take the risk* of potentially receiving a less desirable outcome ($0) if player 2 does not honor her trust. Player 2 clearly has an incentive not to honor player 1's trust because when she does *not* do so, she receives more ($30) than if she does so (only $20). In this case, once player 1 has chosen to take a risk by acting in a trusting manner, her fate is transferred entirely to the hands of player 2. Whether this act of "trust" engenders a more or a less desirable outcome than not trusting depends on player 2's action. If player 2 is "fair" and trustworthy, and honors player 1's trust, then "trusting" is certainly better for player 1; otherwise, not "trusting" is clearly the best choice.

In sharp contrast, in the PD game defection is *always* superior to cooperation—that is, it provides a more desirable outcome—no matter what one's partner does. Whether player 2 cooperates or defects does not affect the benefit player 1 earns from defecting rather than from cooperating. In game-theoretic terms, defection is the dominant behavioral choice for each player in the PD. In the game of trust, however, there is no dominant choice for player 1: the outcome depends solely on whether or not player 2 cooperates.[8]

Although TG succeeds in capturing the critical elements involved in trust and cooperation, it suffers from two significant limitations: it is *static* and *one-sided (or asymmetric)*. The first limitation recently has been removed as researchers have begun to

use a repeated TG rather than the one-shot TG for the study of trust relations. For example, Bolton, Katok, and Ockenfels (2003) study the development (or, more precisely, the maintenance) of trust and trustworthiness by letting the same two subjects play a trust game repeatedly with one another. They also resolve the second major problem with the TG by letting the players alternate between the roles of truster and cooperator during the experiment.

Another popular variant of the trust game, found in experimental economics, is the investment game (IG) developed by Berg, Dickhaut, and McCabe (1995). The IG is played between two players, A and B. As in the trust game, player A decides to trust or not to trust B, and B decides to honor A's trust or not in response. The difference between the TG and the IG is in the nature of the choices for actors A and B. In the TG, both A and B make binary choices: A between trusting and not trusting, B between honoring and not honoring A's trust. In the IG, they make continuous rather than binary choices: player A decides how much trust (indicated by level of investment) he or she will place in B, and player B decides how much to reciprocate the trust placed in him or her by A. Berg et al. (1995) provided A and B with an endowment of $10 each, and asked A to transfer to B *any* amount up to $10. The experimenter tripled the amount of money transferred to B. If, for example, A transferred $4 to B, B received $12. Player B, who received the transferred (and tripled) money in addition to her or his initial endowment of $10, then decided whether to send some or all of the money back to A. The IG thus captures the same elements of trust and cooperation as the TG, with the additional benefit of allowing the researcher to study varying levels of trust and cooperation. Berg et al. (1995) use the IG in a static and one-sided manner, however: it has *not* been employed to study the development of mutual trust and cooperation between the same two partners. The PD/R we introduce here can be regarded as a mutual and repeated version of a variant of the investment game. It enables us to study the emergence of mutual trust between the same pair of players.

---

[8] According to the logic of backward induction used by game theorists, however, "rational" player 1 should not trust because "rational" player 2 is expected not to honor her or his trust.

Like the typical trust game, PD/R allows us to separately measure both trust and cooperation behaviorally. The magnitude of the stake a player chooses is a direct reflection of her or his level of trust in her partner. This decision is clearly distinct from the act of cooperation versus defection. At the same time, the decisions in this game are symmetrical. The PD/R game thus is better suited than the ordinary PD for studying trust formation in dyadic relations or networks of dyads.

We elaborate on the details of the PD/R game[9] in the section on procedures, but the following is a brief overview of the game's structure. In the beginning of each game (or trial) in the PD/R, two players each are given 10 coins and are asked to decide how many of the coins (from one to 10) they want to entrust to their partner. The players make this decision simultaneously. Next they receive information on the number of coins entrusted to them by their partner. Each player then decides whether or not to return the coins entrusted to him or her. When a player returns the coins, the partner receives *double* the number she entrusted. When a player does not return the coins, they become

her gain and her partner's loss. The number of coins entrusted to a partner is the measure of the level of the player's trust in her partner, while the decision whether to return the coins entrusted to her is the measure of cooperation.[10] The PD/R game allows us to distinguish behavioral measures of trust from behavioral measures of cooperation as well as to examine reciprocal trust.

The Development of Trust Relations

The goal of this experimental study is to investigate the role of risk taking in the development of a trust relationship—a relationship in which two players both trust and cooperate at a high level. We aim to achieve this goal by comparing the cooperation levels in a standard PD game with those in the PD/R game described briefly above. Although both the PD game and the PD/R game involve entrusting coins to a partner, there is an important difference.

In the PD game, the number of coins to entrust is determined randomly: the player has no choice. In that game, the player's *only choice* is whether or not to return the coins that were entrusted to him or her. Figure 1 depicts how the PD game we use constitutes a prisoner's dilemma.

---

[9] The initial PD/R game was presented to the subjects in matrix form, but this was too complex for many players to fully comprehend. To alleviate such difficulties, Matsuda and Yamagishi (2001) introduced a new version of PD/R that retained all the relevant features of the original PD/R, while making the game intuitively easier to understand.

[10] Again, we elaborate this point more fully in the section on procedures, when we discuss each condition (and each phase in each condition).

Player 2's Choice

Player 1's Choice	Return	Not Return
Return	10 / 10	15 / 0
Not Return	0 / 15	5 / 5

Figure 1. An Example of the Prisoner's Dilemma Game Used in the Study
Note: In this example, each player is randomly assigned to "entrust" five coins to her or his partner. Each player has only a choice of returning or not returning the five coins entrusted by the partner. When the coins are returned, the number of coins doubles.
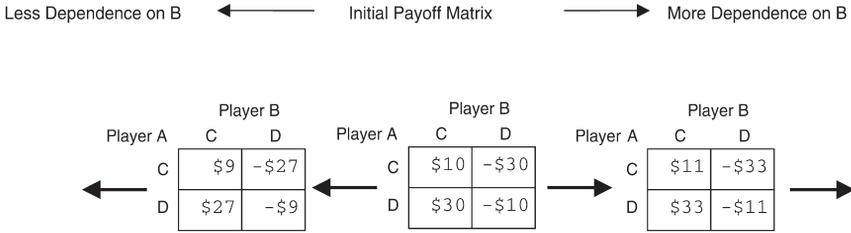
Less Dependence on B ← Initial Payoff Matrix → More Dependence on B

	Player B C	Player B D
Player A C	$9	-$27
D	$27	-$9

	Player B C	Player B D
Player A C	$10	-$30
D	$30	-$10

	Player B C	Player B D
Player A C	$11	-$33
D	$33	-$11

Figure 2. Prisoner's Dilemma with Risk (PD/R) for Player A
Note: Player A chooses to increase or decrease her or his dependence on B.

In contrast, in the PD/R game, each player can choose not only whether or not to return the coins that were entrusted to her, but also how many coins she wishes to entrust to her partner. The main difference between the two games is whether risk taking (whether to entrust a large number of coins) to build trust can take place. (See Figure 2) In the PD game, the player cannot take risks; in the PD/R game, the player can take a risk in deciding how many coins to entrust. By comparing the cooperation rates between the two games—that is, the proportions of the choices to return versus not to return the coins—we can examine whether giving people the opportunity to take a risk and to trust another (by entrusting a large number of coins) helps to develop a trust relationship.

We have argued that in some exchange situations, risk taking enhances cooperation. Here we examine whether this effect is more pronounced among American than among Japanese participants. Given the findings demonstrating a risk-avoidance tendency among the Japanese and Hofstede's (1991) finding that the Japanese are generally higher than Americans in uncertainty avoidance, we expect the Americans to engage in more trusting behavior—that is, to entrust more coins—than the Japanese.

We further investigate whether the development of trust relations will be facilitated by risk taking when a "shadow of the future" (Axelrod 1984) is present, compared with a situation when no such shadow of the future exists. To do this we compare the levels of trust (the number of coins players entrust to their partners) and cooperation (the return rate) in a fixed-partner game as compared with a random-partner game. In the fixed-partner game, the same two players play either the PD game or the PD/R game

repeatedly. In such a game it is possible to gradually increase the level of risk taking and trustworthy responses *within* a relationship. In the random-partner game, each player encounters a new partner each time and plays the PD/R game with that partner. Furthermore, players are not informed of the identity of their current exchange partner. Thus, in the random-partner game, it is not possible to gradually build a trust relationship with a specific person; therefore no "shadow of the future" is present.

Because trust building with a particular person is impossible in the random-partner PD/R, it is doubtful that acting in a trusting manner could improve cooperation rates in this condition. There is one reason, however, to expect a higher level of cooperation in PD/R than in PD even in the random-partner situation, namely the signaling role of trusting behavior. That is, by acting in a trusting manner, a player can signal her or his intention to cooperate.[11]

The prisoner's dilemma and social dilemma literature on cooperation and defection consistently indicates that the choice to cooperate or defect is grounded in two distinct psychological states: greed and fear. On the one hand, those who care only about their own welfare and who are greedy usually defect in one-shot games. On the other, even not-so-greedy people who probably would prefer to cooperate rather than to defect will *defect anyway* because they expect that others will be unwilling to cooperate. In other words, they defect because of a fear of being exploited, not because of greed (See Pruitt

---

[11] In our design, acting in a trusting way is measured by the number of coins a player entrusts to his or her partner. More specifically, acting in a trusting way is what we call risk taking.

and Kimmel 1977; Yamagishi and Sato 1986). In addition to the fear that others will be greedy, a "second-order" fear may exist, namely that others will be similarly fearful and thus may defect for the same reason. Trusting behavior can alleviate this "second-order" fear. *Acting in a trusting manner (entrusting coins, risk taking) signals that a player is not afraid his or her partner will defect.* This action may eliminate the second-order fear in the partner.

Because second-order fear has not been studied until now, we cannot determine in advance its importance in determining the level of cooperation. Tentatively we expect this effect of signaling in reducing second-order fear to be relatively weak at best. The comparison between the fixed-partner and the random-partner PD/R game allows us to examine whether the positive effect of acting in a trusting manner on cooperation rates in the PD/R game is due to trust building in itself or to a simple signaling effect.[12]

We also investigate whether the participant's nationality—American or Japanese—makes a difference even in *random partner exchange*, in which participants interact with a randomly matched partner on every trial. We address whether a greater willingness to act in a trusting manner, as expected more strongly of the American participants than the Japanese, produces greater cooperation in the random-partner PD/R.

THE EXPERIMENT

Participants

Potential Japanese participants were recruited by telephone from a pool of first-year undergraduates enrolled at Hokkaido University. A total of 192 participants, including 115 males and 77 females, were selected and scheduled by phone to participate. American participants were recruited in an email message distributed to undergraduates living on campus at Stanford University. The message directed interested students to a website, where they completed a recruitment form on line. We selected 106 participants, 56 males and 50 females, and scheduled them according to their availability.

Overview of the Experiment

Four, six, or eight participants were scheduled to arrive at the laboratory at a particular time. The scheduler also assigned each subject a separate waiting room and told him or her to wait there for an experimenter. Thus participants were unable to see or talk with one another while they waited.[13] When everyone had arrived, each was taken separately to a workstation consisting of a small room with a chair, a desk, and a desktop computer.[14] Participants were given a consent form to read and sign. They used only the computer during the experiment and could call the experimenter via a help command if necessary. The computer software originally developed by Matsuda and Yamagishi (2001) was used in both countries, with translation from a Japanese display to an English display for the experiment in the United States.

Once the experimenter (located in the control room) started the program from the host computer, the participants were told to read and follow the instructions as they appeared on the screen. They were informed that (1) there were other participants; (2) they would be divided into pairs on each trial and would make decisions concerning exchanges with their partners; (3) they would be paid in accordance with the number of coins they acquired from each exchange; and (4) they would not know with whom they were exchanging, but they would know whether it was a new, randomly selected part-

---

[12] As one reviewer notes aptly, rather than fearing that my partner will defect, I may simply prefer to take a risk. A signaling effect does not distinguish between these two possibilities. The present analysis, as we stated, is simply exploratory. More direct measures would be required to assess the possible role of fear reduction versus simple risk taking, the main focus of this experimental investigation.

[13] A different procedure for scheduling the participants' arrival was used in the Japanese study, with the same effect: they were not allowed to see each other.

[14] In the American version, one room held two participants at the same time. These two workstations, however, were separated by a partition, and participants were brought in separately such that they could not see one another. In addition, the experimenter monitored these rooms closely so that they would not talk with one another.

ner or the same partner as in the previous trial (depending on condition).

Each participant was paired randomly with a *new* partner during each of the first 25 trials, and then for the remaining trials, was paired with the *same* partner (on the basis of cooperation rates) or with *different* partners depending on the experiment condition. After the experiment, each participant completed a short computerized questionnaire and was compensated according to the outcome of the experiment.

The unit of exchange in the computer program was called a "coin." At the end of the study, each coin the participants accumulated during the experiment was converted into cash worth 2 cents. Participants earned about $19 on average, with a minimum of $9 and a maximum of $28. The experiment took an average of 50 minutes to complete, including the post experimental questionnaire. The participants were debriefed at the computer before payment, and then were dismissed separately so that they would not see each other.

Procedure Summary

The experiment included three conditions: PD with a fixed partner, PD/R with a fixed partner, and PD/R with a random partner. Each condition had two phases. Table 1 presents a description of each phase in each of the three conditions.

Phase I

In Phase I, the participants engaged in a standard PD game and were matched with new, random partners on every trial. Phase I was exactly the same for all conditions. It included the first 24 trials in the fixed-partner condition in Japan and the first 25 trials in the random-partner condition in Japan, as well as all of the conditions in the United States.[15] Because players do not have the option to

determine how many coins they wish to entrust in the standard PD game, only cooperation rates (return = cooperate, do not return = defect) were measured in Phase I. At the end of Phase I, we informed each participant of her or his accumulated profit, as well as the amount of the highest profit obtained in the entire group.

We included the first phase in the design of the experiment for two reasons. First, we needed to measure each individual's base rate for her or his general cooperative tendency. The random-matching feature of Phase I prevented participants from engaging in strategic behavior, such as tit-for-tat, aimed at enhancing long-term profits. That is, participants played one-shot PDs repeatedly rather than an iterated PD. Thus Phase I did not include the "shadow of the future" (Axelrod 1984) which often leads a fixed pair who repeatedly play the same PD game to engage in mutual cooperation. The level of cooperation obtained in Phase I should reflect fairly accurately the participants' general cooperative tendencies.

The second reason why we introduced Phase I was that we expected the participants' mutual cooperation to be low during Phase I because of the lack of any "shadow of the future." This experience then would provide a strong motivational basis for building trust relations in Phase II (see Pruitt and Kimmel 1977).

Does cooperation in the PD/R improve among the initially low cooperators or the initially high cooperators in the study? On the one hand, initial cooperation may be low because players have not been given the opportunity to trust their partners independent of the choice to cooperate or defect. Thus, when they receive the option of determining how much to trust their partners, their overall level of cooperation should improve dramatically. On the other hand, the initially low cooperators may be general distrusters who have low expectations regarding other people's trustworthiness; at the same time, they may not be willing to learn from experience. If this is the case, low initial cooperators

---

[15] Although the number of trials was slightly different in Japan and in the United States, we have no reason to believe that this slight difference accounts for any discrepancies between the results obtained in the two countries. The Japanese data were collected first, and each experimental session took about an hour. In the United States, the experimental sessions were

conducted much more rapidly with the same number of trials, so the number of trials was increased slightly.

Table 1. Description of Phases in Each Experimental Condition

Condition	Phase I (24 Trials in Japan, 25 Trials in U.S.)	Phase II (36 Trials in Japan, 50 Trials in U.S.)
PD, Fixed Partner (Condition 1)	*Random partner* on every trial. *Cannot* control the amount to entrust to partner.	*Fixed partner* on every trial. *Cannot* control the amount to entrust to partner
PD/R, Fixed Partner (Condition 2)	*Random partner* on every trial. *Cannot* control the amount to entrust to partner.	*Fixed partner* on every trial. *Can* control the amount to entrust to partner.
PD/R, Random Partner (Condition 3)	*Random partner* on every trial. *Cannot* control the amount to entrust to partner.	*Random partner* on every trial. *Can* control the amount to entrust to partner.

may be unwilling to take risks to break the deadlock of a mutual lack of trust. Then again, because these two factors operate simultaneously, the effects may cancel each other out.

At this stage, we have no specific empirical or theoretical basis for making a particular prediction about these three possible outcomes. The results of the current experiment will provide a valuable basis for further theoretical development concerning this question. Thus we return to these issues after our discussion of the experimental results.

Phase II

In Phase II, participants engaged in either a PD with a fixed partner (condition 1), a PD/R game with a fixed partner (condition 2), or a PD/R game with a random partner (condition 3). Phase II included the remaining 36 trials in the fixed-partner condition in Japan and the remaining 45 trials in the random partner condition in Japan, as well as all of the conditions in the United States.

Condition 1: PD with fixed-partner exchange. In condition 1, Phase II trials consisted of the same PD game as the subjects played in Phase I. The only difference between Phase I and Phase II in condition 1 was that partners were random on each trial in Phase I, while partners remained *the same* on each trial in Phase II. In both phases, participants were unable to choose the amount they wished to entrust to their partners; the

computer determined this amount randomly. Thus only cooperation rates (how often players returned the entrusted coins) were measured.

Condition 2: PD/R with fixed partner exchange. In condition 2, at the end of Phase I, participants were told that they would have the same partners for the remainder of the experiment. We placed subjects in pairs by matching their cooperation rates[16] from Phase I, although we did not tell them so. In addition, in Phase I the subjects played the PD/R game instead of the standard PD game; thus they were allowed to choose the number of coins they wished to entrust to their partner on each trial.

We gave participants 10 coins on each trial, and they decided how many coins (from one to 10) to entrust to their partners. Participants then decided whether to "return" or to "keep" the coins entrusted to them by their partners. When they decided to return them, we doubled the number of coins and gave that number to their partners. When players decided to keep the coins, they kept exactly the number of coins entrusted to them: that is, the coins were not doubled. While the players were deciding whether to return or to keep the coins entrusted to them by their partners, their partners were making

---

[16] Matching on cooperation rates eliminates the potential confounding of differential cooperative tendencies between partners (or, more precisely, differences in their degree of optimism in their assessments of others' cooperativeness).

the same decision. At the end of each trial, participants learned whether their partners had returned the coins entrusted to them.

Condition 3: PD/R with random partner exchange. In condition 3, the Phase II trials consisted of the PD/R game with a random partner. Thus, although participants were able to control the number of coins to entrust to their partner on each trial, they could not use this ability to build a relationship with a single partner because they always received a new, randomly assigned partner after each trial.

In sum, the three experimental conditions were identical during Phase I. In Phase II, either participants could not control entrusting behavior and had a fixed partner (condition 1, PD-fixed); they could control entrusting behavior and had a fixed partner (condition 2, PD/R-fixed); or they could control entrusting behavior but had a randomly assigned partner (condition 3, PD/R-random).

Rules of the game: acquiring profits (all conditions). Participants in every condition acquired profits on each trial in the same way. First, they kept the coins they did not entrust to their partners. Second, they kept the coins their partners entrusted to them if they decided not to return those coins. Third, they received double the number of coins their partners returned to them. Participants were not allowed to use this profit on subsequent trials, however: at the beginning of each trial, they received 10 new coins for exchange. Depending on the experimental condition, either the participants decided simultaneously how many coins to entrust (PD/R), or the computer decided this amount randomly (standard PD). In all conditions, however, participants decided whether to return or to keep the entrusted coins. The computer displayed the number of total coins acquired by each person privately, but not those acquired by others.

The more coins participants entrusted to their partners, the more profit they received if their partners returned them. If their partners did not return them, however, the more coins they entrusted, the more they lost. Suppose a participant entrusts nine of her 10 coins to her partner. If the partner returns

them, the participant receives 18 coins, for a total of 19. If the partner chooses not to return them, however, she loses them and ends up with only one remaining coin. If a participant is afraid that her partner might not return the coins she has entrusted, she may choose instead to entrust only one coin to her partner. Even if her partner returns that coin, the participant receives only two coins and thus ends up with 11 (two plus the remaining nine). Therefore, the more coins a participant entrusts, the greater the potential gain (when the partner returns them) *and* the potential loss (when the partner does not return them).

If a participant is allowed to control the number of coins to entrust to her partner, then the number she chooses to entrust is a direct reflection of her trust in her partner. Trust thus is measured as the number of coins the participant entrusts to the partner. Cooperation is measured by the decision as to whether to return or to keep the coins entrusted by the partner to the participant. To return them is to cooperate; to keep them is to defect.

Hypotheses

Our general theoretical argument suggests, first, that allowing risk taking to play a role helps to build mutually cooperative relationships; and, second, that in building such relationships, risk taking in order to create trust should be more pronounced among Americans than among the Japanese.

The first hypothesis in this study concerns the effect of taking risks as an act of trust in improving cooperation. This hypothesis involves the comparison between the fixed-partner PD and the fixed-partner PD/R conditions. The standard PD allows participants only to choose whether or not to cooperate. In the PD/R, players can choose the amount they are willing to entrust to their partners on each trial *before* deciding whether to cooperate. We expect higher rates of cooperation in the PD/R than in the PD condition as a result.

Hypothesis 1: Both American and Japanese participants will cooperate at a higher level in the fixed-partner PD/R (condition 2) than in the fixed-partner PD setting (condition 1).

On the basis of previous findings concerning differential levels of uncertainty avoidance and commitment formation among Americans and Japanese, we predict that American participants will take larger risks to initiate trust relations (as reflected in their willingness to entrust a larger amount of money to their partners) than will Japanese participants.

Hypothesis 2: American participants will exhibit a higher level of trusting behavior than Japanese participants in both the fixed-partner PD/R (condition 2) and the random-partner PD/R (condition 3).

American participants' greater willingness to take risks and to trust their partners will lead to a higher level of mutual cooperation in the fixed-partner PD/R, in which building trust relationships between particular partners is possible. Although the same effect may occur in the random-partner PD/R condition, it should reflect only general cross-national tendencies toward uncertainty avoidance because the partners change on every trial.

Hypothesis 3: The positive effect of risk taking on cooperation rates predicted in Hypothesis 1 will be more pronounced among American than Japanese participants.

The next hypothesis addresses whether risk taking enhances cooperation, even without a "shadow of the future." Without the possibility of building a trust relationship between a particular pair of partners, taking a risk and trusting one's partner may not exert much effect on cooperation. In contrast, when one has the option of choosing how much to entrust to one's partner before deciding whether to cooperate, it is possible to use trusting behavior as a *signal* to convey one's willingness to cooperate. This option may reduce the partner's possible second-order fear of exploitation or it may simply signal willingness to take a risk on the partner. Thus we predict that the positive effect of choosing the amount to entrust before deciding whether to cooperate will be weaker when no "shadow of the future" is present. This implies:

Hypothesis 4: The cooperation rate in the random-partner PD/R (condition 3) will be lower than in the fixed-partner PD/R (condition 2).

To test whether or not cooperation is enhanced by choosing the level of risk one is willing to take, one can compare cooperation rates in the random-partner PD in Phase I with those in the random-partner PD/R in Phase II. In Phase I the computer determines the amount; in Phase II the participant makes this decision. Assuming that cooperation is improved by a reduction in the second-order fear of exploitation caused by indicating one's willingness to take a risk at some level, we predict:

Hypothesis 5: The cooperation rate in Phase II will be higher than in Phase I in the random-partner PD/R (condition 3).

Are American participants expected to cooperate in the PD/R game more fully than Japanese participants, even when there is no "shadow of the future"? In Hypothesis 2 we predicted that American participants will trust their partners more fully than will Japanese participants even in the random-partner PD/R, in which partners change on each trial. At the same time, we expect the choice of amount to entrust to one's partner to have a weaker effect on cooperation in the random-partner PD/R than in the fixed-partner PD/R. Therefore, we expect that the higher level of trusting behavior (indicated by higher levels of investment) expected of American participants in the random-partner PD/R will not particularly make them more cooperative than the Japanese participants. Given that partners are assigned randomly on each trial, differential levels of risk taking (or investment) should not have any impact on subsequent levels of cooperation. There is no reason to expect a cross-national difference in this effect.

Hypothesis 6: Allowing participants to choose the level of investment in Phase II of the random-partner PD/R (condition 3) will not affect cooperation rates differentially for American and Japanese participants in this condition.

Finally, we offer no specific predictions concerning cultural differences in the participants' behavior in the random-partner PD condition (Phase I of the experiment).

Individual differences in the participants' tendency to trust other people in general (i.e., general trust) are related to the level of cooperation; in addition, Americans, who are higher than Japanese in general trust (Yamagishi and Yamagishi 1994), are more cooperative in the *N*-person version of a PD or a social dilemma (Sato and Yamagishi 1986; Yamagishi 1986, 1988, 1990, 1992). These findings, however, have not been obtained consistently in dyadic PDs. The individual or cultural differences in the tendency to trust other people in general are less relevant in a dyadic relation, in which the participants face a particular partner, than in more diffuse *N*-person relations, where general trust might operate.

FINDINGS

To make the Japanese and the American data compatible, we decided to use only the first 60 of the 70 trials of American data. The 60 decision trials in the experiment were aggregated into 12 blocks, each consisting of five trials. The dependent variables to be analyzed are the cooperation rate[17] and the average number of coins entrusted to the partner in each trial block.[18]

Cooperation Rates in Phase I

Participants in all conditions in Phase I experience the same PD game, with random partners on each trial; thus we have no reason to expect any differences between the three conditions. As shown in Figure 3, however, we observe substantial unexpected differences in the cooperation rates in Phase I. A nationality × condition × trial block repeated-measure analysis of variance revealed a significant effect of the game condition, $F(1, 292) = 10.99$, $p < .0001$. None of the interaction effects involving the game condition were significant. The significance of the main

effect suggests a possible failure in the randomness of assigning participants into conditions. Yet, the lack of significant interaction effects involving the game condition suggests that the differences in the levels of cooperation rates in Phase I are not likely to interact with our other variables. Thus, in analyzing cooperation rates in Phase II below, we control for individual differences in levels of cooperativeness observed in Phase I. Figure 3 presents the average cooperation rate over the 12 trial blocks; Figure 4 depicts the average change in cooperation rate—that is, the difference in the average cooperation rate overall and the average cooperation rate in Phase I for the seven trial blocks in Phase II.

Other significant effects in this repeated-measure ANOVA are the main effect of trial block and the main effect of nationality. The main effect of trial block was highly significant, $F(4, 1168) = 10.37$, $p < .0001$. As shown in Figure 3, the cooperation rate in Phase I declined over trial blocks in all conditions. The interaction between trial blocks and game condition was not significant. The main effect of nationality, however, was significant, $F(1, 292) = 4.43$, $p < .05$. The Japanese participants (.42, sd = .26) were more cooperative than the American participants (.39, sd = .28), though this difference is not large.

Hypotheses 1 and 3

Hypothesis 1: Both American and Japanese participants will cooperate at a higher level in the fixed-partner PD/R (condition 2) than in the fixed-partner PD setting (condition 1)

As shown in Figure 3, the cooperation rate in the fixed-partner PD/R condition in Phase II was much higher than in the fixed-partner PD condition. To test the difference between the two game conditions, we conducted a nationality × game condition × trial block repeated-measure ANOVA in which the game condition included only the relevant conditions, namely the fixed-partner PD and the fixed-partner PD/R conditions. The main effect of the game condition in this ANOVA was highly significant, $F(1, 206) = 19.77$, $p < .0001$ ($F(1, 205) = 27.53$, $p < .0001$, when the cooperation level in Phase I is controlled). Furthermore, the game condition ×

---

[17] This rate transformed the binary response in each trial (returned versus did not return the entrusted coins) into a continuous variable.

[18] The fifth trial block (the last block in Phase I) in the Japanese data included only four trials, and the sixth trial block (the first block in Phase II) included six trials, because Phase I in the Japanese data consisted of 24 trials, not 25.
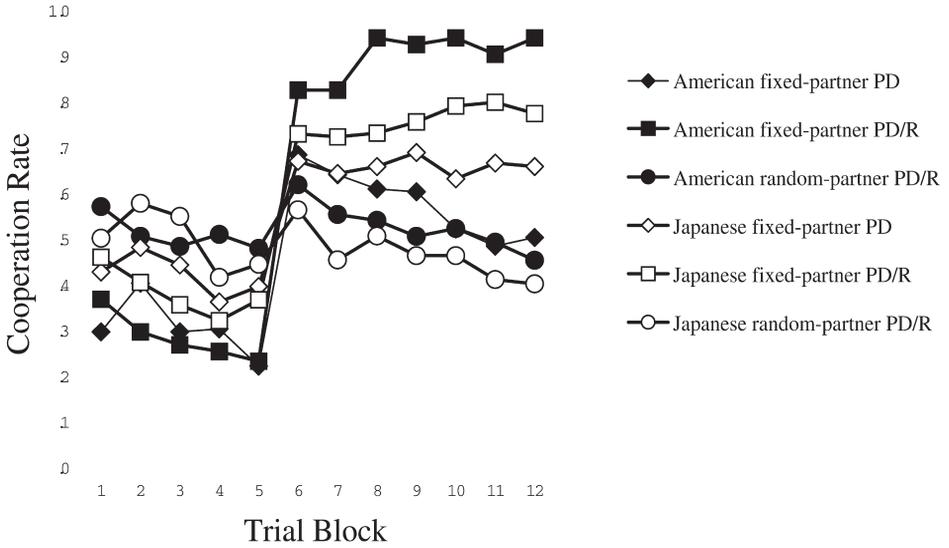
Figure 3. Average Cooperation Rate (Proportion of Coins Returned) Across Trial Blocks, American and Japanese Participants

trial block interaction also was highly significant, $F(1, 1236) = 7.94, p < .0001$. In trial block 6 (the beginning of Phase II), the cooperation rate in the fixed-partner PD/R was 7.5 percentage points higher than in the fixed-partner PD at the same trial block. This difference increased to 21.2 percentage

points by the last trial block (the end of Phase II), indicating that the cooperation rate indeed was much higher by the end of the fixed-partner PD/R (condition 2) than in the fixed-partner PD (condition 1). Hypothesis 1 thus was clearly supported.
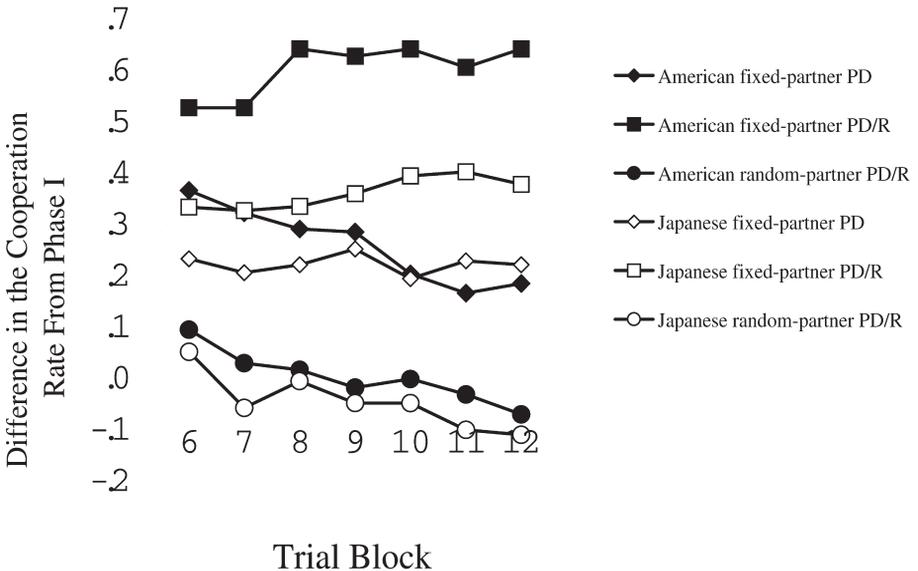
Figure 4. Difference in Cooperation Rate from Phase I Across Trial Blocks in Phase II, American and Japanese Participants

Hypothesis 3: The positive effect of risk taking on cooperation rates predicted in Hypothesis 1 will be more pronounced among American than Japanese participants.

As predicted in Hypothesis 3, the effect of choosing the amount to entrust was stronger among Americans than among our Japanese participants. The effect of the game condition interacted significantly with nationality, $F(1, 206) = 5.59, p < .05$ ($F(1, 205) = 6.09, p < .05$, with the cooperation level in Phase I controlled). During Phase II, the American participants' average cooperation rate was .90 in the fixed-partner PD/R game; .58 in the fixed-partner PD game; this difference was quite large (.32). In contrast, the Japanese participants' average cooperation rate was .76 in the fixed-partner PD/R game and .66 in the fixed-partner PD game, a much smaller difference (.10). The main effect of nationality was not significant, $F(1, 206) = .41$, ns. Finally, the main effect of trial block was not significant, $F(1, 1236) = .87$, ns, whereas the effect of the nationality × game condition × trial block interaction, $F(6, 1236) = 3.06, p < .01$, was significant. The increase in the positive effect on cooperation, of the choice to entrust was observed among the American participants but not among the Japanese (see Figure 3). The American participants cooperated 14.1 percentage points more in the fixed-partner PD/R than in the fixed-partner PD in the first trial block of Phase II (trial block 6); this difference increased to 39.9 percentage points in the last three trial blocks. Among the Japanese participants, however, the difference was 6.0 percentage points in the first trial block of Phase II, and only 11.9 percentage points during the last half of Phase II. These results provide strong support for Hypothesis 3.

Hypothesis 2

Hypothesis 2 predicts that American participants will exhibit a higher level of trusting behavior (will entrust more coins in an act of risk taking) than will Japanese participants in both the fixed-partner PD/R and the random-partner PD/R. As predicted, the American participants' average amount entrusted to others was higher than that of Japanese participants in both the fixed-part-

ner PD/R (8.92 coins versus 7.35 coins) and the random-partner PD/R (6.81 versus 5.06). The main effect of nationality in a nationality × game condition × trial block ANOVA of trusting behavior (the number of coins entrusted by the participants) was highly significant, $F(1, 210) = 18.43, p < .0001$. In this analysis, we used only the fixed-partner PD/R and the random-partner PD/R because no option for trusting behavior (choosing the level to invest) existed in the fixed-partner PD condition. The nationality x game condition interaction effect was not significant, $F(1, 210) = .07$, ns. The main effect of trial block, however, was significant, $F(6, 1260) = 9.87, p < .0001$. The nationality × trial block interaction effect was only marginally significant, $F(6, 1260) = 1.95, p < .08$. As demonstrated in Figure 5, the level of trusting behavior increased over time during Phase II, but this increase occurred primarily among the Americans.

These results clearly support Hypothesis 2. American participants exhibit trusting behavior at a higher level than do the Japanese, whether or not it is possible to build trust relationships with a particular partner. This finding indicates that the Americans' stronger inclination to take a risk to build trust and the Japanese participants' relative reluctance to take such risks do not reflect their differences in desire to build trust relationships. Rather, they seem to reflect general differences in their overall tendencies to avoid uncertainty, as we discussed earlier in this paper.

In addition to the significant effect of nationality, the ANOVA indicates a highly significant effect of game type, $F(1, 210) = 33.70, p < .0001$. Participants entrusted more coins when it was possible to build trust relationships with a particular partner (7.70 coins) than when building such relationships was not possible (5.98 coins). Furthermore, the significant game condition × trial block interaction effect, $F(6, 1260) = 15.89, p < .0001$, indicates (as anticipated) that participants engaged increasingly in trusting behavior over time in the fixed-partner PD/R more than in the random-partner PD/R. Investments in a partner (entrusting more coins) do not pay off in the absence of consecutive repeat play with the same partner.

Hypothesis 4

Hypothesis 4 states that the overall level of cooperation will be lower in the random-partner PD/R than in the fixed-partner PD/R. The main effect of game condition in a nationality × game condition (fixed-partner PD/R versus random-partner PD/R) × trial block ANOVA was highly significant, $F(1, 210) = 53.82, p < .0001$ ($F(1, 209) = 137.02, p < .001$, with control of cooperation in Phase I). As shown in Figure 3, the cooperation rate is much higher in the fixed-partner PD/R than in the random-partner PD/R. Furthermore, the game condition × trial block interaction effect was significant, $F(1, 1260) = 9.32, p < .0001$. This interaction effect shows that participants in the fixed-partner PD/R cooperated more over time than participants in the random-partner PD/R. As Figure 3 demonstrates, cooperation rates increased slowly across trial blocks in the fixed-partner PD/R, while they decreased across blocks in the random-partner PD/R. These results support Hypothesis 4.

Hypotheses 5 and 6

Hypothesis 5 concerns the comparison between the cooperation rates in Phase I and in Phase II in the random-partner PD/R condition. To test this hypothesis, we used the cooperation rates in Phase I and Phase II as a repeated measure in a nationality × phase (I versus II) ANOVA. The main effect of phase was not significant, $F(1, 86) = .12$, ns. The introduction of Phase II (PD/R with random partner) after trial block 5 seems to exert a positive effect on cooperation, as shown in Figure 3, but this positive effect is minor and short-lived. The cooperation rate in Phase II did not exceed the overall cooperation rate in Phase I. As a result, this finding does not support Hypothesis 5.

Hypothesis 6 states that allowing participants to choose the level of investment in Phase II of the random-partner PD/R condition will not affect cooperation rates differentially for American and Japanese participants. Neither the main effect of nationality, $F(1, 86) = .33$, ns, nor the nationality × phase interaction effect, $F(1, 86) = 1.32$, ns, was significant in this ANOVA. The lack of an interaction effect indicates that allowing the choice of levels of risk taking (or investment) does *not* exert differential effects on levels of cooperation for American and Japanese participants. Thus Hypothesis 6 is supported.
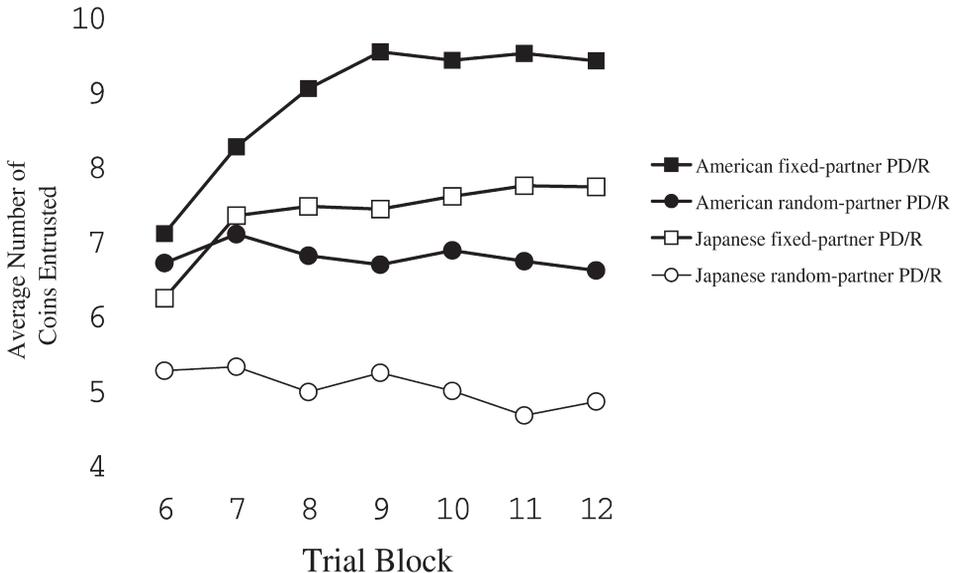
Figure 5. Average Number of Coins Entrusted Over Trial Blocks in Phase II, American and Japanese Participants

Cooperation Rates in the Fixed-Partner PD Condition

We do not offer a specific prediction about the cooperation rates in the fixed-partner PD condition with respect to nationality. The results reported in Figure 3 indicate that the cooperation rate in the fixed-partner PD condition, in which the participants could not determine the number of coins to entrust, was higher among Japanese than among American participants. On average, the cooperation rate was .66 (sd = .35) among Japanese participants but .58 (sd = .31) among Americans. The main effect of nationality in the nationality × trial block ANOVA was not significant $F(1, 82) = 1.15$, ns. The main effect of trial block was significant, $F(6, 492) = 3.86$, $p < .001$; so was the nationality × trial block interaction effect, $F(6, 492) = 3.16$, $p < .01$. These effects reflect the downward trend in cooperation rates over time among the Americans during Phase II. The Japanese cooperation rates, in contrast, stayed at about the same level throughout Phase II. Given that the cooperation rate was higher for the Japanese than for the American participants in Phase I, the Japanese participants seem slightly more willing to cooperate than do the Americans in the absence of the option to select the amount to entrust to others.

Initial Cooperators Versus Initial Defectors

In the introduction we asked whether initial cooperators or initial defectors take more risks to build trust when they are given a chance to do so. Initial cooperators are those who cooperated at a high level (higher than the median cooperation level for the participants of the same nationality and condition) in Phase I, in which they received no opportunity to choose the amount to entrust. Initial defectors are those who cooperated at a low level. In the nationality x game condition (fixed-partner PD/R versus random-partner PD/R) x initial level of cooperation (initial cooperators versus initial defectors) ANOVA of the average amount of money entrusted to a partner, the main effect of the initial level of cooperation was highly significant, $F(1, 206) = 14.79$, $p < .001$. The initial cooperators, more than the initial defectors, entrusted more money (7.78 versus 6.18).

In addition, the game condition x initial level of cooperation interaction was marginal, $F(1, 206) = 3.27$, $p < .08$, and the nationality x game condition x initial level of cooperation interaction was significant, $F(1, 206) = 5.63$, $p < .05$. The initial cooperators' willingness to entrust in comparison with the initial defectors' was more pronounced in the random-partner PD/R (6.96 vs. 4.99) than in the fixed-partner PD/R (8.33 vs. 7.05). This result, however, may have been caused by a ceiling effect. The average amount entrusted was close to 10, the highest possible level, in the fixed-partner PD/R among the initial cooperators. Similarly, the significant three-way interaction seems to be a result of the extremely high amount entrusted by the American participants in the fixed-partner PD/R. In general, in the fixed-partner PD/R involving American participants, including the initial defectors (9.11 coins) and the initial cooperators (8.70 coins), coins were entrusted at very high levels. In contrast, the initial Japanese cooperators entrusted more coins than did the initial Japanese defectors (8.25 vs. 6.37) in the fixed-partner PD/R. In the random-partner PD/R, both American and Japanese initial cooperators (8.15 and 5.77) entrusted more than the initial defectors (5.57 and 4.29).

The option to choose the amount to entrust helped initial defectors more than initial cooperators to achieve a higher level of cooperation over time in the fixed-partner condition but not in the random-partner condition. To analyze the effect of the option to entrust on cooperation, we used the difference in cooperation during Phase II and Phase I: how much the cooperation level improved because of the introduction of the option to entrust different amounts. The main effect of the initial level of cooperation in the nationality x game condition x initial level of cooperation ANOVA of the improvement in cooperation was highly significant, $F(1, 206) = 22.90$, $p < .0001$. The initial defectors' cooperation rate improved by .33, but that of the initial cooperators improved by only .18. The differential effect on cooperation of the option to entrust is not likely to be attributed to regression toward the mean because the differential effect existed only in the fixed-partner condition (.54 vs. .33) and not in the

random-partner condition (.04 vs. –.04). The game condition x initial level of cooperation interaction was significant, $F(1, 206) = 6.60$, $p < .01$. These results indicate that the positive effect of the option to take risks by entrusting different amounts (Hypothesis 1) is more pronounced for initial defectors than for initial cooperators. None of the interaction effects involving nationality and initial level of cooperation were significant.

DISCUSSION AND CONCLUSIONS

The results of this experiment are relatively straightforward. Five of our six hypotheses were clearly supported. Providing an opportunity to choose the level of risk involved in trusting another helped to improve mutual cooperation for both American and Japanese participants (Hypothesis 1). Furthermore, the American participants engaged in a higher level of risk taking to build trust than the Japanese (Hypothesis 2); as a result, they achieved relationships in which the exchange partners trusted each other and honored each other's trust (Hypothesis 3) in a cooperative fashion. These are the core hypotheses we addressed here.

The remaining three hypotheses compared the effects of the choice of level of risk taking on cooperation among fixed pairs of partners as compared with randomly matched partners. The positive effect, on cooperation, of allowing participants to choose the level of risk to take with their partner was found to be much weaker when it was not possible to build a relationship with a particular partner (in the random-partner PD/R condition) than when such a relationship was possible (in the fixed-partner PD/R condition; Hypothesis 4). American participants took more risks than the Japanese and trusted their partners more even in random partner exchanges (Hypothesis 2); this finding supports the general claim that the Japanese are inclined to avoid uncertainty. Even so, American participants were no better than the Japanese at raising the actual level of cooperation (Hypothesis 6).

Only one hypothesis failed to receive empirical support, namely our tentative proposition about the potential reduction in the second-order fear of exploitation by others (Hypothesis 5). We found some indication that allowing participants to signal their level of trust improves cooperation at least temporarily, as indicated by the surge in the cooperation rate at the beginning of Phase II in the PD/R with random-partner condition, but that effect is short-lived. Participants' willingness to take risks and trust their partners engenders greater mutual cooperation only when a trusting relationship can be established gradually with a specific partner.

The results of our experiment indicate that the American participants were more willing than the Japanese to take risks and to trust their partners. This greater willingness helped the Americans more than the Japanese to build trust relations when and only when they engaged continuously in exchanges with the same partners. Japanese participants in fact were more cooperative in the simple PD conditions—that is, in Phase I, in which they played a random-partner PD game, and in the fixed-partner PD condition, in which participants were not allowed to explicitly take risks in order to build trust relations with their partners over time. This difference was reversed in the PD/R game, when the participants were allowed to choose the level of risk to take with their partners so as to build trust.

The message of this study is clear and profound. Risk taking is a critical element in trust building for Americans, but less for the Japanese. Our results provide convincing support for the claim that trust is not the same as the lack of risk taking in social relations. Rather, trust can be built by initial risk taking. As shown by the results from the standard PD condition in our study, past research on trust, which failed to separate trusting behavior from acts of cooperation, was unable to capture the critical role of risk taking in building trust. In fact, in much of the earlier experimental research on trust, trusting and cooperation were confounded both theoretically and empirically. It is very important to distinguish trusting behavior from cooperation and to measure them separately if we are to study trust and trust building in relation to cooperation and to other socially

significant outcomes of trust. This is the most important implication of our findings.

We also discovered important implications of the U.S.-Japan differences in trust and cooperation. In the absence of opportunities to engage in risk taking to demonstrate willingness to trust a partner, American participants were less cooperative than their Japanese counterparts. With the addition of opportunity to take risks to demonstrate such willingness behaviorally, the levels of cooperation between the two groups of participants changed dramatically. Given the opportunity to signal willingness to trust a partner, the Americans not only were more willing than the Japanese to take risks in order to create trust relations with their partners; they also became more cooperative and built stronger trust relations.

The finding that Americans are more successful in building trust relations when risk taking is required is somehow counterintuitive. It contradicts views often expressed by naïve observers of the American and Japanese cultures, who expect more trust in what they consider to be a more collectively oriented society. Nevertheless, this finding is consistent with the conclusions derived from previous cross-societal research on trust. As we discussed in the introduction, the fundamental difference between trust and assurance as a means of dealing with social uncertainty and risk resides in the specific role of risk taking. Trusting initially requires a form of risk taking in which one allows a degree of vulnerability, whereas assurance is a product of risk avoidance or reliance on risk-reducing structural arrangements, such as the formation of commitments.

Finally, the PD/R game used in this study is not only a useful methodological tool for empirically investigating the process of trust building, as demonstrated here. It also provides a clear conceptual framework for analyzing the logic involved in trust building. By separating out the act of risk taking that makes trust relations possible and enhances cooperation in a dynamic context, the PD/R game allows us to analyze the transformation of opportunistic relations into trust relations mediated by the intricate interplay between acts of trust and mutual cooperation. The next step will involve constructing theoretical models that capture this complex interplay and testing them in a wider array of structural and cultural settings. In view of recent events, understanding the role of risk taking in building cross-cultural trust relations could not be more timely.

REFERENCES

Arneson, Richard J. 1982. "The Principle of Fairness and the Free-Rider Problem." *Ethics* 92:616–33.

Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic Books.

Berg, Joyce, John Dickhaut, and Kevin A. McCabe. 1995. "Trust, Reciprocity, and Social History." *Games and Economic Behavior* 10:122–42.

Blau, Peter M. 1964. *Exchange and Power in Social Life*. New York: Wiley.

Bolton, Gary E., Elena Katok, and Axel Ockenfels. 2003. "Cooperation Among Strangers With Limited Information About Reputation." Working paper, Smeal College of Business Administration, Penn State University, Philadelphia.

Buchan, Nancy R., Rachel T.A. Croson, and Robyn M. Dawes. 2002. "Swift Neighbors and Persistent Strangers: A Cross-Cultural Investigation of Trust and Reciprocity in Social Exchange." *American Journal of Sociology* 108:168–206.

Cook, Karen S., ed. 2001. *Trust in Society*. New York: Russell Sage Foundation.

Dasgupta, Partha. 1988. "Trust As a Commodity." Pp. 49–72 in *Trust: Making and Breaking Cooperative Relations*, edited by Diego Gambetta. Oxford: Blackwell.

Deutsch, M. 1973. *The Resolution of Conflict: Constructive and Destructive Processes*. London: Yale University Press.

Etzioni, Amitai. 1967. "A Nonconventional Use of Sociology As Illustrated by Peach Research." Pp. 806–38 in *The Uses of Sociology*, edited by Paul R. Lazarsfeld, William H. Sewell, and Harold L. Wilensky. New York: Basic Books.

————. 1969. "Social Psychological Aspects of International Relations." Pp. 538–601 in *Handbook of Social Psychology*, 2nd ed., vol. 5, edited by Gardner Lindzey and Elliot Aronson. Reading, MA: Addison-Wesley.

Farrell, Henry. 2004. "Trust, Distrust, and Power." Pp. 85–105 in *Distrust.* Edited by Russell Hardin. New York: Russell Sage Foundation.

Hardin, Russell. 2002. *Trust and Trustworthiness*. New York: Russell Sage Foundation.

Hayashi, Chikio, Tatsuzo Suzuki, and Masakatsu Murakami. 1982. *A Study of Japanese*

National Character, vol. 4, Tokyo: Idemitsushoten.

Hofstede, Geert H. 1980. *Culture's Consequences: International Differences in Work-Related Values*. Beverly Hills: Sage.

———. 1991. *Cultures and Organizations: Software of the Mind*. London: McGraw-Hill.

Holmes, John G. and John K. Rempel. 1989. "Trust in Close Relationships." Pp. 187–220 in *Close Relationships*, edited by Clyde Hendrick. Thousand Oaks, CA: Sage.

Hsee, Christopher K. and Elke U. Weber. 1999. "Cross-National Differences in Risk Preference and Lay Predictions." *Journal of Behavioral Decision Making* 12:165–79.

Kakiuchi, Riki and Toshio Yamagishi. 1997. "General Trust and the Dilemma of Variable Interdependency." *Japanese Journal of Social Psychology* 12:212–21.

Kollock, Peter. 1994. "The Emergence of Exchange Structures: An Experimental Study of Uncertainty, Commitment, and Trust." *American Journal of Sociology* 100:313–45.

Kreps, David M. 1990. "Corporate Culture and Economic Theory." Pp. 90–143 in *Perspectives on Positive Political Economy*, edited by James E. Alt and Kenneth A. Shepsle. Cambridge, UK: Cambridge University Press.

Levi, Margaret and Laura Stoker. 2000. "Political Trust and Trustworthiness." *Annual Review of Political Science* 3:475–508.

Lindskold, Svenn. 1978. "Trust Development, the GRIT Proposal, and the Effects of Conciliatory Acts on Conflict and Cooperation." *Psychological Bulletin* 85:772–93.

Matsuda, Masafumi and Toshio Yamagishi. 2001. "Trust and Cooperation: An Experimental Study of PD With Choice of Dependence." *Japanese Journal of Psychology* 72:413–21.

Meeker, Barbara F. 1983. "Cooperative Orientation, Trust, and Reciprocity." *Human Relations* 3:225–43.

Molm, Linda and Karen S. Cook. 1995. "Social Exchange and Exchange Networks." Pp. 209–35 in *Sociological Perspectives on Social Psychology*, edited by Karen S. Cook, Gary Alan Fine, and James S. House. Boston: Allyn and Bacon.

Molm, Linda, Gretchen Peterson, and Nobuyuki Takahashi. 2001. "The Value of Exchange." *Social Forces* 80:159–84.

Orbell, John and Robyn Dawes. 1993. "Social Welfare, Cooperators' Advantage, and the Option of Not Playing the Game." *American Sociological Review* 58:787–800.

Osgood, Charles E. 1962. *An Alternative to War or Surrender*. Urbana: University of Illinois Press.

Pilisuk, Marc and Paul Skolnick. 1968. "Inducing Trust: A Test of the Osgood Proposal." *Journal of Personality and Psychology* 8:121–33.

Pruitt, Dean G. and Melvin J. Kimmel. 1977. "Twenty Years of Experimental Gaming: Critique, Synthesis, and Suggestions for the Future." *Annual Review of Psychology* 28:363–92.

Sato, Kaori and Toshio Yamagishi. 1986. "Two Psychological Factors in the Problem of Public Goods." *Japanese Journal of Experimental Social Psychology* 26:89–95.

Snijders, Chris. 1996. *Trust and Commitments*. Utrecht: Interuniversity Center for Social Science Theory and Methodology.

Solomon, L. 1960. "The Influence of Some Types of Power Relationships and Game Strategies Upon the Development of Interpersonal Trust." *Journal of Abnormal Social Psychology* 61:223–30.

Triandis, Harry C. "The Contingency Model in Cross-National Perspectives." Pp. 167–88 in *Leadership Theory and Research Perspectives and Directions,* edited by Martin M. Chemers and Roya Ayman. San Diego: Academic Press.

Weber, Elke U., Christopher Hsee, and Joanna Sokolowska. 1998. "What Folklore Tells Us About Risk and Risk Taking: Cross-Cultural Comparisons of American, German, and Chinese Proverbs." *Organizational Behavior and Human Decision Processes* 75: 170–86.

Yamagishi, Toshio. 1986. "The Provision of a Sanctioning System As a Public Good." *Journal of Personality and Social Psychology* 51:110–16.

———. 1988. "The Provision of a Sanctioning System in the United States and Japan." *Social Psychology Quarterly* 51:32–42.

———. 1990. "Factors Mediating Residual Effects of Group Size in Social Dilemmas." *The Japanese Journal of Psychology* 61:162–69.

———. 1992. "Group Size and the Provision of a Sanctioning System in a Social Dilemma." Pp. 267–87 in *A Social Psychological Approach to Social Dilemmas*, edited by Wim B.G. Liebrand, David M. Messick, and Henk A.M. Wilke. Oxford: Pergamon.

———. 1998. *The Structure of Trust: The Evolutionary Game of Mind and Society*. Tokyo: University of Tokyo Press.

Yamagishi, Toshio, Karen S. Cook, and Motoki Watabe. 1998. "Uncertainty, Trust and Commitment Formation in the United States and Japan." *American Journal of Sociology* 104:165–94.

Yamagishi, Toshio and Riki Kakiuchi. 2000. "It Takes Venturing Into a Tiger's Cave to Steal a Baby Tiger: Experiments on the Development of Trust Relationships." Pp. 121–23 in *The Management of Durable Relations*, edited by Werner Raub and Jeroen Weesie. Amsterdam: Thela.

Yamagishi, Toshio and Kaori Sato. 1986. "Motivational Bases of the Public Goods Problem." *Journal of Personality and Social Psychology* 50:67–73.

Yamagishi, Toshio and Midori Yamagishi. 1994. "Trust and Commitment in the United States and Japan." *Motivation and Emotion* 18:129–66.

Karen S. Cook *is the Ray Lyman Wilbur Professor of Sociology and Senior Associate Dean of the Social Sciences at Stanford University. She is currently engaged in research on commitment and trust in social exchange relations and is the co-editor of the Trust Series for the Russell Sage Foundation.*

Toshio Yamagishi *is a professor of social psychology in the Department of Behavioral Science, Hokkaido University. His research interests include trust, social intelligence, and culture.*

Coye Cheshire *is a doctoral student in the Department of Sociology at Stanford University. His dissertation examines the emergence of generalized information exchange systems, such as those found on the Internet.*

Robin M. Cooper *is a lecturer in sociology at Stanford University, where she received her PhD in 2004. She wrote her dissertation on cultural and situational effects on personal identity. She has also published on cooperation and trust.*

Masafumi Matsuda *is a researcher at Nippon Telegraph and Telephone Corporation in the NTT Communication Science Laboratories. Currently he is designing e-commerce transaction and reputation systems.*

Rie Mashima *is a doctoral student in the Department of Behavioral Science, Hokkaido University. Her current research interests focus on generalized exchanges and social dilemmas.*