

Does Trust Beget Trustworthiness? Trust and Trustworthiness in Two Games and Two Cultures: A Research Note*

TOKO KIYONARI
McMaster University

TOSHIO YAMAGISHI
Hokkaido University

KAREN S. COOK
Stanford University

COYE CHESHIRE
University of California, Berkeley

An important unanswered question in the empirical literature on trust is whether trusting begets trustworthiness. In two experimental games, with Japanese and American participants, respectively, we compared trust and trustworthiness to provide an answer to this question. The trustee in the standard Trust Game knows that he or she is trusted, whereas the trustee in the Faith Game does not know whether or not this is the case. Except for this fact, the trustee faces the same choice in both situations. If the simple fact that one is trusted by someone else makes a person more trustworthy to the truster, then the trustee in the Trust Game should behave in a more trustworthy manner. Our results indicate that trust does not beget trustworthiness, at least in one-shot games. The results also indicate that trust and trustworthiness are two sides of the same coin but are quite distinct, partially replicating the recent findings of Buchan, Croson, and Dawes. American trusters were more trusting than their Japanese counterparts in the Trust Game, whereas American trustees were less trustworthy. The nationality difference in trust and trustworthiness is less pronounced in the Faith Game. We conclude that trust researchers should consider the limitations of one-shot games in studying the determinants of trust and trustworthiness.

Most scholars as well as ordinary citizens believe that trust is an important lubricant of social relations. Beyond the acknowledgment that trust is important in social and personal life, however, there is not much consensus on the specific nature of trust and its functions in society. A key distinction typically overlooked in the general literature on trust is the difference between the *trust* exhibited by a truster and the *trustworthiness* of a trustee. As Hardin (2002) points out, trust and trustworthiness frequently are confounded in many of the writings on trust. In fact, the term

trust is often used to refer to trustworthiness—a characteristic of the one who is trusted. For example, when people speak of a “decline in trust in American society,” they usually mean that Americans are perceived as less trustworthy now than at some time in the past. At the same time, the term *trust* is used as well to refer to trust or trustfulness; sometimes this is viewed as a psychological trait of the truster, not a characteristic or trait of the trustee. In addition, psychological scales used as indicators of trust (e.g., Rotter 1971) are supposed to measure individual differences in the degree or level of *expecting others to be trustworthy*.

The actual relationship between trust and trustworthiness has escaped the empirical attention of many trust researchers because of this tendency to confound these two concepts in theory as well as in research. In this study we focus on the distinction

*The research reported in this paper was supported by a grant from the Japan Society for the Promotion of Science and by the Russell Sage Foundation Trust Project, New York City. Address correspondence to Toko Kiyonari, Department of Psychology, Neuroscience & Behaviour, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4K1, tokoki@aol.com

between trust and trustworthiness, and we examine whether the fact that a person is trusted makes him or her behave in a more trustworthy manner. In short, does the act of trust beget trustworthiness?

Recent experimental studies provide evidence that under certain conditions people “trust” or take a risk on anonymous others. Those who are trusted in this way honor the trust bestowed on them even in one-shot games, in which such trusting and trustworthy behavior cannot produce future benefits (Berg, Dickhaut, and McCabe 1995; Buchan, Croson, and Dawes 2002; Cox 2001, 2002; Kiyonari and Yamagishi 1999; McCabe and Smith 2000; Wang and Yamagishi 2005).

A typical experimental paradigm for studying trust is the Trust Game (Kreps 1990). This game is played by two parties, a *truster* (often called a *proposer*) and a *trustee* (also known as a *responder*). The truster is given a choice between trusting (T) and not trusting (NT). If she chooses not to trust, she receives a fixed amount of money, say \$10, directly from the experimenter. The trustee also receives \$10 if the truster chooses not to trust. The choice of trust, on the other hand, transfers the power to choose the outcome to the trustee. In this case, the trustee is offered a choice between honoring (H) and not honoring (NH) the trust placed in her. The personal gain involved in choosing NH is greater for the trustee than that in choosing H. For example, the trustee receives \$30 when she chooses NH and only \$20 when she chooses H. The truster receives nothing when the trustee chooses NH, and \$20 when she chooses H.

The standard game-theoretic logic of backward induction predicts that the trustee will choose NH because it provides more money, and that the truster, knowing that not honoring trust (NH) is the rational choice for the trustee, will choose not to trust (NT). This means that the truster is expected not to trust the trustee and the trustee is expected not to honor the trust placed in her.

Despite these predictions, however, the results of experimental studies using the Trust Game and related experiments such as the Investment Game (Berg et al. 1995) or the Faith Game (Kiyonari and Yamagishi 1999) have revealed consistently that a sub-

stantial proportion of trusters choose to trust, and that a substantial proportion of the trustees, when trusted, honor the trust placed in them by behaving in a trustworthy manner.

DOES TRUST BEGET TRUSTWORTHINESS?

The Trust Game described briefly above seems to provide clear-cut behavioral evidence of trust and its relationship to trustworthiness. On closer scrutiny, however, it is not clear what the Trust Game actually measures, for two reasons. First, the trustee who chooses to honor trust (H) may be responding to (or reciprocating) the trust placed in her, or simply may be behaving in a fair or altruistic manner. That is, the trustee’s choice of H may be the reciprocation of trust, or it may be an expression of her concerns for altruism or fairness. Similarly, the choice of T (trust) by the truster may be based on the expectation of reciprocation of her trust or on the expectation of the altruistic and fairness concerns of the trustee. Second, it is also possible that the choice of T by the truster may simply be based on her own altruistic motives: that is, she may be motivated to give the trustee an opportunity to earn more money. Sorting out what is at stake is important for understanding the distinct roles of trust and trustworthiness in social relations.

In an attempt to determine whether being trusted by someone makes the trusted more trustworthy, researchers have tried to discern the motivational bases of the truster’s and trustee’s choices in the Trust Game. Berg et al. (1995), for example, examined the correlation between the amount the truster entrusted to the trustee and the proportion of that amount the trustee sent back to the truster, using a variant of the Trust Game called the “Investment Game.”¹ The propor-

¹ In the Trust Game, the choices of the truster and the trustee are binary. In the Investment Game, these choices are continuous. The truster decides how much of his endowment of X dollars to entrust to the trustee. The amount of money entrusted (say, Y dollars, where $0 \leq Y \leq X$) is tripled and transferred to the trustee. The trustee receives both his own endowment of X dollars and the money entrusted to him by the truster (i.e., 3Y dollars). The trustee then is asked to send any amount, Z dollars, to the truster (where $0 \leq Z \leq X+3Y$).

tion of the endowment entrusted to the trustee is a behavioral measure of how much the truster trusts the recipient. Similarly, the proportion of the money the trustee sends back to the truster is assumed to be a measure of the trustee's trustworthiness.

Berg et al. (1995) examined whether the two ratios (or proportions) were correlated. They reasoned that if the correlation is zero, the trustee's decision is not affected at all by how much he is trusted by his partner. If the trustee reciprocates the trust bestowed on him, the two ratios should show a positive correlation. Berg and his colleagues did not find a significant correlation between these ratios, however.²

In another study with a similar design, Snijders and Keren (1999) also failed to find a positive correlation. Dufwenberg and Gneezy (2000), using a slightly different methodology, obtained the same result. All of these studies produced a negative answer to the question "Does the mere fact of being trusted make one behave in a trustworthy manner toward the truster?" The correlations between how strongly the truster trusts the trustee and the trustee's willingness to reward the truster also do not indicate a significant role for reciprocity in trust relations, as examined in these experimental settings.³

² In a follow-up study, Berg and his colleagues presented the result of this study and asked their participants to play the same game. In that study, they found a positive correlation ($r = .34$).

³ A different correlation—between how much the trustee thinks the truster expects to receive and how much the trustee actually gives to the truster—may indicate more strongly the operation of reciprocity in trust. This expectation and the actual behavior were found by Dufwenberg and Gneezy (2000) to be correlated positively. Their study indicates that the trustees do not reciprocate the trust placed in them, but do respond instead to the trusters' perceived expectations. This correlation of expectation with behavior may be the result of "projection" on the part of the trustees. The trustees who are more altruistic and more concerned with fairness think others have similar concerns and thus expect trusters to behave accordingly. On the other hand, this correlation may indicate that the trustee reciprocates trust only when that behavior is interpreted as an expression of the truster's actual level of trust. Therefore this act of perceived trust motivates the trustee to behave in a trustworthy manner only when she thinks she is being trusted because the truster expects her to be trustworthy.

In an alternative strategy employed to tease out the specific effect of reciprocity on trust, results from the Trust Game are compared with results from the Dictator Game (Camerer and Thaler 1995). The trustee in the Trust Game and the "dictator" in the Dictator Game face the same behavioral choice: The truster must divide a fixed endowment of Z dollars between himself and another player (i.e., the truster) in any way he prefers. The difference between the two games lies in how the endowment is provided to the trustee and to the dictator.

To make the contrast clear, let us use a concrete example in which Z equals \$30. In the Trust Game, the truster has a choice between T (letting the trustee divide \$30 between the two players) and NT (giving each \$10). The trustee thus is given a chance to earn more than \$10, based on the choice made by the truster. In contrast, the dictator in the Dictator Game receives directly from the experimenter this chance to earn more than \$10; the dictator owes nothing to the other party (the recipient).

Exactly that the recipient in the Trust Game owes the truster the chance to receive extra money while the dictator in the Dictator Game owes nothing to the recipient, the two players—the trustee and the dictator—face exactly the same choice of freely dividing a fixed sum of money (\$30) between themselves and another party (the "truster" or the "recipient").⁴ Thus the difference in the amount the trustee or the dictator gives to the truster (the recipient) should reflect the effect of reciprocity on the degree of trust or altruism exhibited by the truster. Dufwenberg and Gneezy (2000) made this specific comparison of games; they found no significant difference in the amount given to the other party by the trustee in the Trust Game and by the dictator in the Dictator Game.

Cox (2001) used a similar logic when he compared the following conditions. Condition A is the standard Investment Game, in which the truster decides how much of her endowment to entrust, and the trustee

⁴ Because the recipient in the Dictator Game has no choice between trusting or not trusting, we call him or her a "recipient" rather than a "truster."

decides how much of the money entrusted to her to return to the truster. Condition C is equivalent to the Dictator Game, in which a “dictator” is given some amount of money directly by the experimenter. The recipient in condition C has no choice of how much to entrust. The amount that the experimenter gives to the recipient in condition C is three times the amount a matched truster in condition A actually entrusts to her partner. The trustee in condition A knows that she has been given a chance to improve her earnings through the truster’s choice, whereas the dictator in condition C knows that she owes nothing to the recipient. In addition to the method of providing the endowment, the actual choice given to the trustee in the Trust Game is the same as that provided to the dictator in the Dictator Game in this experiment. Cox (2001) found that responders in condition A gave more to the truster than the dictator gave in condition C; this finding implies that *the fact that one is trusted makes the trustee reciprocate by behaving in a trustworthy manner*. In a subsequent study, Cox (2002) replicated this result among males, but not females, using the same procedures.

Empirical findings are mixed as to whether the simple fact that one is trusted by someone makes one behave in a more trustworthy manner. We do not yet have a clear, unambiguous answer to this fundamental question. On the one hand, how much a truster entrusts to a trustee is not correlated with how much is returned to the truster (Berg et al. 1995; Dufwenberg and Gneezy 2000; Snijders and Keren 1999). On the other, the positive correlation between expectations and actual behavior (Dufwenberg and Gneezy 2000) may or may not suggest behavioral reciprocity of trust. The results of studies comparing the truster’s behavior in the Trust Game with the dictator’s behavior in the Dictator Game also are mixed: the truster was shown to behave in a more trustworthy manner than the dictator in Cox (2001) and among male trustees in Cox (2002). Yet no significant difference was found between the trustee’s behavior in the Trust Game and the dictator’s behavior in the Dictator Game in the studies conducted by Dufwenberg and Gneezy (2000) and Snijders and Keren (1999), or among female trustees in Cox’s

study (2002). How can these findings be interpreted more generally?

DO TRUSTERS EXPECT THEIR TRUST TO BEGET TRUSTWORTHINESS?

Whether expectations of reciprocity play a role in the truster’s decision has received even less empirical attention in part because the recipient in the Dictator Game has no choice. Thus a comparison of the Trust Game with the Dictator Game cannot be used to address this question. Cox (2001, 2002) used an ingenious design to examine the importance of expectations of reciprocity in the truster’s decision. He added another condition; condition B, to his experiments, in which the truster is given the same choice as in condition A. The trustee, however, does not have the option of sending money back to the truster. Thus the truster in condition B cannot expect a “return” on his trust, and the expectation of reciprocity therefore cannot play a role in the truster’s decision. In this condition, any amount a truster transfers to his trustee must be based on altruism.

Cox (2001) found an interesting difference between conditions A and B. Trusters in condition A transferred more of their endowment to their trustees than did trusters in condition B. The same effect was not replicated in Cox (2002), however.

Unfortunately, in addition to the inconsistency between the two studies, we find that Cox’s ingenious design cannot provide clear evidence on the role of expectations of reciprocal trust because of a significant methodological problem. The difficulty is that the actual amount the truster in condition B transfers to the trustee increases the inequality of earnings between the two players. Thus the truster’s unwillingness to transfer some of his endowment to his trustee can be attributed to inequity aversion (Fehr and Fischbacher 2003, 2004; Fehr and Gächter 2000). If this is the case, the difference between conditions A and B, if it exists at all, may be caused by avoidance of inequity in condition B as well as by expectations of reciprocity in condition A. We explore a potential solution to this problem.

THE TRUST GAME AND THE FAITH GAME

The primary goal of our study is to examine whether or not the expectation of reciprocity plays a role in a truster's decision to transfer some of his or her endowment to a trustee, and whether or not reciprocity plays a role in the trustee's decision to send money back to the truster. To answer these questions, we compare the Faith Game (Kiyonari and Yamagishi 1999) with the Trust Game. This comparison allows us to answer both of these questions at the same time and to eliminate some of the possible confounding factors in prior research.

The Trust Game in our study involves a binary choice by the truster (T and NT), as in the standard Trust Game, and a continuous quantitative choice by the trustee, as in the Investment Game. The truster in the Trust Game decides whether to receive \$10 directly from the experimenter (NT) or to let the trustee divide \$30 between the two (T). If the truster chooses NT, the trustee also receives \$10. The trustee thus receives a chance to earn more than \$10 when the truster chooses to trust (T). The trustee in the Faith Game is offered the same decision to divide \$30 *regardless of the truster's choice*. The trustee in this case in fact is a dictator, as in the Dictator Game.

The Faith Game differs from the Dictator Game only in the choice provided to the truster. Remember that the recipient in the Dictator Game has no choice: she is at the mercy of the dictator. The truster in the Faith Game is informed that the trustee is playing a dictator's role. Yet the truster in the Faith Game differs from the recipient in the Dictator Game in that she can choose between taking the sure \$10 amount and receiving whatever amount the trustee gives her from the \$30 amount. In this sense, the truster's role in the Faith Game is similar to that in the Trust Game. The trustee in the Faith Game, however, is *not* informed that the truster has this choice, and the truster in the Faith Game knows that the trustee is unaware of the truster's choice. From the trustee's viewpoint, then, the Faith Game is identical to a Dictator Game. In this way the comparison of these two games is equivalent

to the comparison between the Trust Game and the Dictator Game discussed earlier.

At the same time, the comparison between these two games can answer the question of whether expectations of reciprocity play a role in the level of trust displayed by the truster. The truster is expected to choose T instead of NT when she expects that the trustee will give her more than \$10 in either condition (or when the truster is motivated by altruism and is willing to reward the trustee at some cost to herself). In the Faith Game, the only logical reason for the truster to expect the trustee to give her more than \$10 is if the trustee is expected to be altruistic or concerned with fairness. In the Trust Game, however, the truster has an additional reason to expect the trustee to reciprocate the trust she bestows on the trustee. Thus any difference in the frequencies of trust choices between trusters in the Faith Game and those in the Trust Game should be attributable to this additional reason: the expectation that the trustee will reciprocate the truster's trust.⁵ By comparing the Trust Game with the Faith Game we can simultaneously test the effects of reciprocity on the trustee's trustworthy behavior and of the expectations of reciprocity on the truster's trusting behavior.

CROSS-CULTURAL DIFFERENCES IN TRUST AND TRUSTWORTHINESS

The second purpose of our study is to replicate the cross-cultural difference in trust and trustworthiness that Buchan et al. (2002) found between American and Japanese participants. Buchan and her colleagues conducted an Investment Game in the United States, Japan, China, and Korea, and found that trust (the truster's choice) and trustworthiness (the trustee's choice) are not correlated highly; they may be orthogonal under some conditions. That is, countries high on trust (the United States and China) are not

⁵ One additional difference is the existence of a focal point of \$10 in the Trust Game. The trustee may use \$10 as such a point, and may return at least \$10 to the truster. The trustee may expect this to happen. We decided to treat the effect of these expectations as part of the reciprocation of the truster's trust ("She will give me back at least \$10 since I'm giving her an opportunity to earn more than \$10").

necessarily high on trustworthiness (Korea and China), and vice versa. Specifically, Buchan et al. (2002) found that Americans are high on trust but low on trustworthiness, Japanese are low on both, Chinese are high on both, and Koreans are low on trust and high on trustworthiness.

Wang and Yamagishi (2005) and Kiyonari, Foddy, and Yamagishi (2004) used the same design including the Trust Game and the Faith Game to measure the truster's levels of trust among Japanese, Chinese, and Australian participants. (They did not measure the trustee's trustworthiness.) Their findings show that the cross-cultural difference in the truster's behavior varies with the game. In the Trust Game (similar to the Investment Game), Chinese are the most trusting group (42.4% trusting), followed by the Japanese (35.4%). The Australians are the least trusting group (32.9%). This order is reversed completely in the Faith Game: here Australians are the most trusting group (54.8%), followed by the Japanese (50%), while the Chinese are the least trusting

(40.3%). In another study comparing American with Japanese trusters and trustees, Kiyonari and Yamagishi (1999) used the Faith Game and the Trust Game. Their results regarding the truster's behavioral choice are consistent with those of Buchan and her colleagues: Americans are more trusting than Japanese. On the other hand, they found that Americans and Japanese do not differ in the level of trustworthiness demonstrated by the trustee. These cross-cultural studies of trust and trustworthiness (Buchan et al. 2002; Kiyonari et al. 2004; Kiyonari and Yamagishi 1999; Wang and Yamagishi 2005;) point to the need to use multiple games in comparing trust with trustworthiness cross-culturally because different psychological bases for trust and trustworthiness associated with different games may be activated in various cultures.

Thus we compare American with Japanese trusters' choices as well as trustees' choices in the Trust Game, and examine whether the findings of Buchan and her colleagues concerning the U.S.-Japan comparison are replicated. We also examine whether

the U.S.-Japan differences in the Trust Game can be replicated in the Faith Game.

GENDER DIFFERENCES IN TRUST AND TRUSTWORTHINESS

The third purpose of the research reported here is to explore gender differences in the role of reciprocity in trust and trustworthiness. Previous findings have been mixed. Cox (2002) found that male trustees reciprocate trust, while female trustees do not. In line with this finding, Wang and Yamagishi (2005) found, in an experiment with Chinese participants, that only male trusters expected reciprocation of their trust from the trustee. In their experiment, female trusters displayed expectations of "reverse reciprocity": they trusted the trustee more in the Faith Game than in the Trust Game. Wang and Yamagishi (2005) explain this gender difference in terms of a gender-based difference in preference for risk: women are afraid that a show of their willingness to trust someone will invite the trustee to exploit them, whereas the same act of trust leads men to expect reciprocal behavior from the trustee.

The pattern observed by Wang and Yamagishi (2005), including the expectations of reverse reciprocity, is replicated by Kiyonari et al. (2004) with Japanese and Australian participants. Unfortunately, both Wang and Yamagishi (2005) and Kiyonari et al. (2004) examined only the truster's behavior in the Faith Game and the Trust Game. Their experiments involved no real trustee.

If Wang and Yamagishi's interpretation of the gender difference is valid, we should find no sex difference in the trustee's behavior in the Faith Game because that trustee is a dictator who has complete control of his or her own fate as well as the truster's. Thus a female trustee does not need to fear her partner. We investigate whether the same sex difference observed by Wang and Yamagishi (2005) and by Kiyonari and her colleagues (2004) actually exists among trustees.

METHOD

We conducted the experiment in Japan and in the United States, with 134 Japanese and 128 American participants. Fifty-two of the Japanese participants were assigned the

role of trustee and 82 the role of truster. In the United States, 56 participants were given the role of trustee, and 72 the role of truster.

The Faith Game and the Trust Game

Participants were assigned randomly the role of truster or of trustee. We call the dictator in the Faith Game “trustee” to avoid the need to refer to the player who divides the endowment by different names—*trustee* or *dictator*—depending on the game. In each game, the trustees divided \$30 in any way they preferred. The truster chose either the sure \$10 (provided by the experimenter) or whatever amount the trustee was to give him or her from a \$30 endowment. Details of the two game conditions are provided below.

The Faith Game is played by two participants who have not met before. This is a one-shot game; therefore each player makes a decision just once. The participants’ decisions are completely anonymous. They do not meet in person before, during, or after the experiment. Their decisions are concealed in envelopes; thus each player’s decisions are anonymous not only to the other player, but also to the experimenter who conducts their study. A different experimenter who sits in the control room and who never sees the participants is the only one who has the chance to examine their decisions.

The participant who is assigned the role of “dictator” or “trustee” (simply called “player B” in the instructions) is provided by the experimenter with \$30 (2,400 yen in Japan) and is asked to divide the money in any way he or she prefers between himself or herself and the other participant, the “truster” (called “player A” in the instructions). The trustee receives the opportunity to freely divide \$30 regardless of the truster’s choice.

The participant who has been assigned the role of truster is first instructed about the role of the trustee, and is offered a choice between receiving whatever amount the trustee might allocate to him or her (T) and receiving \$10 directly from the experimenter (NT). When the truster makes this

choice, he or she is told that the trustee does not know that the truster has the safe option of choosing \$10. Furthermore, the trustee receives whatever amount he or she allocates to self regardless of the truster’s choice. That is, the truster’s choice does not affect the trustee’s earnings; it affects only the truster’s own earnings. The truster makes this choice before learning how much the trustee has allocated to him or her.

The Trust Game, like the Faith Game, is played by two participants and also is a one-shot encounter. As in the Faith Game, the truster makes a binary choice between T and NT. The most important difference between the two games is that the trustee is provided with an opportunity to freely divide \$30 *only* when the truster chooses T. The trustee knows about this choice by the truster and is aware that he or she is provided with a chance to divide \$30 when and only when the truster chooses to trust (T). When the truster chooses NT, both the truster and the trustee receive \$10 each. All of the participants who had been assigned the role of trustee were told that the truster in their pair had chosen T. Thus all of the trustees believed that their partner (the truster) had given them a chance to divide \$30, even when the truster in fact had chosen NT.

As discussed earlier, different motivational bases in the two games affect trust and trustworthiness. The truster’s choice of T in the Faith Game is based on the expectation of altruism or fairness on the trustee’s part: the truster cannot logically expect reciprocity from the trustee for choosing T instead of NT. The truster’s choice of T in the Trust Game, on the other hand, can be based on the expectation of reciprocity in addition to the expectation of altruism or fairness on the part of the trustee. In the same vein, how much the trustee gives to the truster should be based only on his or her altruism or fairness concerns in the Faith Game, because the trustee owes nothing to the truster. The trustee’s behavior in the Trust Game also can be based on his or her motivation to reciprocate trust and the favor bestowed on him or her by the truster, in addition to his or her concerns about altru-

ism or fairness. Thus the greater trust and trustworthiness expected in the Trust Game represents the added effect of reciprocity.

Design of the Experiment and Experimental Procedures

We used a 2 (game structure: Trust Game vs. Faith Game) \times 2 (participants' nationality: American vs. Japanese) \times 2 (gender: male vs. female) factorial design. All three factors are between-subjects factors. The participants were recruited from their respective student populations (Hokkaido University in Japan and Stanford University in the United States) with the prospect of earning money. No class credit was offered to the participants.

Procedures. Four to six students were scheduled for each experimental session. Upon arrival at the laboratory, each participant immediately was led to his or her room without meeting the other participants. The participant stayed in the room throughout the experiment. To secure anonymity with the experimenter who met him or her in person, each participant was provided, upon arrival, with a card carrying an ID number. Participants each picked a card from a box and kept it to themselves. They were instructed, before picking up the card, not to show the ID number to the experimenter.

The experiment was conducted with a set of envelopes delivered to the participants in sequence. When an envelope was delivered, the participant opened it and entered his or her choice (when necessary) on a decision sheet, placed the instructions and the decision sheet inside the envelope, sealed it, and placed it in a box outside his or her room. The experimenter picked up the envelope and delivered the next envelope in the sequence. Another experimenter, who sat in the control room, opened the envelope. This process was repeated several times, as described below.

Envelope 1. When all the scheduled participants arrived, each participant received the first envelope, which contained instructions describing the experiment, the rules of the game, and the role to which the participant had been assigned. No cover story

about the reason for the study was used: participants were told that the purpose was to investigate how people divided money.

Envelope 2. When all the participants finished reading the instructions, they received the second envelope. The contents varied with the game condition and the role assigned. Trusters in the Faith Game were asked whether they wanted to receive whatever amount the trustee gave them or to take the sure \$10 from the experimenter. Trusters in the Trust Game chose whether they wanted to let the trustee divide \$30 (2,400 yen in Japan) or to take the sure \$10 from the experimenter. Trustees in the Faith Game were asked to divide \$30 between themselves and the truster in any way they liked. Trustees in the Trust Game were informed that the truster was making a decision between letting him or her divide \$30 and taking the sure \$10, and then were asked to predict which choice the truster would make. These choices (or their predictions) were written on either a decision sheet or a prediction sheet, whichever was applicable. The participants placed their decision (or prediction) sheets in their envelopes, and the experimenter then collected the envelopes.

Envelope 3. In a few minutes, the third envelope was delivered to each participant. Trustees in the Trust Game each were informed that the truster had decided to let them divide \$30 instead of taking the sure \$10. Then they were asked to divide the money between the two participants. Trusters in the Faith Game were asked to predict how much the trustee would give them. Trusters in the Trust Game were asked how much the truster would give them (if they had chosen to let the trustee divide \$30) or how much the trustee would give them (if they had chosen the sure \$10) if they chose to let the trustee divide \$30. Trustees in the Faith Game were asked how much they thought the truster expected to receive from them. Then the allocation decisions (trusters in the Trust Game) or the expectations (other conditions) were collected.

Envelope 4. The fourth (and last) envelope contained the postexperimental ques-

tionnaire. When a participant filled out this questionnaire and placed it in the fourth envelope, the envelope was collected. Then the participant was paid⁶ and debriefed.

RESULTS

Does Trust Beget Trustworthiness?

Does the simple fact of being trusted by someone make the trusted behave in a trustworthy manner toward the truster? The results of our experiment provide a negative answer to this question, consistent with most of the previous studies (Berg et al. 1995; Buchan et al. 2002; Dufwenberg and Gneezy 2000). The effect of reciprocity should be represented by the difference in the amounts that trustees give to their partners in both the Trust Game and the Faith Game. In fact, however, the difference in the amount (i.e., the proportion of the endowment of \$30) that the trustee gave to the truster was larger in the Faith Game than in the Trust Game—a pattern opposite the predicted effect of reciprocity. Trustees in the Trust Game, on average, gave 33 percent ($sd = 19.6$) of the \$30 (or 2,400 yen) to the truster, whereas they gave 39.3 percent ($sd = 16.4$) in the Faith Game. The main effect of game (Faith Game vs. Trust Game) in the game structure \times participant's nationality \times participant's gender ANOVA was marginally significant; $F(1, 100) = 3.57, p < .07$. The reverse reciprocity previously observed among trusters by Wang and Yamagishi (2005) and by Kiyonari et al. (2004) was found as well in this experiment. It was demonstrated in the trustee's behavior rather than the truster's.

The mean proportions of \$30 provided by the trustee shown in Table 1 indicate that reverse reciprocity exists only among the American trustees. Japanese trustees gave about the same amount of money to the truster in the Trust Game (38.5%, $sd = 17.6$) as in the Faith Game (39.0%, $sd = 16.0$). In contrast, American trustees gave much less to

the truster in the Trust Game (27.3%, $sd = 20.2$) than in the Faith Game (39.5%, $sd = 17.0$). The difference is significant; $t(54) = 2.46, p < .05$. A nationality \times game structure \times gender ANOVA shows that the nationality \times game structure interaction effect ($F(1, 100) = 2.81, p < .10$) is marginally significant, as is the main effect of gender ($F(1, 100) = 3.38, p < .07$). No other effects were statistically significant.

Table 2 shows the proportion of the trustees who gave their partners a fair amount—that is, half or more of their endowment. The difference in proportions between trustees in the Faith Game and those in the Trust Game is more pronounced in Table 2 than in Table 1, especially among American participants. The proportion of fair trustees in the Japanese sample is slightly larger in the Trust Game (53.6%) than in the Faith Game (45.8%), although the difference is not significant; $chi-square(1) = .31, ns$. In contrast, the proportion of fair trustees among the Americans is much smaller in the Trust Game (22.2%) than in the Faith Game (62.1%), and the difference is significant; $chi-square(1) = 9.07, p < .01$. In a logistic regression analysis of the proportion of fair trustees, the nationality \times game structure interaction is significant; $chi-square(1) = 6.43, p < .01$. Other effects including gender are not statistically significant.

We draw two overall conclusions from these analyses: there is no difference between the two games among Japanese trustees, and reverse reciprocity exists among American trustees. A similar reverse reciprocity was observed in earlier studies by Wang and Yamagishi (2005) and by Kiyonari and her colleagues (2004), but their studies involved only trusters. The current experiment is the first to show reverse reciprocity among trustees.

The reverse reciprocity among American participants also seems to be reflected in the correlation between the amount the trustee gave to the truster and the trustee's estimation of the truster's expectation concerning that amount. The correlation is negative ($r = -.34, p < .08$) among American trusters in the Trust Game, while the negative correlation is much smaller ($r = -.13, ns$) among those in the Faith Game. This finding suggests reverse

⁶ To avoid the situation in which some trusters who chose T earned nothing, we paid those who chose T \$15 (1,200 yen in Japan), or half of \$30. Those who chose the sure \$10 received \$10 (or 800 yen in Japan). Trustees were paid the amount they allocated to themselves.

reciprocity in the sense that American responders actually gave less when they thought their partner was expecting more. Such a negative correlation was not observed among Japanese trustees: instead, Japanese trustees gave more when they thought their partner was expecting more ($r = .32, p < .10$). This correlation is positive and much stronger in the Faith Game ($r = .71, p < .001$).

Do Trusters Expect Their Trust to Be Reciprocated?

As shown in Table 3, the proportion of trust choices by the truster is slightly higher in the Trust Game than in the Faith Game, especially among the American participants. Among Japanese trusters, the trust choice is 59.0 percent in the Trust Game and 60.5 percent in the Faith Game; among American trusters, 72.7 percent in the Trust Game and 64.1 percent in the Faith Game. The difference is not significant among either Japanese or American trusters. In a logistic regression analysis, neither the main effect of the game structure, $chi-square(1) = .30, ns$, nor the nationality \times game interaction, $chi-square(1) = .49, ns$, is statistically significant. The trust choice in the Trust Game is higher among American participants (72.7%) than among Japanese participants (59.0%), in keeping

with the finding by Buchan and her colleagues (2002). The difference is not significant, however; $chi-square(1) = 1.49, ns$. No other effects in the logistic regression are statistically significant.

Are Japanese Less Trusting Than Americans?

Buchan and her colleagues (2002) found that American trusters are more trusting than their Japanese counterparts, while the two groups are equivalent in trustees' trustworthiness. Kiyonari and Yamagishi (1999) found a similar pattern. Our results are largely consistent with previous findings concerning the U.S.-Japan comparison in the trusters' behavior. In the Trust Game, 72.7 percent of the Americans made the trust choice and 59 percent of their Japanese counterparts made the trust choice, although the difference is not significant. On the other hand, American trustees were found to be less trustworthy than their Japanese counterparts: American responders gave only 27.3 percent of their endowment to their partners, whereas Japanese responders gave 38.5 percent. This difference is significant; $t(53) = 2.21, p < .05$. Overall, American participants in the Trust Game tended to be more trusting and less trustworthy than their Japanese counterparts.

Table 1. Average Proportions of the Endowment Given by the Trustee to the Truster in the Trust Game and in the Faith Game

	Japanese		American	
	Males	Females	Males	Females
Trust Game	.35 ($n = 13$)	.41 ($n = 15$)	.23 ($n = 16$)	.33 ($n = 11$)
Faith Game	.39 ($n = 11$)	.38 ($n = 13$)	.35 ($n = 17$)	.46 ($n = 12$)

Table 2. Proportions of the Trustees Who Gave Half or More of the Endowment to Their Partner in the Trust Game and in the Faith Game

	Japanese		American	
	Males	Females	Males	Females
Trust Game	.46 ($n = 13$)	.60 ($n = 15$)	.19 ($n = 16$)	.27 ($n = 11$)
Faith Game	.45 ($n = 11$)	.46 ($n = 13$)	.53 ($n = 17$)	.75 ($n = 12$)

Table 3. Proportions of Trust Choices Made by the Truster in the Trust Game and in the Faith Game

	Japanese		American	
	Males	Females	Males	Females
Trust Game	.53 ($n = 19$)	.65 ($n = 20$)	.77 ($n = 13$)	.70 ($n = 20$)
Faith Game	.63 ($n = 24$)	.58 ($n = 19$)	.67 ($n = 18$)	.62 ($n = 21$)

The nationality difference is less pronounced in the Faith Game. The proportion of the trustees in this game who chose the trust option was about the same among American (64.1%) and Japanese (60.5%) participants. American trustees gave 39.5 percent of their endowment to trusters in the Faith Game; their Japanese counterparts gave 39 percent.

Difference by Gender

Gender does not exert a statistically significant main effect on the truster's choice, nor a significant interaction effect with the game structure or the participant's nationality. Wang and Yamagishi's (2005) finding that males are more trusting than females in the Trust Game and less trusting in the Faith Game was not replicated. Females are more trustworthy than males in both games, though the effect of gender on the trustee's behavior is only marginally significant. Furthermore, gender does not interact with nationality or with game structure. Finally, gender has no effect on the trustee's choice.

DISCUSSION AND CONCLUSIONS

Two general conclusions can be drawn from the results of our experiment. First, trust is not a simple reflection of trustworthiness. This is especially apparent among American participants in our study, who are rather high in trust and fairly low in trustworthiness. Furthermore, American participants gave less when their partner gave them the option to divide the money (as in the Trust Game) than when the participants believed that dividing the money was not decided by their partner (as in the Faith Game). As we discuss below, this reverse reciprocity indicates that the one-shot act of trusting someone (and thereby taking a risk) does not necessarily lead the other person to act in a trustworthy manner on a single decision.

Second, subtle differences in the game structure can exert a large effect on choices. We started with the assumption that the trustee in the Trust Game has more motivation for choosing to trust than does the dictator in the Dictator Game: for the latter, altruism and fairness concerns are the *only* motivations for behaving fairly toward the

truster. This assumption of "cumulative" motivational effects is rejected on the basis of the American trustees' behavior: they behaved more fairly in the Faith Game than in the Trust Game. This pattern cannot be explained by the assumption of "cumulative" motivations. Instead, it is consistent with an alternative view of behavior in which some salient aspects of the game structure define the game, with a specific frame activating a particular set of psychological mechanisms (Messick 1999; Weber, Kopelman, and Messick 2004).

Our data show the existence of reverse reciprocity among the American trustees. We suspect it may be due to a reduction in their sense of moral obligation as activated in the one-shot Trust Game. In the Faith Game, the truster's fate is completely subject to the decision of the trustee (who is a dictator, at least in the eyes of the trustee in the Faith Game). This situation makes moral obligation salient to the trustee in the Faith Game. In contrast, trusters in the Trust Game have a means to defend themselves by choosing the safe option of \$10. Through their own choice and their own responsibility, they gave the trustees the chance to freely divide \$30 and give nothing to the trusters in return.

The fact that the trusters willingly trust the trustees thus may operate as a double-edged sword. On the one hand, it may enhance the sense of reciprocal obligation and make the trustees want to reciprocate. On the other, it may release the trustees from the moral obligation to be fair because, after all, the trusters could have defended themselves by taking the sure \$10, but they did not do so. If they get nothing, a trustee might reason, it is their responsibility, not mine. This self-serving logic leads to reverse reciprocity on the part of the trustee.

The effects work in opposite directions. In regard to the reciprocity effect, being trusted makes the trustee more trustworthy to the truster. Yet learning that the truster *willingly* gave up the opportunity for a sure \$10 from the experimenter frees the trustee from the moral obligation to be fair. If this is the case, the comparison of the Trust Game and the Faith Game does not provide valid conclusions concerning the role of reciprocity in trust and trustworthiness. First, the

result may depend on the existence of cues that make one effect more salient than the other. Second, and more important, the trustee's behavior in the Trust Game is not the sum of the two motivations of fairness/altruism and reciprocity. The assumption of the cumulative operation of these two motivations in the Trust Game (an assumption held by previous studies of reciprocity in trust and trustworthiness) may not correctly represent how individuals in these experiments actually make their decisions.

The finding that reverse reciprocity was observed only among the American trustees suggests that the dominance of one effect or another partially reflects cultural differences. One explanation for this difference is that the concept of "self-responsibility" seems to be more salient in the United States than in Japan (e.g., Hamilton and Sanders 1983). Thus subjects behave relatively fairly in the Faith Game, where there is no excuse for taking most of the endowment. In the Trust Game, on the other hand, the truster's assumed self-responsibility can give the trustee an excuse to behave in a self-interested manner.

Another possible reason why the reverse reciprocity effect emerges only among American trustees and not among the Japanese comes from a finding by Hayashi and her colleagues (1999), that the Japanese are more strongly motivated by reciprocity than are Americans. These researchers report experimental evidence that Japanese second players cooperate more often in a one-shot sequential Prisoner's Dilemma game when the first player already has decided to cooperate. When the first player has decided to defect, virtually no Japanese or American second players cooperate. If we assume that the base rate of reciprocity is higher among Japanese than American trustees, the reverse reciprocity engendered in the Trust Game makes reciprocity and reverse reciprocity about equal in strength among Japanese trustees. Among American trustees, however, where the base rate of reciprocity is weaker, the reverse reciprocity effect in the Trust Game is stronger than the reciprocity effect. Further research is required to determine which of these two possible explanations is valid.

Future research also should explore why previous findings on cross-cultural and gender differences in trust and trustworthiness were not replicated in this study. The finding by Buchan et al. (2002) and Kiyonari and Yamagishi (1999) that Americans are significantly more trustful than the Japanese was replicated in part, but only in the Trust Game. Americans trusted the trustee more frequently than did the Japanese in both games, especially in the Trust Game, although the difference did not reach statistical significance. These researchers' findings of no difference in the trustee's behavior also were replicated, but only in the Faith Game. Our findings on the trustee's behavior in the Faith Game are consistent with theirs: American and Japanese trustees gave about the same proportion of their endowments to their partners. The truster's behavior in the Trust Game, however, is inconsistent with the findings of Buchan and her colleagues, who used a variant of the Trust Game called the Investment Game. Despite the similarity of the game structures, American trustees gave much less than did their Japanese counterparts in our game, whereas these two groups of trustees gave about the same proportion of their endowment in the Investment Game of Buchan and her colleagues.

This difference may be based partially on methodological differences between our Trust Game and the Investment Game used by Buchan and colleagues. In our Trust Game, the choice of the truster is binary, either to receive the sure \$10 or to entrust \$30 to the responder. In the Investment Game, the truster's choice is continuous and quantitative: he or she can entrust any amount between zero and \$10. The risk of trusting is greater in our game than in the Investment Game because the choice in our game is binary; thus the truster may earn nothing if he or she chooses to trust the trustee. The truster's greater risk may make the perception of self-responsibility more salient in our game than in the Investment Game. Researchers interested in studying trust and trustworthiness experimentally must be sensitive to subtle variations in the different versions of the Trust Game, especially when studying cross-cultural differences.

We began by asking whether trust begets trustworthiness. Does the fact that a trustee is trusted induce the former to behave in a more trustworthy manner? In one-shot interactions with anonymous partners, such as those used in this study, the answer is no. The act of entrusting some portion of an endowment to someone in a one-time decision does not, by itself, lead to trustworthy behavior on the part of the trustee. This finding is particularly interesting because repeated interactions in which individuals take risks (such as entrusting another person with an important decision) are essential to trust building (Blau 1964; Cook et al. 2005; Holmes and Rempel 1989). In one-shot interactions, individuals have no way to build trust. Yet the use of one-shot games is essential to the investigation of trust and trustworthiness precisely because we need to enhance our understanding of solitary acts of risk taking, as opposed to recurrent acts of risk taking in the formation of trust relations over time. Under some circumstances it may be difficult for trust relations to emerge at all.

REFERENCES

- Berg, Joyce, John Dickhaut, and Kevin 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.
- 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.
- Camerer, Colin and Richard H. Thaler. 1995. "Ultimatums, Dictators and Manners." *Journal of Economic Perspectives* 9:209-19.
- Cook, Karen S., Toshio Yamagishi, Coye Cheshire, Robin Cooper, Masafumi Matsuda and Rie Mashima. 2005. "Trust Building via Risk Taking: A Cross-Societal Experiment." *Social Psychology Quarterly* 68:121-42.
- Cox, James C. 2001. "On the Economics of Reciprocity." Working Paper, Department of Economics, University of Arizona, Tucson.
- . 2002. "Trust, Reciprocity, and Other-Regarding Preferences: Group vs. Individuals and Males vs. Females." Pp. 331-49 in *Experimental Business Research*, edited by Rami Zwick and Amnon Rapoport. Boston: Kluwer.
- Dufwenberg, Martin and Uri Gneezy. 2000. "Measuring Beliefs in an Experimental Lost Wallet Game." *Games and Economic Behavior* 30:163-82.
- Fehr, Ernst and Urs Fischbacher. 2003. "The Nature of Human Altruism: Proximate Patterns and Evolutionary Origins." *Nature* 425:785-91.
- . 2004. "Third Party Punishment and Social Norms." *Evolution and Human Behavior* 25:63-87.
- Fehr, Ernst and Simon Gächter. 2000. "Fairness and Retaliation: The Economics of Reciprocity." *Journal of Economic Perspective* 14:159-81.
- Hamilton, V. Lee and Joseph Sanders. 1983. "Universals in Judging Wrongdoing: Japanese and Americans Compared." *American Sociological Review* 48:199-211.
- Hardin, Russell. 2002. *Trust and Trustworthiness*. New York: Russell Sage Foundation.
- Hayashi, Nahoko, Elinor Ostrom, James Walker, and Toshio Yamagishi. 1999. "Reciprocity, Trust, and the Sense of Control: A Cross-Societal Study." *Rationality and Society* 11:27-46.
- 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.
- Kiyonari, Toko, Margaret Foddy, and Toshio Yamagishi. 2004. "Trusting Reciprocity and Trusting Fairness." Graduate School of Letters, Hokkaido University. Unpublished manuscript.
- Kiyonari, Toko and Toshio Yamagishi. 1999. "A Comparative Study of Trust and Trustworthiness Using the Game of Entrustment" (in Japanese) *Japanese Journal of Social Psychology* 15:100-09.
- Kreps, David M. 1990. "Corporate Culture and Economic Theory." Pp. 90-143 in *Perspectives on Positive Political Economy*, edited by James Alt and Kenneth Shepsle. Cambridge, UK: Cambridge University Press.
- McCabe, Kevin and Vernon Smith. 2000. "A Two-Person Trust Game Played by Naïve and Sophisticated Subjects." *Proceedings of the National Academy of Sciences* 97:3777-81.
- Messick, David M. 1999. "Alternative Logics for Decision Making in Social Settings." *Journal of Economic Behavior and Organization* 39:11-28.
- Rotter, Julian. 1971. "Generalized Expectancies for Interpersonal Trust." *The American Psychologist* 26:443-52.
- Snijders, Chris and Gideon Keren. 1999. "Determinants of Trust." Pp. 355-85 in

- Games and Human Behavior*, edited by David V. Budescu, Ido Erev, and Rami Zwick. Mahwah: Erlbaum.
- Wang, Feixue and Toshio Yamagishi. 2005. "Group-Based Trust and Gender-Difference in China." *Asian Journal of Social Psychology*:199–210.
- Weber, J. Mark, Shirli Kopelman, and David M. Messick. 2004. "A Conceptual Review of Decision Making in Social Dilemmas: Applying a Logic of Appropriateness." *Personality and Social Psychology Review* 8:281–307.

Toko Kiyonari is a post-doctoral researcher in the Department of Psychology, Neuroscience, and Behaviour, McMaster University, Canada. Her research focuses on cooperation and sanctions in social dilemmas, trust, and inter-group conflict. Currently, she is examining ways to combine various approaches to understanding human cooperation, including those used in social psychology, sociology, behavioral economics, evolutionary psychology, and evolutionary biology.

Toshio Yamagishi is the Director of the Center for Cultural and Ecological Foundations of the Mind and Professor of Social Psychology at Hokkaido University, Japan. His current research interests included co-evolution of social institutions and adaptive psychological mechanisms. The topics of his recent publications include cooperation in social dilemmas, punishment as a means to promote cooperation, culture as a system of self-sustaining beliefs, group-based trust, and reciprocity.

Karen S. Cook is the Ray Lyman Wilbur Professor of Sociology and Department Chair at Stanford University. She is also the Director of the Institute for Research in the Social Sciences. She has published widely on the topics of social exchange, power-dependence relations, physician-patient trust, and, more generally, trust networks. Her most recent book (co-authored with R. Hardin and M. Levi) is *Cooperation Without Trust?* 2005, Russell Sage Foundation.

Coye Cheshire is an assistant professor at the UC-Berkeley School of Information. His research focuses on how various forms of exchange are produced and maintained, especially in computer-mediated environments such as the Internet. He is currently working on an NSF-funded study with Karen S. Cook and Alexandra Gerbasi that investigates uncertainty, risk, and shifts in modes of social exchange.