'Trust, Gullibility and Social Intelligence'

Article *in* Asian Journal Of Psychology · April 1999

Impact Factor: 0.83 · DOI: 10.1111/1467-839X.00030

3 authors, including:

Toshio Yamagishi
Hitotsubashi University
159 PUBLICATIONS **6,555** CITATIONS

Asian Journal of Social Psychology (1999) **2**: 145–161

Trust, gullibility, and social intelligence*

Toshio Yamagishi and Masako Kikuchi
Hokkaido University, Japan

Motoko Kosugi
Central Research Institute of Electric Power Industry

A series of experiments conducted in Japan by Yamagishi and his associates are presented, all consistently showing that high trusters (as measured with a general trust scale) are more sensitive than low trusters to information potentially revealing lack of trustworthiness in others and judge other people's choice in a one-shot prisoner's dilemma more accurately. Based on these findings, a new theoretical twist is introduced to the ''emancipation theory of trust'' originally proposed by Yamagishi & Yamagishi (1994), that emphasizes the relation-expansion role of trust in addition to the traditionally noticed relation-fortification role of trust. When opportunity cost for staying in a commitment relation is generally high, it is more advantageous not to stay in secure and stable commitment relations but to explore opportunities that lie outside, and yet such social exploration involves the risk of being exploited by untrustworthy people. It is thus a more gainful strategy to invest ''cognitive resources'' in the nurturing of ''social intelligence'' needed to detect signals of untrustworthiness. General trust may be conceived as a by-product of the development of such social intelligence. Those who have invested in the development of social intelligence can afford to maintain a high level of general trust, whereas those who have not are encouraged to assume that ''everyone is a thief'' and to refrain from pursuing potentially lucrative but risky outside opportunities.

Introduction

Does trust mean gullibility? Are trustful people naïve and credulous? Some people think that they are naïve and credulous, and thus gullible by definition. According to Schlenker, Helm & Tedeschi (1973), for example, ''interpersonal trust may be defined as a reliance upon information received from another person about uncertain environmental states and their

* Earlier versions of this paper have been presented at the First International Conference on Theory and Research in Group Processes, Jagiellonian University, Cracow, Poland, June 18–21, 1996, and the Second Conference of the Asian Association of Social Psychology, Kyoto, August 4–6, 1997. Address correspondence to Toshio Yamagishi, Faculty of Letters, Hokkaido University, N10 W7 Kita-ku, Sapporo, Japan 060-0810, email: Toshio@let.hokudai.ac.jp, fax: +81-11-706-3066.

accompanying outcomes in a risky situation'' (p. 419). There is some experimental evidence to support this view. For example, based on the finding that low trusters (as measured with the Interpersonal Trust Scale or ITS: Rotter, 1967) are quicker than high trusters to recognize adjectives describing a lack of trustworthiness displayed in a tachistscope, Gurtman & Lion (1982) claim that high trust is an indicator of indiscriminate acceptance of information provided by others. Similarly, Garske (1976) claims that high trusters have a less complex cognitive structure ''that is more simplistic and less useful for behavior discrimination and prediction'' (p. 619), and that ''generalized expectancies for interpersonal trust reduce the perceived threat posed by negative others, and concomitantly lessen the necessity for discriminative vigilance in the cognitive sphere'' (p. 618).

While this conception of trust and trustful people is intuitively convincing, most research evidence concerning the relationship between trust and gullibility disagrees with this popular conclusion (see Rotter (1980) for a review of research evidence on this issue). Of course, which answer is correct – whether trust means gullibility or not – depends on the definition of trust. If trust is defined in terms of nonchalance with which one accepts information provided by others, then trustful people are credulous and gullible by definition. Rotter (1980) rejects this definition of trust as credulousness, claiming that trustful people are the ones who have a high level of *general* expectation of other people's trustworthiness. That is, general trust is the ''default'' expectation of other people's trustworthiness. According to this definition of *general* trust, how trustful one is toward other people in general has nothing to do with how credulous one is. General trust is relevant when no specific information is provided concerning a particular person. In contrast, credulousness matters when such information is provided. How trustworthy one expects others to be in general without any relevant information is, at least logically, independent of how much vigilance one exercises in interacting with particular partners.

In this paper, we will provide experimental evidence contrary to the above popular belief that trustful people are in fact more vigilant and prudent in processing information about specific person's trustworthiness. Before presenting such experimental evidence and theoretical explanations for such counter-intuitive findings, we draw the reader's attention to an example of how uncensored applications of the above popular belief equating trust with lack of vigilance could produce a potentially misleading conclusion.

Declining trust in the United States?

Reflecting concerns among the general public, an increasingly large number of writers warn us about the deterioration of trust, and the potential harm it causes in our society. For example, citing results of questionnaire responses extending over the past two decades, Putnam (1993) warns us about possible negative consequences of a lack of ''social capital, such as trust'' that may underlie the failure of many social policies. We will show below that this popular position is at least partly based on a failure to conceptually distinguish trust from absence of vigilance. The questionnaire study cited by Putnam (1993) as an indication of the declining trust in the United States is the General Social Survey that has been annually conducted by the National Opinion Research Center since 1972. Figure 1 shows responses to the question, ''Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?'' As mentioned by Putnam (1993), the proportion of the ''most people can be trusted'' response has substantially decreased during the last two decades. The linear regression coefficient of year on that response between 1972

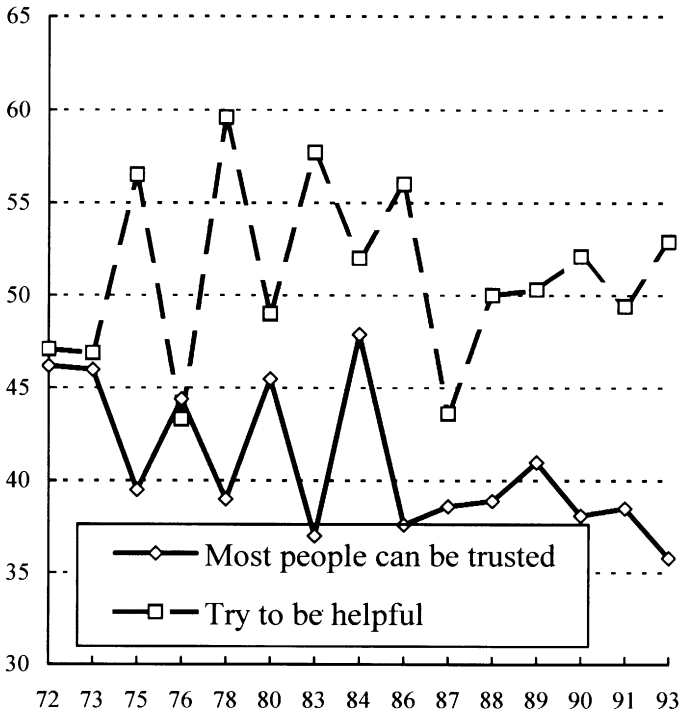

Figure 1. Trends in the ''most people can be trusted'' and ''people try to be helpful'' responses (in percentages) to two GSS items: ''Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with others?'', and ''Would you say that most of the time people try to be helpful, or that they are mostly just looking out for themselves?''

and 1993 is −0.0036, and the linear component of the trend is highly significant, $\chi^2(1)=45.46$, $p<.001$. This result implies that, on average, the ''most people can be trusted'' response to that questionnaire item has been declining since 1972 at an annual rate of 0.36 percent. This trend, according to Hochreich & Rotter (1970), already existed even in the 1960s.

The above survey results seem to unambiguously show that Americans have become distrustful of other people in general during the past two or three decades. However, this seemingly clear interpretation of the result is valid only when the two response categories to the questionnaire – ''most people can be trusted'' and ''you can't be too careful in dealing with people'' – represent two opposite ends of a single dimension, both logically and psychologically. Logically, they do not; the statement that most people can be trusted allows for the possibility that there are a few people who cannot be trusted. It is then prudent to be ready to deal with that small but non-zero probability of encountering untrustworthy people. I should not be blamed for inconsistency if I say that most people can be trusted and yet I am prepared to deal with the small number of untrustworthy ones. Psychologically, factor analysis studies of trust scales often show that these two – trust and need for prudence – constitute two separate factors, rather than representing two ends of a single factor. For example, two of the three factors reported by Kaplan (1973) who analyzed Rotter's (1967) Interpersonal Trust Scale are (1) perceived sincerity of others, and (2) need to be cautious of others. Similarly, analyzing the same ITS, Chun & Campbell (1974) found a factor for

''interpersonal exploitation'' consisting of items concerning ''self-protection or caution based on a perception of others as exploitative and egocentric'' (p.1064). This factor was found to be independent of another factor for honesty (which they call ''reliable role-performance''). Yamagishi and his associates (Yamagishi & Yamagishi, 1989; Yamagishi, 1986, 1988; Yamagishi & Sato, 1986; Yamagishi & Yamagishi, 1994) have repeatedly found two similar factors in Yamagishi's trust scale. The finding of Yamagishi & Yamagishi (1994) in their cross-national questionnaire survey with 501 Americans and 1,136 Japanese that the joint distribution of the two factors, trust and the need for prudence, has a triangular form is particularly interesting in this context. Most (74.5 percent) of the respondents whose level of general trust is low held a firm belief that caution is needed in dealing with others. And yet, the opposite was not true; a substantial proportion (41.3 percent) of those who showed a high level of trust also indicated a high level of caution in dealing with others. These results suggest that being prudent or cautious in dealing with others does not necessarily imply that the person is distrustful of others in general.

If one accepts, at least tentatively, the idea that prudence in dealing with others does not necessarily mean lack of trust *per se*, then the pattern shown in Figure 1 invites an alternative interpretation. That is, Americans have become more prudent, not less trustful of others in general. A clue to determining which of the two interpretations – declining trust or increasing vigilance – is valid is provided by another item in the General Social Survey. Figure 1 also reports the trend in responses to another item: ''Would you say that most of the time people try to be helpful, or that they are mostly just looking out for themselves?'' In contrast to the previous item, which is double-barreled, i.e. measuring trust on the one hand and need for prudence on the other, response categories to this question represent two extremes on a single dimension, belief in human benevolence. In this sense, it is more directly related to general trust that is defined as general, or ''default'' expectancies for the trustworthiness of other people. The figure shows no substantial decline in the positive response to this item, general trust as belief in human benevolence; it shows an increasing rather than declining trend, if there is any systematic trend at all. That is, the linear regression coefficient of year on the ''try to be helpful'' response is positive (b=0.0005).

Why do the trends in responses to those two similar items differ so much? Americans have been increasingly convinced that they ''can't be too careful in dealing with people,'' and yet they are no more convinced that people are selfish and are just looking out for themselves. The only reasonable explanation to this apparent puzzle is that these two items are measuring different things, the first the belief in the need for prudence[1] and the second the belief in the goodness of human nature. If this interpretation is correct, the increase in the need for prudence should have occurred mainly among high trusters. This is because, according to Yamagishi & Yamagishi (1994), most of the low trusters are cautious and the only room for an increase in prudence to occur must thus be among high trusters. That is, what are predicted to have increased are those who are trustful (or believe in the goodness of human nature in general) and yet vigilant when actually dealing with others. This prediction was also consistent with the result shown in Figure 2. Figure 2 reports the proportions (to the total respondents, excluding missing cases) of the cautious and prudent people (''can't be too careful in dealing with people'') within each response category of the general trust item. The proportion of the distrustful (''people are looking out for themselves'') and prudent (''can't be too careful in dealing with people'') respondents does not show any systematic trend over time. The linear regression coefficient of years is 0.0015, and the linear component has only a marginal effect, $\chi^2(1) = 3.70$, $p<.10$. In contrast, the proportion of the trustful (''people try to be helpful'') and yet prudent (''can't be too careful in dealing with people'') respondents

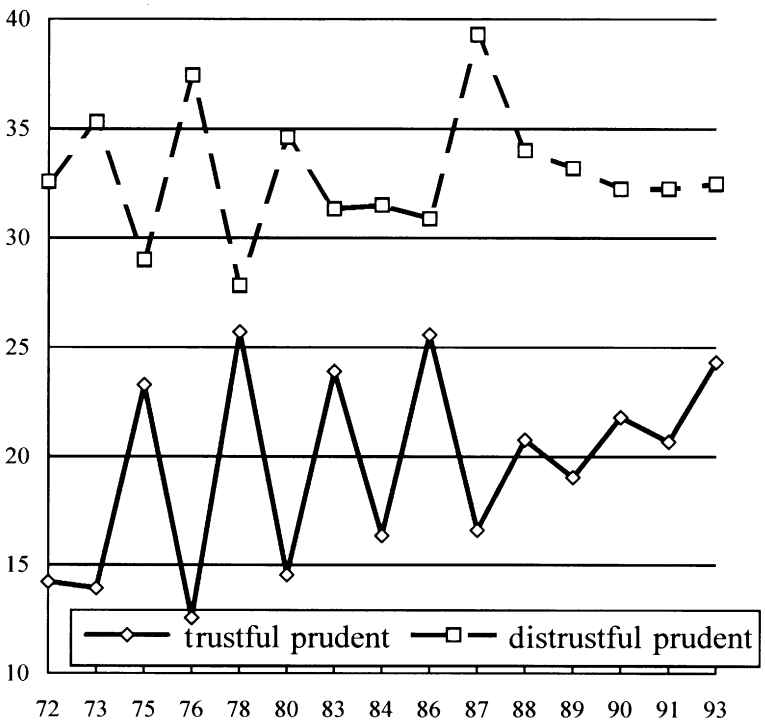

Figure 2. Percentages of trustful and prudent respondents and of distrustful and prudent respondents

shows a clear increase over time, with the linear regression coefficient of year of 0.0050, $\chi^2(1) = 45.72$, $p<.001$.

Have Americans really become distrustful? The trends over time shown in Figures 1 and 2 suggest a negative answer to this question. Americans are in fact becoming more apprehensive in dealing with other people, but this does not necessarily mean that they are becoming less trustful of others in the sense that they are less convinced of the goodness of human nature.

Trust and gullibility

In our casual conception of trust, trustful people are considered to be credulous and gullible. It is rather commonly believed that those who tend to trust others without hard evidence are easy prey to predators in the social jungle. The analysis presented above, however, suggests that those who have a cynical view of human nature are not the only ones who practice prudence in social interactions. Furthermore, experimental studies have repeatedly demonstrated that trustful people are not necessarily credulous and gullible. Following Rotter's (1980) suggestion, let us first define ''general trust'' as *default* expectations of other people's trustworthiness. High trusters are people who assume that people are trustworthy unless proven otherwise. Gullibility, on the other hand, is insensitivity to information revealing untrustworthiness. If we accept these definitions of general trust and gullibility, as Rotter (1980) did in his major review piece on this issue, trust and gullibility – that is,

default expectation of other people's trustworthiness in the absence of information and vigilance in processing trust-related information – are, at least logically, independent of each other.

How about empirically? Rotter (1980) first points out that SAT scores are not related to students' scores on the Interpersonal Trust Scale (ITS: Rotter, 1967), the most standard scale for measuring the respondent's level of general trust. In an earlier study by Rotter (1967), the ITS score was shown not to correlate with assessment of gullibility by members of the same fraternity or sorority. According to Geller (1966), high trusters exhibited more trustful behavior when there was no reason to be suspicious. However, where there was good reason to be suspicious, high trusters were no more trustful than low trusters. Hamsher (1968) and Wrights (1972) found that with an experience of being deceived in a game, high trusters were no more trustful than low trusters. These and other studies cited in Rotter (1971, 1980) provide evidence that high trusters are not necessarily credulous or gullible; what level of expectations one has for the trustworthiness of others in the absence of evidence is independent of how one processes information potentially related to untrustworthiness of a particular interaction partner.

The purpose of this paper, however, is not only to demonstrate that high trusters are not necessarily gullible, adding further evidence to Rotter's claim. We will go beyond this claim and present experimental findings showing that trust and gullibility are in fact interrelated but in the opposite direction to popular belief. That is, the experimental findings presented below show that high trusters are more vigilant in dealing with other people in socially uncertain situations.

Responses to positive and negative information

The first of the series of experiments conducted by Yamagishi and his colleagues is one by Kosugi & Yamagishi (1998) investigating how high trusters and low trusters respond to information potentially revealing trustworthiness or untrustworthiness of a target person. Participants in this experiment were shown scenarios and were asked whether the people depicted in them would act in a trustworthy manner. In one scenario, for example, a person who had stayed at a hotel in a foreign country for a week was charged for only one night when he checked out. He was paying cash, and there was no possibility that the hotel clerks could find out his home address or his next destination. Participants in this experiment were shown scenarios such as this and were asked whether the person depicted in the scenario would act in a trustworthy manner (for example, if the guest would tell the hotel cashier that he had stayed for a week) by indicating the probability of his or her behaving in such a manner. Participants' levels of general trust were measured with a six-item trust scale developed by Yamagishi & Yamagishi (1994).[2] As expected, high trusters believed more strongly than low trusters that the person would act in an honest and trustworthy manner (for example, the guest tells the truth to the cashier) when no information about that person was provided. This finding demonstrates the predictive validity of the trust scale since it is supposed to measure the level of ''default'' expectations for other people's trustworthiness.

In some scenarios, one or two pieces of information were provided that potentially revealed the trustworthiness or untrustworthiness of the focal person. For example, in some scenarios the participant was told that the focal person had picked up rubbish on the street and taken it to a garbage can (positive information). An example of negative information is that the focal person had cut into a waiting line. The purpose of the experiment was to see

how such information about the focal person would affect the participant's estimation of the trustworthiness of that person, and to see if high and low trusters respond in different ways to such information. Each participant made such estimations on 15 scenarios, and these scenarios were randomly combined with five information conditions (no information about the focal person, one piece of positive information, two pieces of positive information, one piece of negative information, and two pieces of negative information). Which information pieces were assigned to a particular scenario in each information condition was also random. Those combinations of the scenarios and information pieces differed from participant to participant. Each participant evaluated three scenarios in each information condition and thus the average of responses to the three scenarios in each condition was used as the participant's overall response to that particular type of information.

Both high and low trusters reduced the estimated likelihood that the focal person would act in a trustworthy manner as negative information was provided, but high trusters reduced the estimation more quickly than did low trusters. When two pieces of negative information were provided, high trusters' estimated likelihood that the focal person would act in a trustworthy manner was in fact *lower* than the estimation of low trusters, though the difference was not statistically significant. The change in high trusters' estimation was also larger than that of low trusters when positive information was provided, but the trust $\times$ information interaction did not reach statistical significance. In short, these results show that high trusters are more sensitive to information potentially revealing untrustworthiness of others, at least more so than low trusters.[3] Instead of being gullible, high trusters were shown to be more carefully attendant to information suggesting untrustworthiness of the target. An additional experiment also reported in Kosugi & Yamagishi (1998) succeeded in replicating the findings (1) that high trusters are more sensitive to negative pieces of information and (2) that, given such negative information, high trusters have a lower estimation of a target person's trustworthiness than do low trusters.

The result of those experiments that high trusters are more sensitive to information potentially revealing the target's untrustworthiness, however, is amenable to an alternative explanation, according to which high trusters' responses in those experiments imply their greater, not smaller, gullibility. This alternative explanation is based on the idea that high trusters changed their estimates of the target's trustworthiness more quickly than did low trusters because they were more acquiescent, taking at face value the information provided by the experimenter. The result of another experiment (Kakiuchi & Yamagishi, 1997; also reported in English in Yamagishi & Kakiuchi, 1997) presented below, however, eliminates this alternative explanation and provides further evidence for the original interpretation.

In this experiment (Kakiuchi & Yamagishi, 1997), two conditions were used. One condition was a standard iterated PD (prisoner's dilemma) in which two players played a PD game for 48 trials. The second condition was what they called the ''tiger's cave game,'' in which the two players had an additional choice, in addition to the ordinary behavioral choice between cooperation and defection, of enlarging or reducing their own payoffs in the payoff matrix. The payoff matrix was constructed in such a way that the player's payoff was negative whenever the partner defected. Enlarging one's own payoff in the payoff matrix was thus tantamount to increasing the stake both in positive and negative directions. By enlarging one's own payoff in the payoff matrix, the player increased the potential gain that accrued when the partner cooperated; at the same time, the player increased the potential loss when the partner defected. If a player expects the partner to cooperate, it is better to enlarge one's own payoff. Otherwise, it is better to reduce it. The choice of payoff size was thus used to measure the development of the player's trust in the partner. The central goal of

the experiment was to demonstrate that trust matters more in the choice of matrix size than in the cooperation/defection choice. This central hypothesis was confirmed.

What is directly relevant to the present issue is the result in the tiger's cave game condition. In the condition in which trust mattered more, high and low trusters behaved quite differently. The participant's level of general trust was measured a few weeks before the experiment using the same trust scale as the one used by Kosugi & Yamagishi (1998). We analyzed how high and low trusters responded to the cooperation–defection choice of the partner in the tiger's cave condition. For this analysis, the choice to enlarge one's own payoff was given a score of one, the choice to reduce it was given a score of minus one, and the choice of no change was given a score of zero. Two average change scores were then calculated for each trial block, one in trials in which the partner cooperated in the previous trial and the other in trials in which the partner had defected in the previous trial. Three trial blocks, each consisting of 16 trials, were used in the analysis. Participants who predominantly chose either C or D (15 trials or more out of 16 trials in each trial block) were excluded from this analysis.

Choice of one's own payoff following defection of the partner in the previous trial was significantly different between low and high trusters in the first trial block (first 16 trials). High trusters reduced their own payoff by .17 per trial, whereas low trusters enlarged their own payoff by .68 per trial following a trial in which the partner had defected. Although this result is contrary to our intuitive understanding of high and low trusters, it is consistent with the findings by Kosugi & Yamagishi (1998) presented above, indicating that high trusters were more sensitive than low trusters to information potentially revealing lack of trustworthiness in the game partner. High trusters were *initially* (i.e. in the first trial block) more prudent and vigilant than low trusters in dealing with a non-cooperative partner. As a result, low trusters' average earnings in the first trial block were negative (−.48 yen per trial) whereas high trusters earned 7.35 yen per trial on average.

The only difference between low and high trusters following a cooperative choice by the partner is in the second trial block. Here, low trusters (who increased the size by .44 per trial) were less willing than high trusters (who increased the size by .82 per trial) to enlarge one's own payoff even when the partner had previously cooperated. The overall picture that emerges from these results is that low trusters were initially less responsive to untrustworthy choices of the partner (i.e. the choice of defection), kept increasing their own payoff, and were exploited by the partner. Ostensibly because of this negative experience in interacting with the partner, low trusters became more hesitant in increasing their own payoff in the second trial block even when they were interacting with a cooperative partner. Low trusters were not prudent enough to guard against the possibility of exploitation by the partner early on, and then became overly cautious, failing to trust the partner when the partner was in fact trustworthy. Low trusters' earnings were kept in the red even in the second trial block (−.83 yen per trial) while high trusters' earnings kept increasing (21.49 yen per trial). In the third trial block, low trusters' earnings recovered into the black (12.38 yen per trial), while tailing behind high trusters (33.36 yen per trial). High trusters were prudent against defectors throughout the experiment, and yet were willing to trust a cooperator.

The result of this experiment is consistent with our interpretation that high trusters are more sensitive than low trusters to information revealing lack of trustworthiness of the interaction partner (i.e. the defection choice). Furthermore, the result also suggests that the way people process information about interaction partners' trustworthiness is related to their default expectation of other people's trustworthiness. The inability of low trusters to properly respond to relevant information made them overly cautious in the subsequent trials.

Low trusters were the ones who ''blew on a cold dish after being burnt by hot soup'' (to use an old Chinese proverb). We will return to this issue after reviewing a few more experiments.

Predicting interaction in partner's trustworthiness

Even if we are willing to accept the conclusion suggested by the above experimental results, i.e. that high trusters are more sensitive than low trusters to trustworthiness-related information, a question still remains: Does that heightened sensitivity among high trusters lead them to more accurate judgments in detecting untrustworthy people? An experiment by Kikuchi, Watanabe & Yamagishi (1997) was designed to provide an answer to this question. Participants in this experiment participated in a 30-minute discussion in six-person groups on garbage collection issues. They then participated in ''another'' experiment. In the second experiment, each participant played a two-person, one-shot prisoner's dilemma game with two of the other five participants. They were not informed as to the identity of their partners in these two games.[4] The prisoner's dilemma game was constructed in the following manner. Each participant was asked either to give 100 yen (about one dollar) to the partner or to take 100 yen from the partner. When the participant gave 100 yen, the partner received 200 yen. When the participant took 100 yen from the partner, the partner lost 200 yen.

Participants made the choice between cooperation (give 100 yen to the partner) and defection (take 100 yen from the partner) without knowing the identity of their partners. After they had made their decisions, they were told who their partners were and were given an opportunity to decline to play the game with each of the two partners. They were told that the game would be played only when both players (the participant and the partner) agreed to participate. The nature of the prisoner's dilemma game was such that one's own payoff was negative whenever the partner defected.[5] Participants were then asked to judge whether each of the two partners had cooperated or defected. They could earn an additional 100 yen for each correct judgment.

The purpose of the experiment was to determine if high trusters were more accurate than low trusters in judging partners' behavior. Participants were classified into three categories of low, medium and high trusters, based on their scores on the trust scale that they had filled out a few weeks before the experiment. There was no statistically significant difference among the three trust levels in the cooperation rate (53 percent, 53 percent, and 48 percent among low, medium, and high trusters, respectively). These three groups of participants also did not differ in their expectation of partners' cooperativeness. These results are consistent with past research findings that the effect of general trust on cooperation or expectation is limited in the n-person, not the two-person, situation. General trust matters less when people interact with a particular partner in a dyadic relation since their attention is focused on that particular person. On the other hand, high trusters were more accurate in estimating the partner's behavior (cooperation versus defection in a one-shot prisoner's dilemma game with anonymous partner) than either low trusters or medium trusters. Furthermore, they were accurate in judging the partner's choice regardless of whether the partner actually cooperated or defected. Of the 16 partners whom high trusters estimated to have cooperated, 12 actually cooperated (accuracy rate of 12/16 or 75 percent). They also estimated that 16 partners had defected, and 10 of those 16 had actually defected (accuracy rate of 10/16 or 62.5 percent). Overall, high trusters' estimations were accurate 22 out of 32 times (69 percent). In contrast, low trusters were accurate only 13 out of 30 times (43 percent) and

medium trusters were accurate only 12 out of 34 times. These differences in the accuracy score were statistically significant. Furthermore, high trusters played more often (91 percent) than medium (65 percent) or low (70 percent) trusters and actually earned more (75 yen) than either medium (−29 yen) or low (−100 yen) trusters. The effect of general trust on the accuracy in descerning defectors from cooperators in the one-shot PD games was also observed in a replication experiment by Kikuchi & Yamagishi (unreported).

The final experiment to be presented in this paper is another unpublished study conducted by Yamagishi and his associates. This is basically a replication of Kikuchi *et al*.'s (1997) experiment presented earlier. The major difference between this and the previous experiment by Kikuchi *et al*. (1997) is the participants. While participants in the Kikuchi *et al*. experiments consisted of strangers, participants in this experiment were acquaintances. The participants in this experiment consisted of two groups, each of which belonged to the same academic program (a smaller unit within a department). They had spent about a year in the same program, frequently meeting in classes.[6] Thus, their insights about other participants' personality traits should be much deeper than those of the participants in Kikuchi *et al*.'s experiment for whom participation in a group discussion for 30 minutes was the only source of information about the other participants.[7] Otherwise, the basic design of the experiment was replicated. First, participants played a PD game with one randomly selected student from a total of 33 participants. When they made a decision whether to cooperate or defect, they did not know who the partner was. They knew only that their partner was randomly selected from among the participants. They then answered a questionnaire including the trust scale and other personality scales. Finally, they judged whether each of the other 13 or 18 students from the same program had chosen to cooperate or defect. The participant was then told who their partners would be.

The overall cooperation rate in this experiment was 0.47. Accuracy of the judgment was calculated for each participant as the unweighted average of the proportions of correct judgments for cooperators and for defectors. For example, the accuracy of judgment score for the one who judged five out of seven actual cooperators as cooperators and three out of eleven actual defectors as defectors is (5/7 + 3/11)/2 = .49. The overall accuracy of the judgment was 0.48. That is, overall, participants' judgments of their friends' behavior in the PD game were no better than random guesswork. This result is consistent with the finding of Ross & Samuels (1993, cited in Ross & Ward, 1995) that dormitory advisers were not able to predict students' behavior in a PD game. However, the accuracy score of high trusters was much higher than that of low-trusters, and the accuracy score of the participants was significantly correlated with their general trust score (r = .48). Furthermore, the accuracy score was significantly correlated with other scales that are conceptually related to general trust. First, correlation with the ''honesty/fairness scale,'' a scale developed by Yamagishi & Yamagishi (1994) to measure the respondent's belief in the personal significance of being honest and fair to others, was significant (r = .43). Those who considered honesty and fairness personally to be more important were able to judge their friends' behavior in a PD game more accurately than those who cared less for being honest and fair to others. Second, the accuracy scale was correlated with the ''sense of interdependence scale'' developed by Jin (1997) to measure the belief that establishing and maintaining mutually cooperative relations is in one's own self-interest (r = .55). Those who had such a ''mentality of generalized exchange'' were more accurate than those who did not. Third, the proportion of their fellow students who judged the participant to have cooperated was marginally correlated to the participant's accuracy score (r=.32). That is, those who were considered by their friends to have cooperated in the experiment were more accurate in judging friends'

behavior in the experiment. The effect of general trust on accuracy of detecting defectors and cooperators among classmates have been replicated in all the three experiments that have been conducted by Yamagishi and his associates since the one presented above. In all of the four experiments using classmates, high trusters were consistently shown to be more accurate than low trusters in discerning defectors from cooperators.

Trust and social intelligence

The results of the series of experiments presented above consistently suggest that high trusters rather than low trusters are more sensitive to trust-related information and are more accurate in judging trustworthiness of others. It is not surprising to see that trust does not mean gullibility once we define trust and gullibility as we did earlier in this paper. However, the experimental findings presented above go beyond this, and imply that high trusters are more prudent and *less* gullible than low trusters. How can we explain this seeming paradox?

Our answer to this question is that general trust is generated and supported by social intelligence. That is, people with high social intelligence – those who are skilled in understanding their own and other people's internal states and use that understanding in social relations – are able to maintain a high level of general trust, whereas those with low social intelligence are not. Before presenting this argument, let me briefly summarize the ''emancipation theory'' of trust proposed by Yamagishi & Yamagishi (1994), since the argument suggested above is based on this theory.

According to the theory of trust developed by Yamagishi and his colleagues (Yamagishi, 1998; Yamagishi, Cook, & Watabe, 1998; Yamagishi & Komiyama, 1995; Yamagishi & Yamagishi, 1994), trust and commitment formation represent alternative solutions to the problem of social uncertainty. Social uncertainty is ubiquitous in human society. In interacting with others, we seek to improve our own welfare while exposing ourselves to a risk of exploitation from interaction partners. Social uncertainty is defined by Yamagishi & Yamagishi (1994) to exist for an actor when (1) the interaction partner has an incentive to act in a way that imposes costs to the actor, and (2) the actor does not have sufficient information to predict if the partner does in fact act in such a way.

Kollock (1994) provides a good example of social uncertainty and of how it promotes commitment formation[8] between particular partners. The example he used concerns rice and rubber markets in Southeast Asia. The quality of rice is immediately apparent upon simple inspection. The buyer has little risk of being cheated on the quality of rice he buys, and thus he faces a low level of social uncertainty. In contrast, the quality of raw rubber is hard to discern; its quality can be known only after it has been processed. Cheating on the quality is easier and the consequence of being cheated in this situation is extremely serious. The buyer of raw rubber thus faces a higher level of social uncertainty than does the buyer of rice. This difference in social uncertainty in the trade of rice and rubber, Kollock argues, explains the observed difference in the dominant form of trade. Rice is usually traded at an open market between relative strangers, whereas rubber is often traded between a particular producer and a broker who have formed a long-term relationship, often extending over generations. A high level of social uncertainty involved in the trading of rubber is the determining factor in the development of such committed relations between rubber producers and brokers.

Kollock conducted an experiment in which a laboratory version of rice and rubber trades was introduced. In one condition (high uncertainty condition), sellers could tell lies to potential buyers as to the quality of the product they sold. In the other condition (low

uncertainty condition), sellers could not lie. The results of Kollock's experiment clearly demonstrated that commitment formation between a particular seller and a particular buyer occurred more frequently in the high uncertainty condition than in the low uncertainty condition. Yamagishi *et al*. (1998) also report similar findings from their experiments that high social uncertainty promotes commitment formation.

Commitment formation as a solution to the problem of social uncertainty, however, has its own shortcomings. While reducing the risk of being duped in interacting with unfamiliar people, commitment restrains the actors from exploring better opportunities that might exist outside of the current relationship. Using terminology borrowed from economics, commitment formation reduces *transaction costs* on the one hand, and imposes *opportunity costs* on the other. In forming a commitment relationship with a particular partner, one obtains security (i.e. reduction in social uncertainty) in exchange for opportunities. Commitment formation is thus an efficient means of reducing uncertainty in a situation in which outside opportunities are limited (i.e. when the general level of opportunity cost for staying in a commitment relationship is low). On the other hand, a commitment relation becomes a liability rather than an asset as people face more and better opportunities outside of the current, mutually committed relationship (i.e. when the general level of opportunity costs for staying in a commitment relationship is high). According to Yamagishi & Yamagishi (1994), general trust (or default expectations of others' trustworthiness) provides a springboard for people who are confined in the security provided by commitment relationships to leap into the outside world of opportunities. General trust emancipates people from the confines of the security of stable relations.

Based on the argument briefly outlined above, Yamagishi & Yamagishi (1994) predicted that Americans would have a higher level of general trust than Japanese. Behind this prediction is an auxiliary assumption that the general level of opportunity costs for staying in a commitment relationship is higher in American society than in Japanese society, the latter being characterized by stable relationships such as the permanent employment system and *keiretsu* business relations. The prediction was based on the logic that having a high level of general trust makes people more willing to venture into the socially uncertain world in pursuit of better opportunities, and such an action is more gainful in American society than in Japanese society where many opportunities are closed to ''outsiders.'' This prediction was supported by the result of a cross-societal questionnaire survey conducted by Yamagishi & Yamagishi (1994). The survey results also supported additional predictions, similarly derived from their argument, that American respondents, compared to Japanese respondents, would consider reputation to be more important and would consider themselves more honest and fair. In contrast, Japanese respondents saw more utility in dealing with others through personal relations.

The above finding that Americans are more trustful of others in general than Japanese is rather counter intuitive, given the widely shared view of Japanese society, and in particular of Japanese business as one characterized by strong bonds of trust. However, a larger scale cross-societal survey conducted by the Institute of Statistical Mathematics (Hayashi, Suzuki, Suzuki & Murakami, 1982) also shows that Americans are more trusting than Japanese. Yamagishi & Yamagishi (1994) claim that the reason the finding appears to be counter-intuitive is because two qualitatively different types of trust – what they call ''trust'' and ''assurance'' – are confused in the commonsense conception of trust. According to their distinction, trust is expectations of benign or cooperative behavior based on the goodwill of the partner. Assurance, on the other hand, refers to expectations of benign behavior for reasons other than the goodwill of the partner. *Trust is based on the inference of the*

interaction partner's personal traits and intentions, whereas assurance is based on the knowledge of the incentive structure surrounding the relationship (p. 132).[9]

The reason for the counter-intuitive finding that Americans are more trustful than Japanese can be easily comprehended once the distinction between trust and assurance is introduced. *What characterizes Japanese society and business is in fact assurance, not trust.* The stability of inter-organizational as well as interpersonal relations in Japanese society makes exploitative, short-term profit maximizing behavior less profitable than in American society because those who desert a relationship for quick money will have a harder time in Japan finding another relationship in which to enjoy an equally comfortable life. In other words, the stable nature of social and organizational relations reduces social uncertainty and thus makes people feel secure inside such relations. This sense of security is what is often considered ''trust'' that characterizes the Japanese scene. Results of the cross-societal questionnaire survey conducted by Yamagishi & Yamagishi (1994) provide partial support to our claim that assurance rather than trust characterizes Japanese society. That is, Japanese respondents expressed a stronger belief than did Americans that they could benefit from using personal connections in dealing with others, or that preferential treatment of insiders was a matter of social fact that they could count on in everyday life. However, once such socio-relational bases of security are removed, Japanese would feel more insecure than Americans. They have not developed a high level of belief in human benevolence, and they are more distrustful of strangers outside of established relations. In short, in contrast to Americans, Japanese feel secure within established and stable relationships but are more distrustful of outsiders to such relationships.

Trust in this sense, that is, trust as a positive cognitive bias in processing information concerning trustworthiness of potential interaction partners, implies credulousness and thus gullibility especially when the world outside of the established relations is filled with nasty ''predators.'' The emancipation theory of trust is in this sense a theory of unintended benefits of credulousness. According to their conception, a ''trusting person is the one who overestimates the benignity of other people's intentions beyond the level warranted by the prudent assessment of the available information'' (Yamagishi & Yamagishi, 1994; p.135). The experimental findings presented in this paper, however, are not consistent with this conception of trust. Given information potentially revealing the target person's untrustworthiness, high trusters in Kosugi & Yamagishi's (1998) experiment more quickly reduced their judgment of the target's trustworthiness than did low trusters. Furthermore, high trusters in an experiment by Kikuchi *et al*. (1997) and in another, unpublished, experiment made more accurate estimations of interaction partners' behavior than did low trusters. Facing such experimental evidence against the ''cognitive bias view'' of general trust, Yamagishi (1998) proposed an investment model of trust development to explain the counter-intuitive findings that high trusters are more prudent and less gullible than low trusters.

Having a high level of general trust (or high ''default'' expectations of human benevolence) helps people to leave the security of established relations and look for better outside opportunities. In a social environment full of opportunities for those ''deserters'' of commitment relations, having a high level of general trust provides an advantage. And yet, exiting the current secure relations and exploring better outside opportunities entails risks of being exploited. In the social environment assumed in the emancipation theory to favor development of general trust – an environment characterized by a relatively high level of social uncertainly and opportunity costs for staying in a particular relation – people thus face the need for discerning trustworthiness of potential interaction partners. This implies that

investment of cognitive resources that is needed to discern trustworthiness of others will yield greater returns in such an environment than in another environment characterized by networks of stable relationships. Thus, it is predicted that people will invest more in cognitive resources, paying more attention and more carefully processing information concerning signals of trustworthiness in the high-uncertainty-high-opportunity-cost environment than in the low-uncertainty-low-opportunity-cost environment. As a result of these cognitive investments, skills needed for discerning trustworthiness will develop. General trust or the default expectations of trustworthiness of others can be a by-product of the investment in cognitive resources for improving ''social intelligence.'' Those who have invested a large amount of cognitive resources in developing such skills needed for discerning trustworthiness in other people can afford to maintain high default expectations of other people's trustworthiness. By maintaining these high expectations, they can enjoy the advantage of being able to fully explore opportunities that lie outside of the established relations. At the same time, they can quickly pull out of a risky relation at the first sign of danger. Those who have not made such a cognitive investment, on the other hand, are slow in detecting signals of untrustworthiness and thus are ill-suited to explore potentially fruitful but risky relations. It is prudent for them simply to assume that all people are untrustworthy, or to quote a Japanese saying, ''It's best to assume everyone is a thief.'' By assuming that (or by keeping the level of default expectations for trustworthiness of others low), they can avoid being exploited.

Examples of peasants and merchants would help. Peasants in isolated mountain villages in the pre-modern era lived in the village throughout their lives and had practically no opportunities for dealing with outsiders. For them, the outside world provided no better opportunities, and thus they were better advised to stay in the village and not to leave it unless forced to do so. Because they dealt exclusively with the same set of villagers, and because security was provided within such a small community, investments of cognitive resources for developing ''social intelligence'' to detect untrustworthy people generated no returns; it was a wasteful investment. As a consequence, they just assumed that anyone outside of their small community or all ''strangers'' were untrustworthy and avoided interacting with them. By doing this they lost practically nothing and could avoid the potential risk of being exploited. Both general trust and social intelligence is thus kept at a low level in such an environment, where opportunity costs for staying in secure relations are small. In contrast, opportunity costs play a much more important role in the life of merchants. They may prefer to deal with only a limited number of ''trustworthy'' trade partners. However, if they do so they have to pay opportunity costs; they could have obtained higher profits if they had dealt with new trade partners. One of the most critical decisions which merchants faced, especially in the pre-modern era when legal protection was not as efficient, was one between the security of dealing exclusively with trading partners of long-standing relations and expanding trade to new partners. The need for discerning the trustworthiness of new partners is much greater for merchants than for peasants, and thus investment of cognitive resources in the development of ''social intelligence'' is a more lucrative investment for the former than for the latter. Opportunity costs for staying in secure and stable relations are the major driving force for the development of such ''social intelligence.''

The above discussion of social intelligence and general trust further suggests an interesting twist in the trust and gullibility issue. According to the discussion, high trusters who have a higher level of social intelligence are less gullible than low trusters *per social interaction*. On the other hand, they may be more gullible in total than low trusters, since

they are more willing to enter into such risky social interactions. The popular view of high trusters being gullible may be derived from this difference; most people who are exploited are high trusters, since they are the ones who enter into such risky and yet potentially profitable social interactions. What people often fail to see is the other side of the coin – benefit forgone to low trusters. Using a Chinese saying, one has to venture into a tiger's cave to steal a baby tiger. People who have acquired a high level of social intelligence to discern trustworthiness in others are like those who have acquired skills for detecting the presence of adult tigers in the cave. Most of the time they can successfully get a baby tiger and make a fortune. On the other hand, they are the only ones who get killed by adult tigers, since those without such skills would not venture into tigers' caves. Whether or not it is better to develop that skill and venture into tigers' caves depends on how precious a baby tiger is. Neither social intelligence nor general trust will develop in a social environment in which there is no market for a baby tiger; i.e. in a social environment in which leaving the security of commitment relations provides no better opportunities than staying there.

Notes

1. More precisely, the first item measures a mix of general trust (or default expectations of human goodness) and the need for prudence, whereas the second item measures only general trust.
2. This trust scale and its earlier versions has been consistently successful in predicting people's behavior in a situation requiring trust (i.e. cooperation in social dilemma situations), demonstrating the predictive validity of this scale (Yamagishi, 1986, 1988, 1992; Yamagishi & Cook, 1993; Yamagishi and Sato, 1986).
3. An alternative interpretation of the findings is that negative information was more salient to high trusters than to low trusters since such information was unexpected for them.
4. The experiment presented here is only one of two conditions used in this experiment. The other condition is irrelevant to the discussion in this paper and thus is not reported.
5. Each participant was given an endowment of 500 yen as at the beginning of the experiment.
6. Each program (such as the Social Psychology Program) consists of up to 20 or so students per cohort. Due to course requirements, students on the same course see each other frequently. They also often have ''parties.'' Their bonding is much stronger than that among students in the same major in American universities.
7. Participants in this experiment did not have a discussion session before the experiment.
8. The term ''commitment'' in this paper is used in a strictly behaviorist manner. That is, one is defined to be committed to a relationship to the degree that he or she forgoes better (better at the moment) alternatives. Mutual attraction and liking, and the sense of loyalty to each other, may emerge in such a committed relation, and when they do, they will certainly strengthen the commitment. However, such psychological factors, however strongly related to commitment, are not commitment itself in this sense.
9. Shapiro, Sheppard & Cheraskin (1992) call assurance in this sense ''deterrence-based trust.''

References

Chun, K., & Campbell, J. B. (1974). Dimensionality of the Rotter Interpersonal Trust Scale. *Psychological Reports, 35*, 1059–1070.

Garske, John P. (1976). Personality and generalized expectancies for interpersonal trust. *Psychological Reports, 39*, 649–650.

Geller, J. D. (1966). Some personal and situational determinants of interpersonal trust. Unpublished

doctoral dissertation, University of Connecticut.
Gurtman, M. B., & Lion, C. (1982). Interpersonal trust and perceptual vigilance for trustworthiness descriptors. *Journal of Research in Personality, 16*, 108–117.
Hamsher, J. H., Jr. (1968). Validity of personality inventories as a function of disguise of purpose. Unpublished doctoral dissertation, University of Connecticut.
Hayashi, C., Suzuki, T., Suzuki, G., & Murakami, M. (1982). *A study of Japanese national character, Volume 4*. Tokyo: Idemitsushoten. (In Japanese with an English summary.)
Hochreich, D. J., & Rotter, J. B. (1970). Have college students become less trusting? *Journal of Personality and Social Psychology, 15*, 211–214.
Jin, N. (1997). Reconsidering social identity theory: an empirical study of within-group reciprocity. Doctoral dissertation, Hokkaido University, Sapporo, Japan. (In Japanese.)
Kakiuchi, R., & Yamagishi, T. (1997). General trust and the dilemma of variable interdependency. *Japanese Journal of Social Psychology, 12*, 212–221. (In Japanese with an English abstract.)
Kaplan, R. M. (1973). Components of trust: Note on use of Rotter's Scale. *Psychological Reports, 33*, 13–14.
Kikuchi, M., Watanabe, Y., & Yamagishi, T. (1997). Judgment accuracy of other's trustworthiness and general trust: An experimental study. *Japanese Journal of Experimental Social Psychology, 37*, 23–36. (In Japanese with an English abstract.)
Kollock, P. (1994). The emergence of exchange structures: An experimental study of uncertainty, commitment, and trust. *American Journal of Sociology, 100*, 313–345.
Kosugi, M., & Yamagishi, T. (1998). General trust and judgments of trustworthiness. *The Japanese Journal of Psychology, 69*, 349–357. (In Japanese with an English abstract.)
Putnam, R. (1993). The prosperous community: social capital and public affairs. *The American Prospect*, spring, 35–42.
Ross, L., & Samuels, S. M. (1993). The predictive power of personal reputation vs. labels and construal in the Prisoner's Dilemma game. Unpublished manuscript, Stanford University, Palo Alto, CA.
Ross, L., & Ward, A. (1995). Psychological barriers to dispute resolution. *Advances in Experimental Social Psychology, 27*, 355–304.
Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality, 35*, 651–665.
Rotter, J.B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist, 26*, 443–452.
Rotter, J.B. (1980). Interpersonal trust, trustworthiness, and gullibility. *American Psychologist, 35*, 1–7.
Schlenker, B.R., Helm, B. & Tedeschi, J.T. (1973). The effects of personality and situational variables on behavioral trust. *Journal of Personality and Social Psychology, 3*, 419–427.
Shapiro, D.L., Sheppard, B.H., & Cheraskin, L. (1992). Business on a handshake. *Negotiation Journal, 8*, 365–377.
Wrights, T.L. (1972). Situational and personality parameters of interpersonal in a modified prisoner's dilemma game. Unpublished doctoral dissertation, University of Connecticut.
Yamagishi, M., & Yamagishi, T. (1989). Trust, commitment, and the development of network structures. Paper presented at the Workshop for the Beyond Bureaucracy Research Project, December 18–21, Hong Kong.
Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology, 51*, 110–116.
Yamagishi, T. (1988). The provision of a sanctioning system in the United States and Japan. *Social Psychology Quarterly, 51*, 265–271.
Yamagishi, T. (1992). Group size and the provision of a sanctioning system in a social dilemma. In W.B.G. Liebrand, D.M. Messick, & H.A.M. Wilke (Eds.), *A social psychological approach to social dilemmas*. New York: Pergamon Press.
Yamagishi, T. (1998). *The structure of trust: The evolutionary games of mind and society*. Tokyo: Tokyo University Press. (In Japanese.)

Yamagishi, T., & Cook, K.S. (1993). Generalized exchange and social dilemmas. *Social Psychology Quarterly, 56*, 235–248.

Yamagishi, T., & Kakiuchi, R. (1997). It takes venturing into a tiger's cave to steal a baby tiger: Experiments on the development of trust relationships. Paper presented at the Management of Durable Relations: Theoretical and Empirical Models for Households and Organizations, June 26–28, Utrecht, The Netherlands.

Yamagishi, T., & Komiyama, H. (1995). Significance and the structure of trust: theoretical and empirical research on trust and commitment relations. *INSS Journal, 2*, 1–59. (In Japanese.)

Yamagishi, T., & Sato, K. (1986). Motivational bases of the public goods problem. *Journal of Personality and Social Psychology, 50*, 67–73.

Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion, 18*, 1994, 129–166.

Yamagishi, T., Cook, K. S., & Watabe, M. (1998). Uncertainty, trust and commitment formation in the United States and Japan. *American Journal of Sociology, 104*, 165–194.