THE SOCIAL EXCHANGE HEURISTIC

MANAGING ERRORS IN SOCIAL EXCHANGE

Toshio Yamagishi, Shigeru Terai, Toko Kiyonari, Nobuhiro Mifune, and Satoshi Kanazawa

ABSTRACT

We extend the logic of Haselton and Buss's (2000) error management theory to the domain of social exchange and propose that a psychological mechanism, referred to as the *social exchange heuristic* (SEH), produces certain cognitive biases that affect *how* individuals manage these errors. We predicted that the SEH would remain dormant in individuals who failed to realize that they were in a situation that involved social exchange. In the first experiment ($n = 78$), PD players who had a chance, before they played the game, to think both about the nature of the game they were playing and about their partner's choice were less cooperative compared to players of the standard one-shot PDG. In the second experiment ($n = 105$), PD players were again less cooperative when they made their decision before they were matched with a particular partner than after they had been matched with a partner. Results strongly suggest the operation of a SEH.

KEY WORDS • social exchange • prisoner's dilemma • social dilemma • heuristic • cooperation

Introduction

There are two consistent and yet puzzling findings in the prisoner's dilemma game (hereafter called PDG) research. First, non-trivial cooperation in the one-shot PDG is consistently observed. Second, researchers often find that the partner's (actual or expected) choice has an effect on the player's behavior. The purpose of this article is to present an argument, along with empirical evidence, claiming that a psychological mechanism,

Rationality and Society Copyright © 2007 Sage Publications. Vol. 19(3): 259–291
http://rss.sagepub.com DOI: 10.1177/1043463107080449

which we call the 'social exchange heuristic' (Kiyonari et al. 2000), provides an explanation for these puzzling findings.

Players often choose to cooperate in one-shot PDGs played between two unrelated players, forgoing personal rewards even under complete anonymity. Although the cooperation rate in one-shot PDGs varies greatly from experiment to experiment, rates close to 50% or above are not unusual. According to the standard model used in classic game theory, players should not cooperate. It is understandable for a player to cooperate if he or she believes that his or her *reputation* is at stake (e.g., Greif 1989; Milinski et al. 2001; Nowak and Sigmund 1998; Wedekind and Milinski 2000). What is puzzling, however, is the fact that many players cooperate in a one-shot PDG even when the reputation of their decision will not be disseminated.

The effect that a player's *expectations* of her partner's behavior have on her decision to cooperate is also puzzling. Since defection is the dominant choice in a one-shot PDG, a rational player who cares only for her own welfare should be indifferent to the choice of her partner. It is not surprising for players in *repeated games* to seek out what their partners have done in the past and then to adjust their own behavior accordingly. However, even players in a one-shot game are concerned with predicting what their partner might do so that they can adjust their behavior in accordance with the expected behavior of their partners. Researchers have repeatedly found that in the PDG and social dilemma game (SDG; *n*-person version of the PDG), the expectations that individuals have of the choices of other players are strongly related to their own choice between cooperation and defection (e.g., Alcock and Mansell 1977; Dawes et al. 1977; Fox and Guyer 1978; Marwell and Ames 1979; Orbell and Dawes 1991, 1993; Sato and Yamagishi 1986; Yamagishi 1986, 1988a, 1988b; Yamagishi and Sato 1986). In a review of research on the PDG, Pruitt and Kimmel (1977) concluded that the expectation that one's partner is willing to cooperate was the critical factor behind a player's cooperative choice in the PDG.

While the correlation between expectations and behavior does not necessarily mean that expectations cause behavior, results of one-shot, sequential PDG experiments clearly demonstrate that other players' choices do affect a player's own behavior even in a one-shot game. In a one-shot, sequential PDG, the first player makes a choice between cooperation (hereafter called the choice of C) or defection (the choice of D), and then the second player, who has been informed of the first player's choice, makes his or her decision. It is important to note that each makes the decision only once in this one-shot, sequential PDG. For the second

player, D produces an immediately better outcome than C and the choice of D does not affect the future choices of the first player. Thus, assuming that the second player is rational, she should choose D no matter what the first player chooses. Knowing that a rational second player will choose D no matter what the first player chooses, the first player should also choose D since it produces an immediately better outcome than C. Thus, based on the logic of backward induction used in game theory, the choice predicted for both players in one-shot, sequential PDGs as well as simultaneously played PDGs should be the same: defection.

The key for this prediction is that the second player's choice is independent of the first player's choice. In sharp contrast to this prediction, second players in one-shot, sequential PDG studies overwhelmingly chose C when the first player's choice was C. In response to the first player's choice of C, 75% of Japanese participants (Watabe et al. 1996; 62% in Kiyonari et al. 2000), 61% of American participants (Hayashi et al. 1999), and 73% of Korean participants (Cho and Choi 1999) made the choice of C as second players. When the first player's choice was D, on the other hand, practically all (88% of Japanese participants in Hayashi et al. 1999, and 100% of American and Korean participants) of the second players chose D. These findings clearly indicate that a substantial majority of second players reciprocated the first player's choice, even when the game was not repeated. Furthermore, data from the above experiments suggested that first players expected second players to reciprocate their behavior. The fact that the rate of cooperation by first players in the above experiments was substantially higher than the rate of cooperation among players who played simultaneous PDG (i.e., without knowledge of the partner's choice) supports this conclusion (83% vs. 56% of Japanese, 59% vs. 38% of Japanese, 56% vs. 36% of Americans, and 55% vs. 46% of Koreans; Cho and Choi 1999; Hayashi et al. 1999; Kiyonari et al. 2000). Thus, the majority of second players reciprocated their partner's decision: cooperation with cooperation and defection with defection. At the same time, the first player expected the second player would reciprocate even when there was no 'shadow of the future' (Axelrod 1984). Some researchers have argued that the relationship between expectations and behavior is produced via the projection of the player's own behavior onto that of their partner (Orbell and Dawes 1991, 1993). However, the results of the one-shot, sequential PDG studies cited above clearly indicate that people do adjust their behavior to the behavior of their partner even in one-shot games.

Social Heuristics

Heuristics play a decisive role in the decision-making process that takes place in experimental games as well as everyday life, and the use of heuristics is often more adaptive than deliberate decision-making (e.g., Bargh and Chartrand 1999; Gigerenzer and Todd 1999; McCabe 2003; Weber et al. 2004). McCabe (2003), for example, argues that people should have 'evolved cognitive mechanisms for capturing gains from exchange' (McCabe 2003: 153), which he calls 'goodwill accounting'. Goodwill accounting is a 'reputation-based scoring mechanism whereby people keep mental accounts of the extent to which potential trading partners can be relied on to establish a trust relationship in exchange settings' (p. 149). Players cooperate when they expect that their partner is willing to cooperate, not because they derive utility from their partner's outcome but because they have cognitive mechanisms that are designed to respond to goodwill offered by their exchange partner.

Other heuristics that have been argued to encourage cooperation in a one-shot PDG include the matching heuristic and illusion of control (Morris et al. 1998). The matching heuristic is 'related to the basic social norm of reciprocity' (Morris et al. 1998: 496). When activated, the matching heuristic motivates people to seek reciprocity *as an end in itself*. The second heuristic that Morris et al. (1998) suggest might be behind cooperative behavior in one-shot PD is illusion of control or 'quasi-magical thinking' (Shafir and Tversky 1992), whereby players believe they can control the choice of their partner. Having a sense of control encourages use of a tit-for-tat or similar strategy, and thus can be instrumental in producing mutual cooperation.

The heuristics described above – goodwill accounting, matching, and illusions of control – are considered to be adaptive in repeated games, and the use of them in one-shot situations is an overgeneralization of the heuristics to inappropriate situations. Social exchanges are characteristically repeated in nature, in contrast to purely economic exchanges, as pointed out by social exchange theorists (Blau 1964; Emerson 1976); thus using such heuristics as default decision rules for social exchange rarely results in a significant loss. This suggests that social heuristics are likely to be activated by cues that are typical of social exchange. Thus, cooperation in a one-shot PDG will depend on whether such cues exist in the game situation. As suggested by Messick (1999) and Weber et al. (2004), sensitivity to environmental cues is a critical factor in comparing the consequence-based versus rule-based models of cooperation in one-shot games.

Social Exchange Heuristic and Error Management

Overgeneralization of heuristics that lead players to cooperate in a one-shot PDG may produce maladaptive consequences. It is often argued that overgeneralized use of such heuristics is a result of cognitive misering (Fisk and Taylor 1991; McCabe 2003; Orbell and Dawes 1991). In contrast, we argue that the use of such heuristics may involve more than what the simple cognitive miser model suggests. We argue below that the use of such heuristics improves total gains under situations where one-shot games are mixed with repeated games.

Social Exchange Heuristic

Faced with the non-rational nature of reciprocal behavior, Kiyonari et al. (2000) sought an explanation for what they called the 'social exchange heuristic'. Yamagishi and his colleagues (Kiyonari et al. 2000; Yamagishi et al. 1999; Yamagishi and Kiyonari 1997) developed an earlier explanation for cooperation in one-shot games in terms of an illusion of control (Karp et al. 1993) and later proposed what they call the social exchange heuristic. They used the term 'social exchange heuristic' to emphasize the fact that it is uniquely used in social exchange situations. The social exchange heuristic prompts people to cooperate in one-shot games once it has been activated. Furthermore, they argued that it helps people reduce the likelihood of one type of error while increasing the likelihood of another type of error in social exchanges.

The logic of the social exchange heuristic (SEH) resembles Haselton and Buss's (2000) error management theory (EMT) in a few important ways.[1] Both SEH and EMT provide an adaptationist explanation for cognitive biases and heuristics. Both begin with the observation that decision-making under uncertainty often results in erroneous inferences, but some errors are more costly than others. An inference system applied across many decision-making tasks that minimizes the total *cost* of errors, instead of their total *number*, produces a net outcome that is most adaptive. Like Haselton and Buss (2000), we believe that understanding the adaptive function of a particular psychological mechanism can provide the insight necessary to formulate testable hypotheses concerning their operation. We agree with Haselton and Buss (2000) that many of our psychological mechanisms have been shaped by the adaptive problems we face, or, according to them, those that were faced by our ancestors. Haselton and Buss (2000) explicitly claim that the process that has shaped our psychological mechanisms is natural and sexual selection. The adaptationist approach we adopt in this article

does not have a position on this process. The process may be based on individual learning, or it may also be based on cultural evolution (i.e., imitation of successful strategies). The purpose of this study is to demonstrate the operation of the SEH. How it came to exist will be a topic of future study.

Error Management in Social Exchange

While Haselton and Buss (2000) apply their EMT exclusively to the domain of intersexual mind reading (inferences about the sexual interest of potential mates), the adaptationist logic used in EMT, we believe, is applicable to other domains of inference.[2] In this article, we continue our earlier work (Kiyonari et al. 2000; Yamagishi et al. 1999; Yamagishi and Kiyonari 1997) and extend the adaptationist logic we share with EMT to an entirely different domain of social exchange. Moreover, we propose that the SEH is a cognitive device for error management within this domain. The SEH could explain, among other things, how mutual cooperation can emerge with strangers even when the pressure to maintain one's reputation is absent. In this article, we discuss the results of a previous experiment and present data from two new experiments. The results of two experiments collectively support the hypothesis regarding the operation of the SEH.

The key to understanding both cooperation *and* the prominent role of the partner's behavior as a function of the SEH is the *uncertain* nature of exchange situations. In a PDG, as an example of social exchange, each party pays or does not pay a cost, c, to provide the other with a benefit, b, where $b > c$. Paying the cost, c, corresponds to the choice of C (cooperation) in the PDG, and not paying the cost corresponds to the choice of D (defection). This exchange of cost for benefit between two players produces four outcomes characterizing the PDG matrix shown in Figure 1. That is, the 'temptation payoff' of b (receiving the benefit without paying the cost) is greater than the 'reward payoff' of $b - c$ (paying the cost and receiving the benefit), which in turn is greater than the 'punishment payoff' of zero (not paying the cost and not receiving the benefit), which finally is greater than the 'sucker payoff' of $-c$ (paying the cost without receiving the benefit).

Quite unlike laboratory experiments with explicit instructions and clear explanations about the structure and rules of the game, encounters with potential exchange partners in the 'real' environment involve a great deal of uncertainty. Let's say you encounter someone for potential exchange. Is this truly a one-shot game with no possibility of future interactions? If so, then free-riding, or exploitation of your exchange

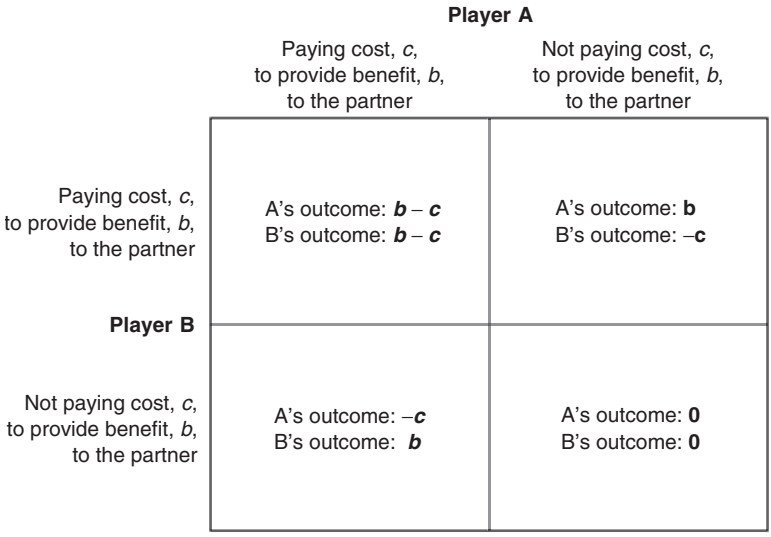

Figure 1. A PDG matrix representing four outcomes of a social exchange

partner, is possible and beneficial to you, and you should defect on her if you want to maximize your gain. Or, instead, is this situation merely the beginning of a long-term relationship in which you and she will exchange repeatedly in the future? In the latter situation, as Axelrod (1984) has shown, mutual cooperation is more profitable than mutual defection. Therefore, if this interaction is the first of many interactions with a partner, you should choose to cooperate as the first move of the *tit-for-tat* strategy. Moreover, even if this is just a one-shot encounter and you will never exchange with this partner again, information about your behavior could potentially spread to other members of the community who may use it to infer what you will do in future encounters. Under these circumstances, it is once again rational to cooperate, as a large number of studies have shown (Axelrod 1984; Davis and Holt 1993: 391–6). If the game with the same exchange partner is repeated in the future, or if there are reputation effects, then it is less possible to free-ride on your partner and get away with it.

The adaptive task faced in the above example is to estimate the likelihood of successful free-riding under uncertain conditions. Figure 2 posits the consequences of four possible states in this exchange relationship. In the diagonal cells of Figure 2, no error is involved. In the lower-right cell, the player correctly infers that attempts to free-ride (i.e., not paying the

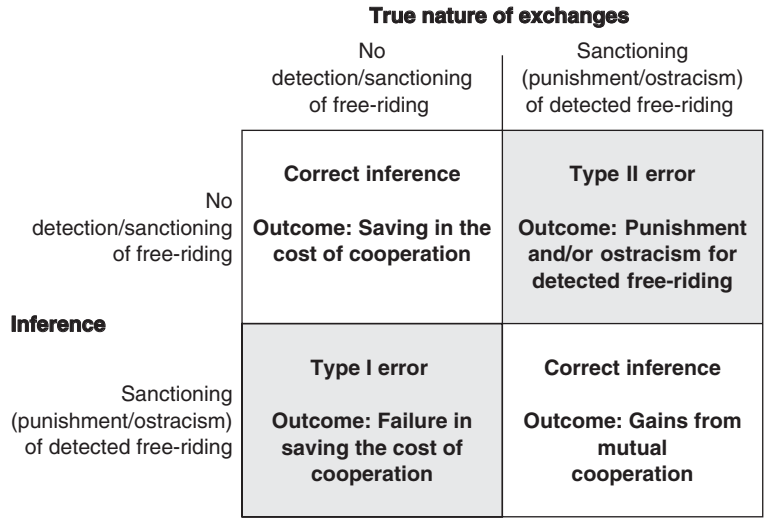

Figure 2. The consequences of four possible states of the social exchange

cost, *c*, to receive the benefit, *b*, will be detected. With this inference, the player chooses to pay the cost to provide benefits to the partner. Assuming that the exchange partner who is facing the same reality infers similarly, both the player and the partner benefit from the gains of mutual cooperation. In the upper-left cell, the player correctly infers that free-riding is possible without being detected (or the consequence of detection is not serious). Then she can save the cost of cooperation, *c*, which is, in fact, not needed to obtain the benefit from the other party.

Errors arise in cases represented by the off-diagonal cells. In the lower-left cell, the player infers that not only will attempts to free-ride be detected but also that the consequence of being detected will be serious even though this is *not* true. In this case, the player commits what Yamagishi et al. (1999) and Haselton and Buss (2000: 81–2) call a *Type I error*, following the language of inferential statistics. (The 'null hypothesis' in this case is that free-riding is possible.) The consequence of making this type of error is that the player pays the cost of cooperation, *c*, even though this is unnecessary to receive the benefit, *b*. The loss associated with making a Type I error is strictly limited to *c*, the cost of cooperation.

In the upper-right cell, the player makes the opposite (*Type II*) error. Here she infers that free-riding is possible without being detected (or that the consequence of being detected for free-riding is not serious) *even*

though free-riding is likely to be detected with serious consequences. In this case, one may have mistaken the exchange partner for a *stranger* with whom there would be no possibility of future interaction, or, perhaps, one might have defected in a setting in which other community members discovered one's objectionable behavior. Consequently, one's reputation within the community could suffer, and it is possible that one might face sanctions or be avoided as an exchange partner in the future. If sanctions involve ostracism from the community or exclusion from future exchange relations, the loss associated with making a Type II error is the sum of future benefits, $b - c$, that would have resulted from continued exchange relations.

How much Type I error an individual is willing to accept depends on the seriousness of committing a Type I vis-à-vis a Type II error, similar to the decision problem in inferential statistics. Minimizing the probability of committing Type I errors results in a higher probability of committing Type II errors. This would be preferable when the consequences of committing the former error are more serious than committing the latter, as is often the case in medical research testing a particular medicine that may involve deadly side-effects. We suggest that humans face the same kind of error management task in social exchange. In cases where the cost of sanctions, including ostracism from the community, is greater than the one-time savings, c, achieved by defection, a cognitive bias that perceives free-riding in exchange situations as neither possible nor desirable is adaptive. We call this cognitive bias the *social exchange heuristic*.

However, the SEH does not make the player cooperate unconditionally. When it is clear that the partner is not seeking mutual cooperation, then attainment of the 'reward' payoff – i.e., the future sum of $b - c$ – is impossible. In such cases, the SEH is unlikely to prevent the player from defecting since there is no benefit in cooperation. An individual should defect to avoid the 'sucker payoff'. We thus suggest that humans perceive PDGs as if they are Assurance Games (AG) when the SEH is operating. In an AG, choosing C produces a better payoff than choosing D insofar as the partner's choice is C. On the other hand, the payoff of choosing D is better than that of choosing C when the partner's choice is D. Players of the AG choose C when they expect their partner to choose C, and choose D when they expect their partner to choose D. As presented earlier, a series of experimental studies has demonstrated that the majority of players approach the game as if it were an AG rather than a PDG, preferring mutual cooperation over unilateral defection (Kiyonari et al. 2000: Table 2; Kollock 1997). Participants face the objective payoff matrix of a PDG, yet they subjectively perceive the situation as if it were

an AG. Furthermore, the fact that the majority of second players in the sequential PDG choose C rather than D when the first player has chosen C (Cho and Choi 1999; Hayashi et al. 1999; Kiyonari et al. 2000; Watabe et al. 1996), combined with the fact that brain activity indicates greater pleasure when mutual cooperation is achieved compared to successful free-riding (Rilling et al. 2002), suggest that people often play a PDG as if it were an AG.

Cooperation in One-Shot PDG and SEH

We mentioned two puzzling findings in research using the one-shot PDG at the beginning of this article: non-trivial cooperation and the effect of the partner's expected or real behavior on cooperation. The SEH helps explain both of these findings. Mutual cooperation is preferred to free-riding by players whose SEH has been activated, motivating them to choose C insofar as they expect that their partner will also choose C. Furthermore, when the SEH is activated, whether the partner chooses C or D is of critical importance since the exchange is perceived as an AG, in which there is no dominant choice; in an AG, the player's choice depends on the (expected) choice of his or her partner.

Our last point concerning the operation of the SEH is that it may affect expectations of the partner's behavior as well. Achieving and maintaining mutual cooperation are the ultimate goal of the SEH. This goal is unattainable to those who do not trust their potential exchange partners. Thus, the transformation of a PDG into an AG by means of the SEH is meaningless to individuals who expect their potential exchange partners to defect. In contrast, those who expect their potential exchange partners to cooperate can achieve mutual cooperation. Therefore, in addition to motivating an individual to cooperate in social exchange, the SEH should enhance the expectation that the partner will also cooperate. For individuals whose SEH remains unactivated, the exchange is perceived as a PDG, and mutual cooperation should not be a primary goal. They should be indifferent to their partner's choice and, thus, should always choose the dominant behavior: to defect.

Environmental Cues That Activate the SEH

When is the SEH activated? The operation of the SEH should be specific to the domain of social exchange, and, thus, cooperation or defection in a one-shot PDG depends on the presence or absence of social exchange cues. Human cognitive biases do not have to be activated constantly. Certain environmental cues may trigger (activate or deactivate) the SEH. We hypothe-

size that the cues most likely to activate the SEH, and thus induce people to cooperate, are those associated with, or suggestive of, social exchange. The defining feature of social exchange is mutual interdependence; the benefit each partner obtains from the relationship depends on the behavior of the partner. We thus predict that the perception that one's relationship with an exchange partner is interdependent, and that one is already in an ongoing social exchange, will facilitate the activation of the SEH. Activation will then lead one to perceive a PDG as an AG and to prefer mutual cooperation over unilateral defection.

Kiyonari et al. (2000) tested the effect of interdependence on the activation of the SEH. Contrary to the prediction made by the microeconomic theory of utility maximization, their experiments demonstrate that participants cooperate more when they are interdependent (the player's reward is a true function of his and his partner's choices) and *defect* more when they are not interdependent (when the outcome of their exchange involves symbolic points). They further show that the rate of cooperation among players of sequential PDGs is higher than the rate of cooperation among players of simultaneous PDGs. Kiyonari et al. (2000) argue that this occurred because sequential PDGs foster a sense of contingency among players. The difference in rates of cooperation between sequential and simultaneous PDGs is greater when the players were truly interdependent. Kiyonari et al.'s experiment supports the prediction that interdependence is an environmental trigger of the SEH. In the following section, we report the results of two experiments that test the effect of an ongoing social exchange on activation of the SEH.

Experiment 1

Purpose

The purpose of the two experiments we present below was to demonstrate that the cooperation rate in the PDG is higher for players who feel that they are in an exchange situation with their partner than for those who do not feel this way. This prediction is derived from the basic premise of our SEH account of cooperation in social exchange. Conversely, SEH is unlikely to be activated when players clearly realize that the PDG they are playing is not an instance of social exchange. In the two experiments, we manipulated the situation in which a one-shot PDG was played to make players realize that they were not in a social exchange situation with their partner. We did this in the first experiment by providing our participants with an option to choose which of the two games they preferred to play. As they were

making their choice, we asked them to think of what their partner would do in each of the games. In other words, participants were induced to think of the game while they were in no particular relationship with another person and, thus, the SEH was unlikely to have been activated at the time. We predict that the cooperation rate in this delayed-play condition (in which they first think of the game before playing) will be lower than the cooperation rate in the immediate-play, standard PDG condition.

Method

Participants

All participants were Japanese college freshmen. Seventy-eight students (54 male, 24 female) were recruited from a large participant pool at a major research university in Japan. The participant pool consisted of approximately 1,500 freshmen recruited from various classes on campus on the basis of their desire to earn money. Their participation was not part of a course requirement. The experimenter called and scheduled the participants for a particular experimental session. The participants arrived individually at a reception desk in the entrance lobby where they received an ID number and were told that they would be identified by this number throughout the experiment to preserve their anonymity, even to the experimenters. An experimenter then escorted each participant to his or her own individual compartment in the laboratory. Each participant stayed in his or her own room and did not meet any of the other participants before, during, or after the experiment, unless they happened to arrive at the reception desk simultaneously.

The Prisoner's Dilemma Game

Participants were handed the first set of instructions as they were individually escorted to their compartment. The first set of instructions told the participants that the purpose of the experiment was to study how people would behave in interpersonal transactions (*torihiki* in Japanese). Then the instructions explained the nature of the transaction (i.e., the PDG). The PDG used in this study takes the exchange format rather than the matrix format. The transaction starts as each participant receives an endowment of 400 yen (about $3.50). Next, the participant chooses between keeping the endowment or giving the endowment to his or her partner. If given to the partner, the endowment is doubled by the experimenter before the partner receives it. Using the same notations as before, the cost of cooperation, c, in this transaction was 400 yen and the benefit, b, was 800 yen. After reading the instructions, participants

answered a few quiz items that tested their understanding of the PDG. The experimenter checked the participants' answers to the quiz. If a participant made an error, the experimenter provided additional explanations until the participant understood the nature of the task. Afterwards the experimenter asked the participant to read the last page of the first set of instructions explaining each of the two types of games. In the *simultaneous* game, the two players make their choices simultaneously (without knowing what the other player had chosen before making their choice). In the *sequential* game, one player made his or her choice first, then the second player learned what that choice was, and then made his or her own choice. Participants were then told that the specific game they were going to play would be determined randomly by drawing a lot.

Manipulation of the Experimental Conditions

When all participants in the same experimental session had finished reading the last page of the first set of instructions, each drew a lot to determine which game they were to play. The participant was then randomly assigned to one of three between-subjects conditions: the immediate-play/simultaneous ($n = 25$) game condition, the immediate-play/sequential game condition ($n = 26$), or the delayed-play condition ($n = 27$). All the participants in the sequential game condition were assigned the role of the first player; no one was assigned the role of the second player. The participants in the simultaneous game condition were told that they must decide whether or not to give 400 yen to their partner. The participants in the delayed-play condition were given a chance to decide which type of game they wanted to play: either simultaneously with a partner, or as the first player in a sequential game. The choices provided to them did not include one for being the second player in a sequential game, implying that the second player would be assigned by the experimenter. Participants were given this choice in order for us to assess the activation of the SEH in the delayed-play condition. We hypothesize that being in the delayed-play condition suppresses the activation of the SEH, and thus participants in this condition are less likely to be reciprocal and hence more likely to choose the simultaneous game in which reciprocation is impossible.

The theoretically important difference between the immediate-play and the delayed-play conditions is whether or not the participant is already in an exchange relation with a particular partner *when she thinks about the nature of the relationship* and speculates as to what his or her partner would do. If the participant thinks about the game and the payoff consequences of their choice (C or D) before she is already in a

relationship with a particular exchange partner, the SEH is unlikely to be activated and thus she is more likely to make the strictly 'rational' choice of D. On the other hand, if the participant thinks about the game and the consequences of her choice after she has been in a particular game with a particular exchange partner, the SEH is more likely to be activated and she will be more reciprocal and more likely to choose C. Following assignment to the immediate-play condition, for either simultaneous or sequential games, participants were asked immediately what they expected their partner to do (C or D). The participants then made their decision about the endowment of 400 yen. The participants in the immediate-play condition were, therefore, already in a particular exchange relationship (a particular type of game) with a partner when they made their choice whether or not to give the endowment to their partner. In contrast, participants in the delayed-play condition were encouraged to think about the nature of the game (simultaneous or sequential) and to speculate on the possible choices of their partner *before* they actually played the game. Their inferences about their partner's behavior were thus made before they were committed to a particular relationship. We hypothesize that the participant's inference about his or her partner's behavior and his or her perception of the nature of the relationship should not be affected by the SEH in the delayed-play condition. Thus, participants in this condition should be more likely to perceive the relationship as a PDG rather than as an AG, and to choose D instead of C in the game of their choice.

Dependent Measures

We used two separate dependent measures in this experiment: 1) participants' expectation of their partner's behavior, and 2) participants' own behavioral choice (cooperation vs. defection). After the participants were assigned to a condition, an additional page of instructions corresponding to the assigned condition was handed to them. When all participants had finished reading the additional instructions, they were all asked to complete a short questionnaire. Two of the questionnaire items asked them to estimate the probability that their partner would cooperate in response to cooperation or defection on their part *in the sequential game*.[3] Specifically, the participants answered the following question: 'Suppose you have decided to give your 400 yen to your partner, and that your decision has been told to your partner. What do you estimate is the probability that your partner will give his or her 400 yen to you? Please give your probability estimate in the form of a percentage.' Similarly, they answered another question concerning the probability of cooperation given defection

by the first player. We also asked participants about their exchange partner's motives. The participants placed the completed questionnaire in an envelope, sealed it, and handed it to the experimenter.

When all participants – both in the immediate- and the delayed-play conditions – had completed the short questionnaire, they were provided with a decision sheet. The participants in the immediate-play condition (both the simultaneous and sequential game conditions) made their choice to either give the endowment of 400 yen to their partner or to keep the money. The participants in the delayed-play condition chose between playing a simultaneous and playing a sequential game. After a few minutes, the participants in the delayed condition were given another decision sheet asking them to decide whether to give the endowment to their partner or to keep it for themselves. The participant put the decision sheet with his or her decision marked on it in an envelope, sealed it, and handed it to the experimenter. The participant was then asked to complete a post-experimental questionnaire and place it in an envelope upon completion. Afterwards participants were individually paid the amount they had earned in the experiment,[4] and then left the laboratory separately.

Results

Expectations of Reciprocal Cooperation

We first analyzed the participants' responses to the pre-decisional questions concerning their expectations of their partner's behavior. In the pre-decisional short questionnaire, we first asked the participants to imagine that they would play the sequential game as first players. Next, we asked them how likely it was that their partner would cooperate in the sequential game if they themselves cooperated as the first players. (This question was asked to all participants regardless of the actual condition to which they were assigned.) In the analyses below we made two planned orthogonal comparisons of the participants' responses to this question. First, we compared the participants in the immediate- and delayed-play conditions, and, second, we compared the participants in the simultaneous and sequential games *within* the immediate-play condition. As predicted by the logic of the SEH, the participants in the immediate-play condition were more likely to expect cooperation (average likelihood of 56.34%, $sd = 25.66$) from their exchange partner than those in the delayed-play condition (40.26%, $sd = 23.58$), and the difference was significant, $F(1, 75) = 7.33$, $p < .01$ (see Figure 3). Since participants in both conditions (immediate- and delayed-play

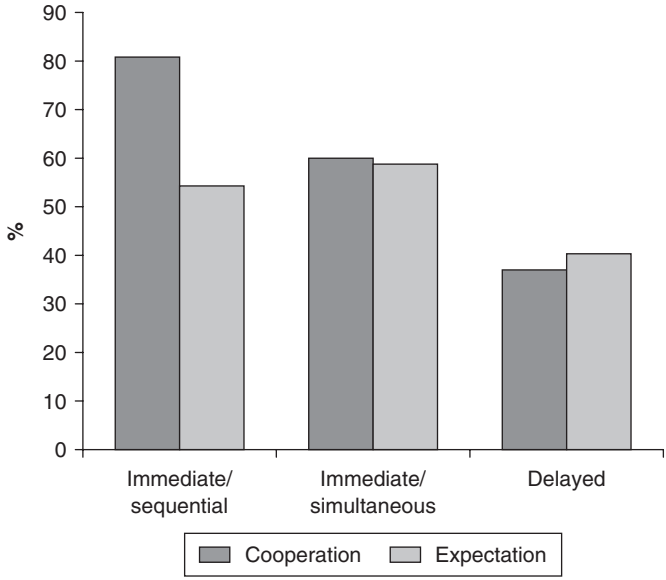

Figure 3. Cooperation rates and the expectations of reciprocal cooperation (Experiment 1)

conditions) were asked after receiving the initial instructions about the nature of the two games, there was no difference in the timing of this question between the two conditions. That is, there was no delay in the delayed-play condition regarding the answer to this question. Thus, this difference cannot be attributed to more time being given to the participants in the delayed-play condition as seemingly implied by the term 'delayed-play'.

Within the immediate-play condition, those playing the simultaneous game (58.60%, $sd = 22.25$) did not differ significantly in their expectation of their partner's cooperation from those playing the sequential game (54.17%, $sd = 25.03$), $F(1, 75) = .42$, *ns*. When the gender of the participant and the interaction effects of gender and the two contrasts were entered into the model, the main effect of sex was significant, $F(1, 72) = 4.21$, $p < .05$, but the two interaction effects were not significant. Women were more likely to expect their partner's cooperation (57.29%, $sd = 26.17$) compared to men (47.88%, $sd = 24.67$).

We then analyzed the difference between the participants' estimate of their partner's cooperation given their own cooperation, and the same estimate given their own defection. This difference is a measure of the participants' expectation that their exchange partner would behave in a

reciprocal manner, as rational players do *in an AG but not in a PDG*. Essentially, this is a measure of the extent to which the participants perceive the objective PDG that they are playing subjectively as an AG. Once again, as predicted, the participants in the immediate-play condition (mean difference of 51.43 percentage points, $sd = 24.53$) were more likely to expect reciprocal behavior from their exchange partner than those in the delayed-play condition (36.63 percentage points, $sd = 26.53$), and the difference was significant, $F(1, 75) = 6.26$, $p < .05$. As before, within the immediate-play condition, those playing the simultaneous game did not differ significantly from those playing the sequential game (48.77 versus 54.20), $F(1, 75) = .59$, *ns*. When the gender of the participant and the interaction effects of gender and the two contrasts were entered into the model, the main effect of gender was significant, $F(1, 72) = 5.52$, $p < .05$, but the two interaction effects did not reach the significance level. Women had a higher expectation of their partners' reciprocal behavior (53.54%, $sd = 26.15$) than men (43.09%, $sd = 25.59$).

In the same pre-decisional questionnaire, we also asked participants about their perception of their exchange partner's motives. Specifically, we asked them whether they perceived their exchange partner (as a second player) to prefer mutual cooperation over unilateral defection in the sequential PDG. The participants in the immediate-play condition (average response of 3.73, $sd = 1.73$, on a 7-point scale on which 1 indicated the preference for mutual cooperation and 7 indicated unilateral defection) were more likely to believe that their exchange partner would prefer mutual cooperation, while those in the delayed-play condition (5.00, $sd = 1.74$) were more likely to believe that their exchange partner would prefer unilateral defection. The difference between the conditions was statistically significant, $F(1, 72) = 6.16$, $p < .01$. The difference between the two immediate-play conditions (simultaneous and sequential games) was not significant, $F(1, 72) = 0.63$, *ns*. Neither the main effect of gender nor the two interaction effects involving gender were significant. These analyses provide strong evidence that letting participants think about the games before actually playing suppresses the activation of the SEH and, thus, the expectation that their partners will cooperate or behave in a reciprocal manner.

Choice of the Game

An overwhelming majority (74.1%; significantly greater than an even split, χ^2 (1) = 6.26, $p < .05$) of the participants in the delayed-play condition chose to play a simultaneous game rather than a sequential game. The finding that a majority of the participants chose the simultaneous

game is worthy of attention, especially considering the additional finding (detailed below) that the participants in the immediate-play condition generally expected reciprocal behavior from their partner and the majority of them cooperated. If the player expects reciprocal behavior from their partner and is willing to cooperate anyway, it is rational for them to choose the role of the first player in the sequential game rather than the simultaneous game since the chances will be greater that their cooperation will be reciprocated by their partner in the former role than in the latter. Yet when the participants were given a choice of which game to play, they avoided putting themselves in the position of the first player in the sequential game. Their overwhelming preference for the simultaneous game in the delayed-play condition provides further support for our claim that the participants' expectation of cooperation was low, due, presumably, to an inactive SEH, when such an inference was made outside of an ongoing exchange relationship. Without activation of the SEH, the participants in the delayed-play condition did not expect a reciprocal move from their partner (see the result above) and, thus, were unwilling to choose the role of initiator for the purpose of reciprocal cooperation.

Choice of Cooperation/Defection

Consistent with our prediction that being in an ongoing exchange relationship activates the SEH and prompts actors to seek mutual cooperation, 70.6% of the participants in the immediate-play condition chose to cooperate with their exchange partner, while only 37.0% of those in the delayed-play condition did so (see Figure 3). The comparison between the immediate-play and the delayed-play conditions was statistically significant, χ^2 (1) = 8.21, $p < .01$. Between the two immediate-play conditions, the cooperation rate (proportion of participants who chose to give) was higher in the sequential game than in the simultaneous game, as shown in Figure 3, replicating the pattern observed in previous studies (Cho and Choi 1999; Hayashi et al. 1999; Kiyonari et al. 2000; Watabe et al. 1996; Yamagishi and Kiyonari 2000). However, the difference found in this experiment was not statistically significant, χ^2 (1) = 2.65, *ns*.

The strong effect of the SEH in promoting cooperation was particularly evident when we compared players in the simultaneous PDGs (where, unlike in the sequential game, free-riding is a realistic possibility) across conditions. While 60.0% of players in the simultaneous/immediate-play condition chose to cooperate, only 20.0% of those in the delayed-play condition who chose to play the simultaneous game did so, χ^2 (1) = 7.27, $p < .01$. The rates of cooperation across the game-choice conditions

among players of the sequential game were not statistically different from each other (80.0% vs 85.7%, χ^2 (1) = .09, *ns*).

Discussion

The goal of the first experiment was to demonstrate that the cooperation rate in the one-shot PDG would be reduced greatly if the player deliberated on the nature of the game and on his or her partner's motives before actually playing the game. This goal was clearly achieved. Participants in the delayed-play condition who thought about the PDG and their partner's motives cooperated at a much lower rate than those who played the same game immediately without deliberation. In addition, we also demonstrated that participants expected much less reciprocity from their partners in the delayed-play condition than in the immediate-play condition.

While the manipulation in the first experiment produced results that clearly support our hypothesis, the manipulation used leaves us with some ambiguity. We assumed that participants in the delayed-play condition defined the PDG as a non-exchange situation when they thought about the situation and their partner's behavior before playing the game. That is, we assumed that the participants' thoughts about the game beforehand defined the game as a non-exchange situation in their mind, and that their later behavior in the game was framed by this definition of the situation.[5] A critic might argue that, regardless of how they perceived the situation before they played the game, they were actually in a social exchange relationship when they made their choice. We do not have direct evidence to show that our assumption is correct, or that the would-be critic's is not, except for the fact that the cooperation rate was much lower in the delayed-play condition than in the immediate-play condition. Furthermore, the results leave room for some alternative interpretations. For example, comparing the two games carefully may have made the rational nature of defection more salient in the delayed-choice condition than in the immediate-choice condition. Although two games were presented to the participants in the immediate-play conditions to make them compatible with the delayed-choice condition, those in the delayed-play condition may have thought about the two games more carefully when they faced a choice between the two. If this was the case, the lower cooperation rate in the delayed-play condition may be due to the fact that they were induced to think about the games more carefully rather than simply to the fact that they were not in a particular exchange relation. The ambiguity in the manipulation concerning

whether the participant was subjectively in a social exchange relationship or not, and the potential alternative explanation of the findings, prevent us from drawing a firm conclusion about the operation of the SEH in social exchange. We thus designed a second experiment in which participants in one condition were clearly in a particular relationship with a partner, while those in another condition were clearly not.

Experiment 2

Purpose

The purpose of the second experiment was to conceptually replicate the findings of the first experiment, and to demonstrate that the cooperation rate in a one-shot PDG decreases when the player makes the cooperation–defection decision outside of a social exchange situation, with a new manipulation that involves a less ambiguous interpretation. The control condition (to be called the 'partner-specified condition') is a standard one-shot PDG, in which a participant is matched with another participant whose personal identity is unknown to them, and decides how much of an 800 yen endowment to give to the partner. In the experimental condition (to be called the 'partner-unspecified condition'), participants choose the level of cooperation *before they are matched with a particular partner* for the outcome. They play the game with the understanding that their choices will be matched with another participant's *after everyone has made their decision*. Except for this seemingly minor difference – players make decisions before or after they are matched to specific partners – these two conditions are identical in every other way. Thus, the manipulation involves very little ambiguity concerning whether the participants are inside or outside of a particular exchange relationship. We predict that the SEH will be deactivated and, thus, the level of cooperation will be lower in the partner-unspecified condition than in the partner-specified condition.

Method

Participants

As in the first experiment, all participants were Japanese college freshmen ($n = 105$; 62 men and 43 women) recruited from a participant pool at the same university. The subject pool was formed through the same recruitment procedures used in the first experiment. As in the first experiment, participants arrived individually and were given ID cards to be

used throughout the experiment. Each participant stayed in his or her own room and did not meet any of the other participants before, during, or after the experiment unless they happened to arrive at the reception desk simultaneously.

Procedure

Participants were randomly assigned to either the partner-specified condition ($n = 53$; 21 women and 32 men) or the partner-unspecified condition ($n = 52$; 22 women and 30 men). All the participants played a one-shot PDG, similar to the one used in the first experiment, although in this experiment a participant's decision regarding the endowment was continuous rather than binary (see below for the details of the PDG used in the second experiment). Each participant was provided with an endowment of 800 yen and then decided how much of the money to give to another participant and how much to keep for him- or herself. As in the first experiment, all instructions were delivered in envelopes, and participants' decisions were collected in sealed envelopes that were opened by another experimenter to ensure anonymity of the participants' decisions. The first envelope contained the general instructions and introduced the experimental manipulation. At the beginning of the instructions, participants in the partner-specified condition were told that they had been matched with another participant, and that they would engage in a transaction (*yaritori* in Japanese) with that person. Participants in the partner-unspecified condition were simply told that they would engage in a transaction for money with someone in the group, without mentioning that they had already been matched with another participant. When they had finished reading the rules of the transaction, participants in the partner-unspecified condition were told that they would be matched with one of the other participants after all participants had made their decisions.

Once all participants had finished reading the instructions, the experimenter delivered the second envelope containing the decision sheet. Participants indicated how much of the 800 yen endowment they would give to their partner. The experimenter collected the envelopes in which the participants placed the completed decision sheets, and delivered the third envelope containing the post-experimental questionnaire. At the beginning of the post-experimental questionnaire, participants were asked how much they expected their partners to give to them.

The Prisoner's Dilemma Game

As in the first experiment, the PDG used in the second study took the exchange format rather than the matrix format. A player's choice in the

second experiment was continuous rather than binary. Each player decided how much of the endowment to give to his or her partner (in increments of 10 yen). The money given was doubled by the experimenter and then given to the partner. The player kept the remaining amount. For example, if a player gave 500 yen, his or her partner would receive 1,000 yen and the player would keep 300 yen. If, furthermore, the partner gave 300 yen, the player would receive 600 yen and the partner would keep 500 yen. As a consequence of the two players' choices, the first player would earn 900 yen (300 yen from the endowment and 600 yen from the partner) and the partner would earn 1,500 yen (500 yen from the endowment and 1,000 yen from the player).

Manipulation of the Partner Conditions

As stated above, participants in the partner-*specified* condition were told that they had already been matched with another participant to engage in a transaction (though they would never meet that person). Then they were told the following rules of the transaction: (1) They were each provided with an endowment of 800 yen, and their task was to decide how much of that money to give to the partner. They could keep the rest for themselves. (2) The money they gave to the partner (*aite*)[6] would be doubled. (3) The matched partner would make the same decision, and they would receive twice the amount the partner gave them.

Participants in the partner-*unspecified* condition were told that they would engage in an interaction with one of the other participants, but they were not told that they had already been matched with a partner. The rules of the transaction were then explained with some changes in wording. The rules were: (1) They would each be provided with an endowment of 800 yen, and their task was to decide how much of that money to give to someone. They could keep the rest for themselves. (2) The money they gave would be doubled and given to *someone* (one of the participants as opposed to a *matched partner*). (3) All the participants make the same decision. It should be emphasized that the participants in this condition did know that they would play with another participant; they were not different from participants in the partner-specified condition in this respect. The only difference was that they were explicitly told that the matching with their partner would be made after they had made their decisions, whereas those in the partner-specified condition were told that they had been matched with another participant.

After these initial instructions, participants in the partner-unspecified condition were informed that the match between participants (whose

money goes to whom) would be randomly determined after all of the participants had made their decision. The important point in this instruction is that participants were explicitly told that the matching of the game players would be done *after* they had made their decisions. In other words, when participants made their decision, they were not in an exchange relationship with another participant. In contrast, participants in the partner-specified condition were reminded that they were engaged in a transaction that was characterized by an already-formed pair. The final page of the instructions reiterated these rules of the transaction, again emphasizing the presence or absence of a matched partner in the two respective conditions.

Results

Two participants expressed strong suspicion of the experimental procedures by indicating in the post-experimental questionnaire that they thought: (1) that the money they were going to receive was predetermined and independent of their decision, (2) that the experimenter would not let them leave the laboratory empty-handed, and (3) that they were the only participants and other participants did not exist. They answered 7 on the 7-point scale on all three questions. We decided to eliminate these two participants from our analysis.

The Level of Cooperation

The comparison of cooperation levels (i.e., the amount of money participants gave to their partners) in the two conditions provided clear support for our hypothesis (see Figure 4). The cooperation level was much lower among participants in the partner-unspecified condition (219.42 yen out of the endowment of 800 yen, $sd = 230.31$) than in the partner-specified condition (334.51 yen, $sd = 258.11$). The main effect of the condition in the condition × gender ANOVA was significant, $F(1, 99) = 4.56$, $p < .05$. The main effect of gender was also significant, $F(1, 99) = 7.45$, $p < .01$. Men (330.00 yen, $sd = 280.54$) were more cooperative than women (201.63 yen, $sd = 177.35$). The condition × gender interaction was marginally significant, $F(1, 99) = 2.98$, $p < .09$. The effect of the experimental conditions tended to be more pronounced among men (420.67 yen in the partner-specified condition and 239.33 yen in the partner-unspecified condition) than among women (211.43 yen in the partner-specified condition and 199.27 yen in the partner-unspecified condition).

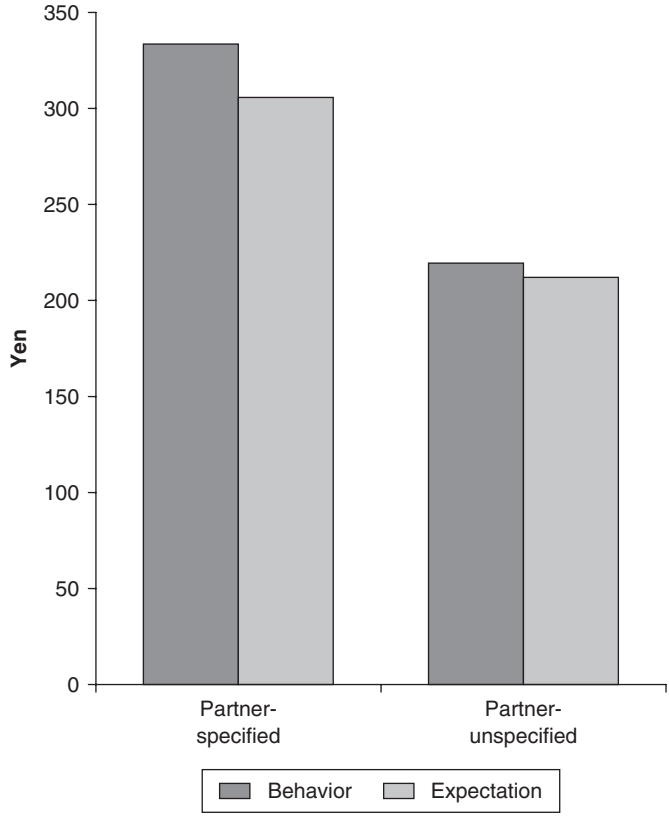

Figure 4. The average cooperation level (money given to the partner) and the expectation of the partner's cooperation level (Experiment 2)

Expectations of Cooperation

Participants were asked how much they expected their matched partner to give to them. Their expectations were consistent with the behavioral data and support our hypothesis. Participants in the partner-unspecified condition expected a lower level of cooperation (211.54 yen, $sd = 201.77$) from their partners than those in the partner-specified condition (305.10 yen, $sd = 209.32$). The main effect of the condition was significant, $F(1, 99) = 4.43$, $p < .05$. The main effect of gender was also significant, $F(1, 99) = 6.90$, $p < .01$. Men expected more cooperation from their partners (301.67 yen, $sd = 239.46$) than women (196.74 yen, $sd = 140.88$). The condition × gender interaction was not significant, $F(1, 99) = 1.72$, *ns.*

Post-experimental Questionnaire

Matrix transformation. We asked our participants in the post-experimental questionnaire to rank the order of personal satisfaction they would feel given the four extreme outcomes (both give 800 yen, both give nothing, you give 800 yen and your partner gives nothing, you give nothing and your partner gives 800 yen). We defined transformers as those who ranked mutual cooperation higher than unilateral defection. That is, those who were subjectively playing an AG should have felt more satisfaction from mutual cooperation, while those who were subjectively playing a PDG should have felt greater satisfaction from unilateral defection. If our hypothesis concerning operation of the SEH is correct, subjective transformation of the matrix should take place more often in the partner-specified condition than in the partner-unspecified condition. This prediction was confirmed by the participants' questionnaire responses. The overwhelming majority (94.1%) of the participants in the partner-specified condition were matrix transformers, stating that they would feel more satisfaction from the outcome of mutual cooperation than from unilateral defection. The proportion of transformers in the partner-unspecified condition was much smaller (73.1%). The difference was significant, χ^2 (1) = 8.27, $p < .01$. Furthermore, no one, either in the partner-specified or partner-unspecified condition, stated that he or she would feel more satisfaction from unilateral cooperation than from mutual defection. That is, no one felt they would be satisfied with the 'sucker payoff'.

Empathy and role taking. We included a five-item role-taking scale in the post-experimental questionnaire. The role-taking scale was taken from Tanida and Yamagishi (2004), who administered an empathy scale consisting of 25 items taken from the IRI (Davis 1983) and 25 items taken from the QMEE (Mehrabian and Epstein 1972) to 764 students, and obtained four factors of empathy. Among the four factors – an other-directed emotional responses factor (corresponding to the empathic concern factor of IRI); a self-directed emotional responses factor (corresponding to the personal distress factor of IRI); an imagination factor (corresponding to the fantasy factor of IRI); and a role-taking factor (corresponding to the perspective-taking factor of IRI) – we expected that the role-taking factor would be related to activation of the SEH. The five items of role taking that we included in the post-experimental questionnaire were: 'When I'm upset at someone, I usually try to "put myself in his shoes" for a while'; 'I sometimes try to understand my friends better by imagining how things look from their perspective'; 'I believe that there are two sides to every question and try to look at them both';

'Before criticizing somebody, I try to imagine how I would feel if I were in their place'; and 'I try to look at everybody's side of a disagreement before I make a decision'. We expected that activation of the SEH would enhance the participants' endorsement of these items because establishment and maintenance of a long-term cooperative relationship requires role taking (i.e., seeing the relationship from the point of view of the partner). As expected, the average role-taking score for these five items was significantly higher among participants in the partner-specified condition (5.05, $sd = 1.12$) than among those in the partner-unspecified condition (4.60, $sd = .98$; $t(101) = 2.15$, $p < .05$). Controlling for gender in a condition × gender ANOVA yielded a similar result, $F(1, 99) = 4.11$, $p < .05$.

Discussion

The results of the second experiment provide clearer support for our hypothesis than the results of the first experiment. Whether or not the participant is in an exchange relationship with a particular partner is clearly manipulated, and the resulting difference in the level of cooperation is clear-cut. Despite the fact that the nature of the game was exactly the same across the experimental conditions, participants in the partner-specified condition gave more money to their partners than those in the partner-unspecified condition. A similar difference was also observed in the participants' expectations of the partner's level of cooperation. Furthermore, the post-experimental questionnaire data indicated that the subjective transformation of the PD matrix into an AG matrix occurred more frequently in the partner-specified condition than in the partner-unspecified condition, and that participants in the partner-specified condition felt more strongly than those in the partner-unspecified condition that they were taking the perspective of others.

General Discussion

The results of the two experiments reported above are clear. We hypothesized that the SEH would more likely be activated and, thus, a PD player would more likely cooperate when the game situation was perceived to be one involving social exchange. The goal of the experiments we presented above was to test this hypothesis by getting the participants to make their behavioral decisions while they were outside a particular exchange relationship, and by comparing the resultant levels of

cooperation in this condition with that in a standard PDG. In the first experiment we accomplished this by allowing the participants in the delayed-play condition to think about the PDG and deliberate on the choice of their partners before they actually were in a particular game. We hypothesized that they would be less likely to frame the PDG as an instance of social exchange while thinking about it outside a particular game (simultaneous or sequential), and thus their cooperation rate would be lower than in the immediate-play condition. This prediction was supported by the results of the first experiment.

We assumed that having an opportunity to think about the game while not in a particular interdependent situation would suppress the sense of being in a social exchange, and that this definition of the situation would suppress activation of the SEH. The validity of this assumption was not directly tested in the first experiment, however, and, thus, the manipulation leaves some ambiguity in its interpretation. The second experiment was designed to address this problem. Participants in the partner-specified condition were told that they had been matched with one of the other participants, were given the instructions for the rules of the game, and then were asked to make a behavioral decision. In contrast, participants in the partner-unspecified condition were told that they would be matched with a partner only *after* all of the participants had made their decision. Participants in either condition made a decision not knowing what their partner would do. The seemingly trivial difference in the timing of the matching produced a large difference in the level of cooperation.

Expectations of the partner's cooperation were also consistent with our hypothesis. We predicted that the SEH would enhance the expectation of cooperation from the partner. As expected, participants in conditions where we did *not* expect SEH to be activated (the delayed-play condition in Experiment 1 and the partner-unspecified condition in Experiment 2) were less likely to expect their partners to cooperate than participants in conditions where we *did* expect SEH to be activated (the immediate-play condition in Experiment 1 and the partner-specified condition in Experiment 2).

Additionally, the post-experimental questionnaire data from Experiment 2 provided further evidence that activation of the SEH is suppressed if the fact that the participant is not in an exchange relationship is made salient. The first piece of evidence comes from the data on the desirability of the four possible outcomes. The proportion of participants who were subjectively playing an AG rather than a PDG was lower in the partner-unspecified condition than in the partner-specified condition. Second, the participants in the partner-unspecified condition

were less willing than those in the partner-specified condition to take the perspective of another person.

We attribute the above findings to activation of the SEH. However, an alternative interpretation may attribute them to differences in self-orientation: personal- versus social-self. Once in an exchange relationship, the participant is a part of a dyad, which could lead to inclusion of the exchange partner as part of the in-group. The increased level of cooperation observed in exchange relationships thus reflects a shared social identity. This alternative interpretation, however, implies unconditional, rather than conditional, cooperation. That is, heightened social identity may blur the distinction between the individual and the in-group member (i.e., the exchange partner) and induce the participant to treat the partner more favorably. Contrary to this prediction, the questionnaire data indicated that the participants preferred cooperation to defection only when their partner cooperated. No one preferred cooperation to defection when the partner defected. Treating the exchange partner favorably was conditional on the partner's cooperation. Social identity cannot explain this fact. An alternative theorization of group identity is that it enhances the desirability of *mutual cooperation* rather than making people altruistic toward the exchange partner (in which case cooperation should be unconditional). This conception of group identity is consistent with Yamagishi and Kiyonari's (2000) approach to the group as a 'container of generalized reciprocity'. Facing a group, people perceive it to be a venue in which they help each other, and come to adopt mutual cooperation (not being altruistic to other group members) as a goal of their group behavior. This interpretation provides a new and potentially highly productive approach to social identity from a social exchange theoretic perspective, and broadens the scope of the SEH hypothesis.

We propose that the SEH is a reasonable candidate for a psychological mechanism responsible for the subjective matrix transformation (i.e., cooperation becomes the default decision heuristic in the domain of social exchange insofar as the player expects that his or her partner will cooperate as well). In other words, the behavioral choice of cooperation itself, rather than its outcome, assumes a positive utility when a player feels that he or she is in a social exchange relationship. If the partner defects, it is clear that no social exchange relationship exists and, thus, cooperation loses its positive utility. The most important message from the current study is that the subjective matrix transformation occurs when PD players are (or perceive that they are) in social exchange. Identification of the particular mechanisms through which the transformation occurs constitutes the next stage of research.

While we succeeded in demonstrating that the removal of environmental cues that suggest the game situation is an instance of social exchange reduces the cooperation level, it was not reduced to zero. A substantial level of cooperation was observed in both experiments even when the participants played the game outside an exchange relationship. This, we suspect, indicates that the current manipulation was not strong enough to completely eliminate the operation of the SEH, as suggested by the fact that matrix transformation, as observed in post-experimental questions, was observed among a substantial number of participants even in the non-exchange conditions. The SEH is a robust psychological mechanism which makes people seek mutual cooperation in social exchange.

In presenting the SEH hypothesis, we avoided the question of whether or not the SEH is an *evolutionarily based*, domain-specific cognitive module. We do not believe that we have sufficient empirical evidence to answer this question. Before seriously posing this question, we need to collect data from multiple societies. The study reported in this article and the study conducted by Kiyonari et al. (2000) were both conducted in Japan with Japanese participants. Evidence of the SEH from multiple societies would be consistent with, but would not provide the necessary support for, the argument that the SEH is an evolutionarily selected cognitive module in the way that Haselton and Buss (2000) claim in their EMT. The social environment favoring the reduction of Type II, rather than Type I, errors may characterize the overwhelming majority of contemporary as well as traditional societies. If this is the case, the activation of the SEH (or the phenomena we characterize as consequences of the SEH) can spread simply because it is adaptive in such a social environment.

On the other hand, the consequences of committing Type II errors vis-à-vis Type I errors may vary from society to society. The consequences of committing Type II errors are expected to be more serious in collectivist societies in which ostracized group members have nowhere else to go and, thus, the associated cost of being ostracized from a system of generalized exchange within a group is extremely high (Greif 1989; Yamagishi 1998; Yamagishi et al. 1999). Similarly, environmental cues that activate the SEH may be different from society to society, reflecting the dominant form of social exchange in each society. For example, in a recent experiment Hayashi et al. (1999) demonstrated that expectations of reciprocal cooperation from one's partner in the PDG come from different sources in different societies. These expectations were more strongly based on generalized exchanges among American participants and direct exchanges among

Japanese participants. Whether or not the SEH is a product of evolution, illustrating how the SEH operates in different societies is, in and of itself, an important endeavor.

Acknowledgments

The research reported in this article has been supported by grants from the Japan Society for the Promotion of Science. We thank colleagues at Hokkaido University who helped us recruit potential participants from their classes.

NOTES

1. What Messick and Kramer (2001) call 'shallow morality' also shares many important aspects with the SEH. However, the error management aspects are more explicit in the SEH than in the 'shallow morality' argument.
2. In fact, Haselton and Buss (2000: 90) themselves note that the adaptationist logic used in EMT is applicable to other domains.
3. All participants were given instructions that explained both types of games. This question was asked after they had been told which game they were to play, but before they actually made the decision.
4. Since, in reality, the participants did not play against anyone, each participant received the payoff assuming that she had played against herself (800 yen if she cooperated, 400 yen if she defected).
5. The SEH may be conceived of as a special form of framing effect, in which the game situation is framed as an instance of social exchange rather than a game in which the goal is to win a competition.
6. The Japanese word *aite* is used to refer to 'partner'. *Aite* in Japanese does not have a positive connotation like the word 'partner' does in English. It is a neutral word used to describe the person who is at the other end of a relationship.

REFERENCES

Alcock, J. E. and D. Mansell. 1977. 'Predisposition and Behavior in a Collective Dilemma.' *Journal of Conflict Resolution* 21: 443–57.

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

Bargh, J. A. and T. L. Chartrand. 1999. 'The Unbearable Automaticity of Being.' *American Psychologists* 54: 462–79.

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

Cho, K. and B. Choi. 1999. 'A Cross-Society Study of Trust and Reciprocity: Korea, Japan and the U.S.A.' Paper presented at the WOW II, Workshop for Political Theory and Policy Analysis, Indiana University, June 16–19.

Davis, D. D. and C. A. Holt. 1993. *Experimental Economics.* Princeton, NJ: Princeton University Press.

Davis, M. H. 1983. 'Measuring Individual Differences in Empathy: Evidence for a Multidimensional Approach.' *Journal of Personality and Social Psychology* 44: 113–26.

Dawes, R. M., J. McTavish, and H. Shaklee. 1977. 'Behavior, Communication, and Assumptions about Other People's Behavior in a Commons Dilemma Situation.' *Journal of Personality and Social Psychology* 35: 1–11.

Emerson, R. M. 1976. 'Social Exchange Theory.' *Annual Review of Sociology* 2: 335–62.

Fisk, S. T. and S. E. Taylor. 1991. *Social Cognition* (2nd edn). New York: McGraw Hill.

Fox, J. and M. Guyer. 1978. '"Public" Choice and Cooperation in N-Person Prisoner's Dilemma.' *Journal of Conflict Resolution* 22: 469–81.

Gigerenzer, G. and P. M. Todd. eds. 1999. *Simple Heuristics That Make Us Smart*. New York: Oxford University Press.

Greif, A. 1989. 'Reputation and Coalitions in Medieval Trade: Evidence on the Maghribi Traders.' *Journal of Economic History* 49: 857–82.

Haselton, M. G. and D. M. Buss. 2000. 'Error Management Theory: A New Perspective on Biases in Cross-Sex Mind Reading.' *Journal of Personality and Social Psychology* 78: 81–91.

Hayashi, N., E. Ostrom, J. Walker, and T. Yamagishi. 1999. 'Reciprocity, Trust, and the Sense of Control: A Cross-Societal Study.' *Rationality and Society* 11: 27–46.

Karp, David, Nobuhito Jin, Hiromi Shinotsuka, and Toshio Yamagishi. 1993. 'Raising the Minimum in the Minimal Group Paradigm.' *Japanese Journal of Experimental Social Psychology* 32: 231–40.

Kiyonari, T., S. Tanida, and T. Yamagishi. 2000. 'Social Exchange and Reciprocity: Confusion or Heuristic?' *Evolution and Human Behavior* 21: 411–27.

Kollock, P. 1997. 'Transforming Social Dilemmas: Group Identity and Cooperation.' In *Modeling Rational and Moral Agents*, ed. P. Danielson, pp. 186–210. Oxford: Oxford University Press.

McCabe, K. A. 2003. 'A Cognitive Theory of Reciprocal Exchange.' In *Trust and Reciprocity: Interdisciplinary Lessons from Experimental Research*, eds E. Ostrom and J. Walker, pp. 147–69. New York: Russell Sage Foundation.

Marwell, G. and R. E. Ames. 1979. 'Experiments on the Provision of Public Goods. I: Resources, Interest, Group Size, and the Free-Rider Problem.' *American Journal of Sociology* 84: 1335–60.

Mehrabian, A. and N. Epstein. 1972. 'A Measure of Emotional Empathy.' *Journal of Personality*, 40: 525–43.

Messick, D. M. 1999. 'Alternative Logics for Decision Making in Social Settings.' *Journal of Economic Behavior and Organization* 39: 11–28.

Messick, D. M. and R. M. Kramer. 2001. 'Trust as a Form of Shallow Morality.' In *Trust in Society*, ed. K. S. Cook, pp. 89–118. New York: Russell Sage Foundation.

Milinski, M., D. Semmann, and H. J. Krambeck. 2001. 'Reputation Helps Solve the Tragedy of the Commons.' *Nature* 415: 424–6.

Morris, M. W., L. H. Sim, and V. Girotto. 1998. 'Distinguishing Sources of Cooperation in the One-Round Prisoner's Dilemma: Evidence for Cooperative Decisions Based on the Illusion of Control.' *Journal of Experimental Social Psychology* 34: 494–512.

Nowak, M. A. and K. Sigmund. 1998. 'The Dynamics of Indirect Reciprocity.' *Journal of Theoretical Biology* 194: 561–74.

Orbell, J. M. and R. M. Dawes. 1991. 'A "Cognitive Miser" Theory of Cooperators' Advantage.' *American Political Science Review* 85: 515–28.

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

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

Rilling, J. K., D. A. Gutman, T. R. Zeh, G. Pagnoni, G. S. Berns, and C. D. Kilts. 2002. 'A Neural Basis for Social Cooperation.' *Neuron* 35: 395–405.

Sato, K. and T. Yamagishi. 1986. 'Psychological Factors in the Public Goods Problem: Free-Riding and the Lack of Trust.' *Japanese Journal of Experimental Social Psychology* 26: 89–95 (in Japanese with an English abstract).

Shafir, E. and A. Tversky. 1992. 'Thinking through Uncertainty: Nonconsequential Reasoning and Choice.' *Cognitive Psychology* 24: 449–74.

Tanida, S. and T. Yamagishi. 2004. 'The Effect of Empathy on Accuracy in the Prediction of Behavior in Social Exchange.' *The Japanese Journal of Psychology*, 74: 148–55 (in Japanese with an English abstract).

Watabe, M., S. Terai, N. Hayashi, and T. Yamagishi. 1996. 'Cooperation in the One-Shot Prisoner's Dilemma Based on Expectations of Reciprocity.' *Japanese Journal of Experimental Social Psychology* 36: 183–96 (in Japanese with an English abstract).

Weber, J. M., S. Kopelman, and D. 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.

Wedekind, C. and M. Milinski. 2000. 'Cooperation through Image Scoring in Humans.' *Science* 288: 850–2.

Yamagishi, T. 1986. 'The Provision of a Sanctioning System as a Public Good.' *Journal of Personality and Social Psychology* 51: 110–16.

Yamagishi, T. 1988a. 'Seriousness of Social Dilemmas and the Provision of a Sanctioning System.' *Social Psychology Quarterly* 51: 32–42.

Yamagishi, T. 1988b. 'The Provision of a Sanctioning System in the United States and Japan.' *Social Psychology Quarterly* 51: 265–71.

Yamagishi, T. 1998. *The Structure of Trust*: *The Evolutionary Games of Mind and Society*. Tokyo: Tokyo University Press (in Japanese; English translation is available at http://lynx.let.hokudai.ac.jp/members/yamagishi/).

Yamagishi, T. and T. Kiyonari. 1997. 'Playing a Prisoner's Dilemma as an Assurance Game: Matrix Transformation and Production of Trust.' Paper presented at the Russell Sage Trust Conference, November 14–16, New York.

Yamagishi, T. and T. Kiyonari. 2000. 'The Group as the Container of Generalized Reciprocity.' *Social Psychology Quarterly* 63: 116–32.

Yamagishi, T. and K. Sato. 1986. 'Motivational Bases of the Public Goods Problem.' *Journal of Personality and Social Psychology* 50: 67–73.

Yamagishi, T., N. Jin, and T. Kiyonari. 1999. 'Bounded Generalized Reciprocity: Ingroup Favoritism and Ingroup Boasting.' *Advances in Group Processes* 16: 161–97.

TOSHIO YAMAGISHI is director of the Center for Experimental Research in Social Sciences and professor of social psychology at Hokkaido University, Japan. His current research interests include co-evolution of social institutions and adaptive psychological mechanisms. The topics of his recent publications have been cooperation in social dilemmas, punishment as a means to promote cooperation, culture as a

self-sustaining system of beliefs, trust and culture, group-based trust, and direct reciprocity and generalized reciprocity.

ADDRESS: Graduate School of Letters, Hokkaido University, N10 W7 Kita-ku, Sapporo, Japan 060–0810 [email: toshio@let.hokudai.ac.jp].

SHIGERU TERAI is a post-doctoral research fellow in the Department of Behavioral Science, Hokkaido University. He has published papers on development of trust in a selective-play environment.

TOKO KIYONARI is a post-doctoral researcher in the Department of Management, University of Antwerp, Belgium. Her research focus is on cooperation and sanctions in social dilemmas, trust, and inter-group conflict. She aspires to combine various approaches to human cooperation, including those used in social psychology, sociology, behavioral economics, evolutionary psychology, and evolutionary biology.

NOBUHIRO MIFUNE is a graduate student in the Department of Behavioral Science, Hokkaido University. He is currently working on a research paper about an experiment in which dictators in a dictator game gave more money to an ingroup recipient than an outgroup recipient when the group membership of the two players was common knowledge. When only the dictator knew the group membership of the recipient, he/she did not favor an ingroup recipient.

SATOSHI KANAZAWA is Reader in Management and Research Methodology at the London School of Economics and Political Science. His work on evolutionary psychology has appeared in peer-reviewed journals in all the social sciences (psychology, sociology, economics, political science, and anthropology) and in biology.