We Can't Return Evil for Good: The Comparison Between Direct and Indirect Reciprocity

Yuma Shiraki[1, 2, 3, *], Tasuku Igarashi[1]

[1]Graduate School of Education and Human Development, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8601, Japan

[2]Japan Society for the Promotion of Science, Tokyo, Japan

[3]National University of Singapore, Singapore

*Author for correspondence (shiraki.yuma@nagoya-u.jp)

There are two distinct evolutionary mechanisms of altruistic behavior: the disposition to return a benefit to a former benefactor (direct reciprocity) and the disposition to discriminate partners of social interaction based on their reputation (indirect reciprocity). Humans are motivated not only to reciprocate benefits to benefactors but also to behave altruistically toward a cooperative person, not an uncooperative person. This study compared the two evolutionary mechanisms of altruistic behavior. Three scenario-based experiments on diverse samples (Japanese undergraduates in Experiment 1, Japanese crowdsourcing workers in Experiment 2, and crowdsourcing workers worldwide in Experiment 3) were conducted by manipulating (1) reciprocity between participants and a colleague (reciprocal or non-reciprocal) and (2) the colleague's reputation in the workplace (good or bad). When participants received a reciprocal request from their colleague to help, they tended to accept it, even if the colleague had a bad reputation among others. On the other hand, participants were less accepting of a non-reciprocal request from a colleague with a bad reputation than a colleague with a good reputation. These results clearly indicate that people do not rely on the partner's reputation when they have a direct reciprocal relationship with the partner. In other words, humans prioritize the maintenance of direct reciprocal relationships over partner discrimination.

Keywords
direct reciprocity, indirect reciprocity, reputation

Introduction

Humans help others, as can be seen in such examples as donating money to charity or helping colleagues with their tasks. This is so-called altruistic behavior in which we impose some costs on ourselves in order to benefit others (Hamilton, 1963). Although altruistic behavior can be seen in daily life, it is not adaptive from the viewpoint of evolution. Why have we evolved with such a non-adaptive behavioral system?

There are two standard explanations for this question: direct and indirect reciprocity (Nowak, 2006). Direct reciprocity is characterized as a situation where "A helps B, then B helps A." With direct reciprocal relationships, altruistic behavior can evolve because it can become its own benefit when repaid by the beneficiary. Trivers (1971) referred to this mutual cooperation as reciprocal altruism and highlighted the role of emotion in its process. For example, gratitude is assumed to have evolved to maintain the reciprocal relationship between a benefactor and a beneficiary. Although direct reciprocity explains the evolution of altruistic behavior in close relationships, such as between friends, it is not clear why our behavior can also be altruistic toward strangers.

The mechanism of indirect reciprocity clarifies this issue. Indirect reciprocity is expressed by a situation such that "A helps B, then C helps A," where an altruistic behavior will be repaid not by the beneficiary, but by strangers. In this framework, reputations play a pivotal role on the stabilization of the structure. People share a strong belief, which is called a reputation-making norm, to cooperate selectively with someone (e.g., Mashima & Takahashi, 2005). In order to maintain the indirect reciprocity mechanism, we need to discriminate partners of social interaction based on their reputation and cooperate not with a "bad" person who deceived others but with a "good" person who behaved altruistically toward others.

In these ways, direct and indirect reciprocity explains the evolution of the different types of altruistic behavior. However, it is still unclear which reciprocity is more important for humans in a situation where both reciprocal mechanisms conflict. On one hand, it is true that we can achieve both of them at the same time. For example, where all members in a group share reputations of others, their repayment toward someone who helped them does not cause any conflict because the benefactor is regarded as a "good" person among all members. On the other hand, when members in a group do not observe interactions among others, a conflict can occur. For instance, when a bad man who is uncooperative toward all other members helps someone in the absence of others, his reputation remains bad among others. At this time, if the benefactor asks the beneficiary for help in front of the others, the beneficiary would hesitate to accept the offer, because the beneficiary is willing to accept the offer in order to reciprocate benefits to the benefactor, but the benefactor is judged as a "bad" person among others. Should the beneficiary accept the request from the benefactor to achieve reciprocal help (direct reciprocity), or should he/she discriminate partners based on their reputation and refuse the request from benefactor (indirect reciprocity)?

The Present Research

This study conducted three scenario-based experiments to examine whether humans prioritize the maintenance of a reciprocal relationship or the discrimination of partners when the different types of reciprocity are in conflict. To broaden the generalizability of findings, each experiment used completely different samples.

Experiment 1

(a) Method

Participants and design.
Japanese undergraduates (N = 154) voluntarily participated in an online experiment. The participants were randomly assigned to one of four conditions, 2 (reciprocity: benefactor vs. control) × 2 (reputation: good vs. bad) in a between-participants design. After the participants read a scenario where a colleague at a company asked for help, they rated how likely they were to accept the request.

Scenario and measure.
In the story, the participants were office workers in a department of a company. When participants were having lunch with their colleagues, "Person B," who had just been transferred to the department, approached them and asked his new colleagues to cover his night shift that day. Participants were neither familiar with Person B nor on intimate terms with each other.

We manipulated the reciprocity between participants and Person B (reciprocal or non-reciprocal) and Person B's reputation at the department (good or bad). In the reciprocal condition, participants had asked Person B to cover their night shift and Person B had accepted the request. In the non-reciprocal condition, participants had not received any help from him/her before. Person B's reputation among other colleagues was manipulated by telling participants that Person B was "cooperative and popular" in the good reputation condition or "uncooperative and unpopular" in the bad reputation condition.

After reading the scenario, participants rated how likely they were to cover Person B's night shift on a 6-point Likert scale, ranging from 1 (I would not cover) to 6 (I would cover).

(b) Results and Discussion

We excluded 35 participants who did not comprehend the scenario and only analyzed the data from 119 participants (83.19% female; M_{age} = 19.60, SD =1.16). The reciprocity and reputation manipulations were both effective; participants in the reciprocal condition were more likely to accept Person B's request (M = 5.34, SD = 0.85) than those in the non-reciprocal condition (M = 4.53, SD = 1.34), F (1, 115) = 17.53, p < .001, partial η^2 = .132. Moreover, participants in the good reputation condition were more likely to accept Person B's request (M = 5.21, SD = 0.83) than those in the bad reputation condition (M = 4.70, SD = 1.39), F (1, 115) = 6.64, p = .011, partial η^2 = .055.

These main effects were qualified by a significant reciprocity × reputation interaction, F (1, 115) = 5.90, p = .017, partial η^2 = .048 (see Figure 1). In the non-reciprocal condition, participants were more likely to accept the request from Person B with a good reputation than with a bad reputation, F (1, 115) = 12.66, p = .001, suggesting that

participants avoided selecting a bad person as a partner. On the other hand, Person B's reputation did not affect the acceptance of the request in the reciprocal condition, F (1, 115) = 0.04, p = .841. These results indicate that people do not rely on the partner's reputation when they have a direct reciprocal relationship with the partner. However, the generalizability of the current findings is limited because only Japanese undergraduates, who may be unfamiliar with a company situation, participated in the experiment. To resolve this issue, we conducted Experiment 2 to replicate the current findings with participants from a wide age range.
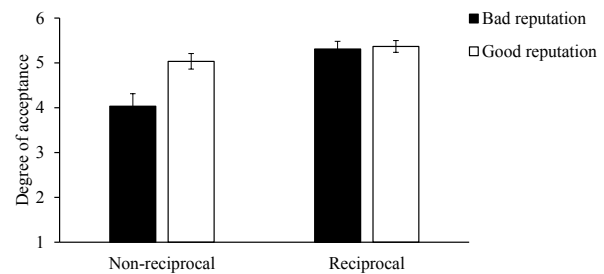

Figure 1. Acceptance of request from Person B (Study 1).

Experiment 2

(a) Method

Participants and design.
Japanese crowdsourcing workers (N = 300) recruited from Lancers participated in the experiment. The procedures were the same as Experiment 1. Each participant received 50 JPY ($0.43) as remuneration.

(b) Results and Discussion

We used the data from 281 participants (55.87% female; M_{age} = 38.95, SD = 9.93) who passed scenario comprehension questions. Both of the reciprocity and reputation manipulations were effective. Participants in the reciprocal condition were more accepting of Person B's request (M = 4.99, SD = 0.85) than those in the control condition (M = 3.73, SD = 1.31), F (1,277) = 112.9, p < .001, partial η^2 = .256. Participants in the good reputation condition were more accepting of Person B's request (M = 4.50, SD = 1.20) than those in the bad reputation condition (M = 4.21, SD = 1.33), F (1,277) = 5.82, p = .027, partial η^2 = .021. Furthermore, there was a significant reciprocity × reputation interaction, F (1, 277) = 4.96, p = .027, partial η^2 = .017 (see Figure 2). A good reputation increased the acceptance tendency of Person B's request in the control condition, F (1, 277) = 11.02, p = .020, but not in the reciprocal condition, F (1, 277) = 0.03, p = .882.

Experiment 2 extended the generalizability of the original findings in Experiment 1 by recruiting participants from a wide age range, from 19 to 73 years of age. However, these findings were obtained only from Japanese individuals. People in interdependent cultures tend to feel indebtedness (Shen, Wan, & Wyer, 2011), a negative emotion that promotes repayment behavior, more often than people in independent cultures. Thus, there may be a possibility that the application of these findings is limited to contexts of interdependent cultures, such as Japan. To reject this explanation, we conducted Experiment 3, recruiting participants from different cultures.
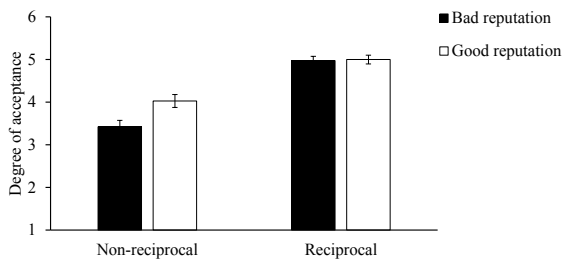
Figure 2. Acceptance of request from Person B (Study 2).

Experiment 3

(a) Method
Participants and design.
Participants were 400 crowdsourcing workers from 45 countries recruited from CrowdFlower. The material used in Experiment 3 was the same as Experiment 2, except for a slight change to the situation (covering a night shift in a hospital, rather than a company) which seems more relatable to people around the world. Each participant received $0.20 as remuneration.

(b) Results and Discussion
We used the data from 262 participants (34.35% female; M_{age} = 34.52, SD = 9.05) who passed the Instructional Manipulation Check (Oppenheimer, Meyvis, & Davidenko, 2009) and scenario comprehension questions. Both of the reciprocity and reputation manipulations were effective. Participants in the reciprocal condition accepted Person B's request (M = 5.54, SD = 0.69) more often than those in the non-reciprocal condition (M = 4.48, SD = 1.40), F (1, 258) = 63.50, p < .001, partial η^2 = .197. Participants in the good reputation condition were more likely to accept Person B's request (M = 5.23, SD = 0.90) than those in the bad reputation condition (M = 4.88, SD = 1.36), F (1, 258) = 6.85, p < .001, partial η^2 = .019. There was also a significant reciprocity × reputation interaction, F (1, 258) = 4.58, p = .033, partial η^2 = .017 (see Figure 3). Participants tended to accept the request from Person B based on his/her reputation in the non-reciprocal condition, F (1, 277) = 11.02, p = .020, but did not in the reciprocal condition, F (1, 277) = 0.03, p = .882. As in Experiments 1 and 2, participants gave back help to Person B, regardless of his/her reputation.

Overall, the successful replication of the findings that humans prioritize reciprocation over discrimination of partners suggests the generalizability of the mechanism to the global population.
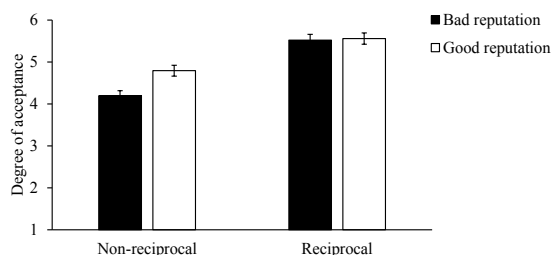

Figure 3. Acceptance of request from Person B (Study 3).

General Discussion
The current study examined whether humans prioritize the maintenance of reciprocal relationships with a partner or the discrimination of the partner based on the reputation. Throughout the three experiments including diverse samples, participants accepted a request from a "bad" benefactor as often as that from a "good" benefactor. The maintenance of a mutual bond with the benefactor comes before the reputation-based discrimination of him/her.

The difference in the mechanisms of direct and indirect reciprocity gives a possible explanation for the findings. Reputational indirect reciprocity is driven by cognitive mechanisms to calculate benefits and costs. A neuroimaging study revealed that a brain region related to cognitive calculations is activated when people judge whether to donate money to another person with his/her behavioral history in the presence of a third person (Watanabe et al., 2014). On the contrary, direct reciprocity is rooted in emotional mechanisms, such as feelings of gratitude (Trivers, 1971). Gratitude promotes paying-it-forward, or repayment behavior, from the beneficiary to a third person (Shiraki & Igarashi, 2016). Paying-it-forward is not rational, as it will not increase any adaptive fitness, but our emotional mechanisms could drive such irrational behavior beyond cognitive mechanisms. Considering these points, participants might have accepted the request from the bad benefactor due to the emotional mechanisms that drive (somewhat irrational) direct reciprocal behavior in this context.

There is a limitation in the current study. In the scenario, we did not control the reputations of the colleagues who were the targets of Person B's uncooperative behavior. If Person B had not cooperated with the colleagues based on their "bad" reputations, the uncooperative behavior could be justified and regarded as "good" behavior in some reputation making norms. The ambiguity of the second-order information makes it unclear how participants interpreted Person B's uncooperative behavior. Future research should incorporate the information to validate the generalizability of the current findings.

References
Hamilton, W. D. (1963). The evolution of altruistic behavior. The American Naturalist, 97, 354-356. (doi: 10.1086/497114)

Mashima, R., & Takahashi, N. (2005). The emergence of indirect reciprocity: Evolutionary foundation of altruistic behavior based on "strict discriminator." The Japanese Journal of Psychology, 76, 436-444. (doi: 10.4992/jjpsy.76.436) (in Japanese with an English abstract)

Nowak, M. A. (2006). Five rules for the evolution of cooperation. Science, 314, 1560-1563. (doi: 10.1126/science.1133755)

Oppenheimer, D. M., Meyvis, T., & Davidenko, N.

(2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. Journal of Experimental Social Psychology, 45, 867-872. (doi: 10.1016/j.jesp.2009.03.009)

Shen, H., Wan, F., & Wyer. R. S. Jr. (2011). Cross-cultural differences in the refusal to accept a small gift: The differential influence of reciprocity norms on Asians and North Americans. Journal of Personality and Social Psychology, 100, 271-281. (doi: 10.1037/a0021201)

Shiraki, Y., & Igarashi, T. (2016). What factors of prosocial behavior evoke recipients' gratitude and indebtedness? An experimental examination. The Japanese Journal of Psychology, 5, 474-484. (doi: 10.4992/jjpsy.87.15040) (in Japanese with an English abstract)

Trivers, R. L. (1971). The evolution of reciprocal altruism. The Quarterly Review of Biology, 46, 35-57. (doi: 10.1086/406755)

Watanabe, T., Takezawa, M., Nakawake, Y., Kunimatsu, A., Yamasue, H., Nakamura, M., … Masuda, N. (2014). Two distinct neural mechanisms underlying indirect reciprocity. Proceedings of the National Academy of Sciences of the United States of America, 111, 3990-3995. (doi: 10.1073/pnas.1318570111)