



ARTICLE

<https://doi.org/10.1057/s41599-018-0203-8>

OPEN

Effects of beliefs about sanctions on promoting cooperation in a public goods game

Hitoshi Yamamoto ¹ & Takahisa Suzuki²

ABSTRACT Cooperative behaviour among people facing social dilemmas remains an unsolved puzzle. Sanction systems such as punishment and reward are well-known solutions to social dilemmas. On the one hand, it is theorised that peer sanctions cannot maintain cooperation because of the intrusion of second-order free riders. On the other hand, experimental studies have widely reported that cooperation is sustainable by first-order sanction systems. To understand the divergence between theory and experimental results, we focus on the effects of beliefs about sanctions on promoting cooperation. While previous studies have revealed effects of beliefs about other people's cooperation in public goods games (PGGs), the effect of beliefs about sanctions have not been considered. We conducted a scenario-based experiment using one-shot PGGs with and without sanction systems. The results revealed that beliefs about sanctions promote cooperation in a PGG when types of sanctions and some psychological attitudes to cooperation are controlled for. Our results indicate that the beliefs of actors promote cooperation despite the possible presence of second-order free riders. A belief about sanctions differs from a preference for cooperative behaviour and the amount a player pays to exercise a sanction. It is necessary to consider players' beliefs about a sanction when studying its effectiveness.

¹ Faculty of Business Administration, Ritssho University, Tokyo, Japan. ² College of Policy Studies, Tsuda University, Tokyo, Japan. Correspondence and requests for materials should be addressed to H.Y. (email: hitoshi@ris.ac.jp)

Introduction

Altruistic behaviour, such as helping others and contributing to a public good, is observed in society despite a competitive environment. However, promoting and maintaining cooperation is difficult because those who exploit others' contributions could gain greater benefits than cooperators. Clarification of mechanisms to facilitate cooperative behaviour would stabilise social capital in a fluid society and have social significance.

Many ways to promote cooperative behaviour in relation to social dilemmas have been discussed. Many theoretical and experimental studies have investigated the evolution of cooperation (Nowak and Sigmund, 2005; Sigmund, 2011). There has been research on "reputation systems" (Dellarocas, 2003; Bakos and Dellarocas, 2011), such as those used on auction web sites. In addition, it has been suggested that cooperation evolves through the development of various types of network (Nowak, 2006). On the other hand, studies have examined the promotion of cooperation by imposing regulations or playing games. For example, third parties can conduct surveillance and signal when they observe certain behaviours or sanctions (Fehr et al., 2002; Fehr and Gächter, 2002; Sutter et al., 2010; Milinski and Rockenbach, 2011), and such sanctions can be meted out to participants in society depending on whether they cooperate (Bateson et al., 2006).

The effects of punishment and reward as sanctions on promoting cooperation have been verified. Opportunists can be induced to cooperate because cooperation can yield greater final benefits to cooperating players than to non-cooperating players under conditions where sanctions are possible. Some studies have reported that people tend to prefer rewards to punishments when they can choose a sanction system (Sefton et al., 2007; Ule et al., 2009) and they find rewards to be more effective (McCusker and Carnevale, 1995). However, other studies showed that punishment promotes cooperation more effectively than reward (Andreoni et al., 2003; Güreker et al., 2006; Sefton et al., 2007). Yet another study claimed that the combination of both sanctions is most efficient (Dong et al., 2016). A meta-analysis (Balliet et al., 2011) was conducted to analyse the difference in effectiveness between reward and punishment, and showed that both rewards and punishment had a significant positive effect on cooperation, although the differences were not statistically significant. It may be that these sanction systems have different effects on expectations of cooperation, beliefs about sanctions or ethical concerns over social dilemmas.

Theoretically, higher-order dilemmas occur in systems where sanctions exist. Thus, it is considered that cooperation cannot be achieved with only first-order sanctions because of the occurrence of second-order free riders (Axelrod, 1986; Galán and Izquierdo, 2005; Okada et al., 2015; Yamamoto and Okada, 2016). We define first/second-order free riders and first/second-order sanctions as follows. The first-order free riders are those who do not contribute to a public good. Punishments for first-order free riders and rewards for contributors are defined as first-order sanctions. Those who do not contribute to the first-order sanctions are second-order free riders. Additionally, punishments of those who do not impose a first-order punishment or rewards for those who have rewarded cooperators are defined as second-order sanctions.

The reason why higher-order dilemmas occur is as follows. If only first-order sanctions exist, cooperation temporarily increases in the population because it is beneficial; i.e., a cooperator can be rewarded or a non-cooperator punished. However, a sanctioner's payoff is less than that of a pure cooperator who cooperates in relation to the public good but does not participate in sanctions. In the first-order punishment case, cooperative sanctioners must pay the cost of punishment for non-compliers so that pure cooperators can enter the population. Such pure cooperators are second-order free riders because they shirk responsibility for

watching and punishing other players' actions and their payoffs are higher than those for cooperative punishers. These free riders impair the function of the punishment; therefore, the cooperative regime becomes fragile. In the first-order reward case, however, cooperative rewarders must pay the cost of rewarding all cooperators. As a corollary, the payoffs for cooperative rewarders are lower than those for pure cooperators who are free riders on the first-order rewards. These free riders impair the function of rewards; therefore, the cooperative regime becomes fragile.

As mentioned above, however, many studies have revealed that first-order sanction systems could promote cooperation. Furthermore, there have been studies showing that second-order dilemmas do not occur in punishment systems (Onoda et al., 2009) and that reward systems prevent second-order dilemmas (Kiyonari and Barclay, 2008).

Why is there such divergence between the theory and these experimental results? Many studies have focused on differences in the effects of punishment and reward to date; however, there has been almost no discussion of the mechanisms by which cooperation is achieved under first-order sanctions. Nevertheless, it is known that beliefs about the cooperative behaviour of others in a sanctionless public goods game (PGG) affect the players' cooperative behaviour (Fischbacher and Gächter, 2010). To verify the effectiveness of sanctions, it is necessary to verify the effects of the players' beliefs about sanctions. Such beliefs are defined in relation to the willingness of other players to pay the cost of sanctions. If beliefs about the behaviour of others affect cooperative behaviour, then beliefs about sanctions will have a similar effect. Focusing on punishment, (Onoda et al., 2009) conducted a fundamental analysis of beliefs about sanctions. They clarified that an opportunist who cooperates only when a sanction system exists overestimates the degree of punishment. Based on the view that beliefs about sanctions affect cooperative behaviour in both punishment and reward sanction systems, we derived the following hypothesis.

H.1: Beliefs about sanctions have a positive effect on cooperative behaviour.

Do beliefs differ from traditional psychological explanations for altruistic behaviour? It is known that altruistic behaviour is brought about by psychological attitudes such as reciprocity and generalised trust (Yamagishi and Yamagishi, 1994; Yamagishi and Kiyonari, 2000). It has also been observed concerning the exercise of sanctions that people who cooperate in a PGG pay a greater cost to impose a sanction (Bowles and Gintis, 2004; Fehr et al., 2002).

It is assumed that these attitudes are human personality characteristics formed over a long period. In contrast, beliefs have the characteristic of changing in each situation, because they are dependent on short-term experiences (Fischbacher and Gächter, 2010) and frames (Dufwenberg et al., 2011). Therefore, beliefs should affect the intention to cooperate, independent of reciprocity and generalised trust. If beliefs are fully explained by these attitudes, the effects of beliefs in H.1 should disappear when these attitudes are controlled for. Thus, we also derived H.2.

H.2: Even when generalised reciprocity and generalised trust are controlled for, beliefs about sanctions have a positive effect on cooperative behaviour.

Methods

Overview of experiments. To explore the effects of beliefs about sanctions, we developed a scenario-based experiment that uses a one-shot PGG.

We conducted two experiments (Experiments 1 and 2), which both consisted of one-shot PGGs with and without sanctions. The purpose of this study was to analyse the effects of beliefs about sanctions on the intention to cooperate. Therefore, to avoid the

effects of feedback from the behaviour of other players in the game, no feedback was given. Experiment 1 used a between-subjects design for sanctions. In addition to a PGG without sanctions, participants played a PGG with the sanction of punishment or reward. Experiment 2 used a within-subjects design. In addition to playing a PGG without sanctions, subjects played both a PGG with punishment and one with reward. For beliefs about sanctions, the players were asked to guess subjectively the average amount of money used by other players for the sanction. This value was used as the player's belief about the sanction.

Experiment procedures. In Experiment 1, participants were divided into groups of three and they played two PGGs: one with no sanction (N-PGG) and one with a sanction (S-PGG). In each game, 800 yen was distributed as capital to every player in the group. In the N-PGG, if one player gave 800 yen to the organiser, two times that amount, i.e., 1600 yen, would be evenly distributed to the other two players. Participants determined whether they would give capital to the organiser. In the S-PGG, 400 yen as capital for sanctions was distributed to every player in the group. Participants could pay the cost of sanctions up to a ceiling of 400 yen. In the punishment condition, double the amount of money paid was drawn from non-collaborators. In the reward condition, double the amount of money paid was given to collaborators. At the beginning of the S-PGG, participants were asked to answer the following questions: Will you cooperate in the S-PGG? Will you pay the cost of the sanction? If you pay, how much will you pay for it? How much would you think people participating in this experiment would pay?

In Experiment 2, participants would play three games: one with no sanction (N-PGG), one with punishment (P-PGG) and with reward (R-PGG). To eliminate order effects on the participants, the order of P-PGG and R-PGG was switched between the two classes of student participants. The other settings were the same as in Experiment 1.

At the end of both experiments, participants were asked to respond to four questions about generalised trust and reciprocity. The statements were as follows: I can trust most people; If they are trusted by others, most people trust other parties in the same way; I have confidence in distinguishing people that I can trust from those I cannot trust; When someone helps me, I also help someone else; I believe that good things eventually come back to me when I am kind to others.

Experiment 1 was conducted with two classes of undergraduate students (a total of 110 participants) at a university in Tokyo on October 23, 2015. Only 100 participants (of which 34 were female) gave valid responses. Those who did not answer or stopped answering in the middle of the games were excluded. Three participants who chose to cooperate without sanctions and not to cooperate with sanctions were also judged not to have understood the experimental situation, and their answers were regarded as invalid. The number of valid responses assigned to the punishment condition was 43, and 57 were assigned to the reward condition. The participants were told about the experiment when a lecture started, and the experiment was carried out in the last 20 min of the lecture.

Experiment 2 was conducted with two classes of undergraduate students (a total of 127) at a university in Tokyo on May 27, 2016. Only 96 participants gave valid responses. Those who did not answer or stopped answering in the middle of a game were excluded. The first class (46 participants) carried out the experiment in the order of N-PGG, P-PGG and R-PGG, referred to as "NRP". The other class (50 participants) carried out the experiment in the order of N-PGG, R-PGG and P-PGG, referred to as "NRP". Both experiments were approved by the ethics

committee of Risho University and were performed in accordance with the relevant guidelines and regulations. All participants gave written informed consent before participation and were debriefed upon completion of each experiment. They were permitted to stop participating in the experiment whenever they wished.

Results

Validation of procedures and settings. In the analysis of Experiment 1, to confirm that there was no difference in the original cooperation rate between the punishment and reward condition groups, the cooperation rate in the N-PGG was analysed (the cooperation ratio was 23.3% under the punishment condition, and 33.3% under the reward condition). The results showed no significant difference ($F(1, 98) = 1.199, p = .276, \eta^2 = .012$). Thus, it can be inferred that there was no difference in the cooperation rates between the two classes before the experiment.

In the analysis of Experiment 2, there was no difference between the two groups in the cooperation rate under the no-sanction condition. No significant difference was observed in the N-PGG in the order of NPR or NRP (13.0% NPR, 18.0% NRP, $F(1, 94) = .439, p = .509, \eta^2 = .005$). Thus, it can be inferred that there was no difference in the cooperation rate between the two groups before the experiment. Moreover, no significant difference was found in the P-PGG (the cooperation ratio was 52.2% under NPR, and 66.0% under NRP, $F(1, 94) = 1.897, p = .172, \eta^2 = .020$) and R-PGG (the cooperation ratio was 34.8% under NPR, and 40.0% under NRP, $F(1, 95) = .273, p = .602, \eta^2 = .003$). Similarly, there was no significant difference in the estimated sanction amount for P-PGG (NPR: mean = 182.6, SD = 116.1; NRP: mean = 198.0, SD = 125.3, $F(1, 94) = .388, p = .535, \eta^2 = .004$) or R-PGG (NPR: mean = 143.5, SD = 136.1; NRP mean = 154.0, SD = 99.4, $F(1, 94) = .189, p = .665, \eta^2 = .002$). Because no effects attributable to the sanction order could be determined, the subsequent analysis was carried out on combined NPR and NRP data.

In Experiment 1, the N-PGG and S-PGG were designed to be repeated measures within participants. Punishment and reward designs were assigned between participants. In Experiment 2, the N-PGG, P-PGG and R-PGG were designed to be repeated measures within participants. Where appropriate, the Chi-Muller correction for non-sphericity was applied to the analyses.

Basic analysis on sanction systems. We first analysed the effects of sanction systems and observed beliefs about sanctions, which is the essential variable for our paper. To test whether punishment or reward has the effect of maintaining cooperation in a dilemma situation, a two-way analysis of variance (ANOVA) of the presence or absence of sanction and the type of sanction (punishment/reward) was carried out in Experiment 1. The presence or absence of a sanction had a significant main effect ($F(1, 98) = 51.011, p = .000, \eta^2 = .096$), and its interaction effect was also significant ($F(1, 98) = 12.404, p = .001, \eta^2 = .023$) (Fig. 1a). These results revealed that the existence of a sanction increased the cooperation rate and the punishment type of sanction had a greater effect. For Experiment 2, the cooperation rates in the N-PGG, P-PGG, and R-PGG were compared ($F(1.82, 172.68) = 37.174, p = .000, \eta^2 = .136$) (Fig. 1b). Multiple comparisons using Holm's method revealed differences between the groups (N-PGG vs. R-PGG; $t(95) = 5.158, p = .000, d = .510$, R-PGG vs. P-PGG; $t(95) = 3.789, p = .000, d = .45$). The cooperation rates were ranked N-PGG < R-PGG < P-PGG, and the punishment condition showed the highest cooperation rate.

Subsequently, to analyse the fundamental characteristics of beliefs about sanctions, we analysed the estimated rewards and punishments (Fig. 1c, d). Punishment had the highest estimated

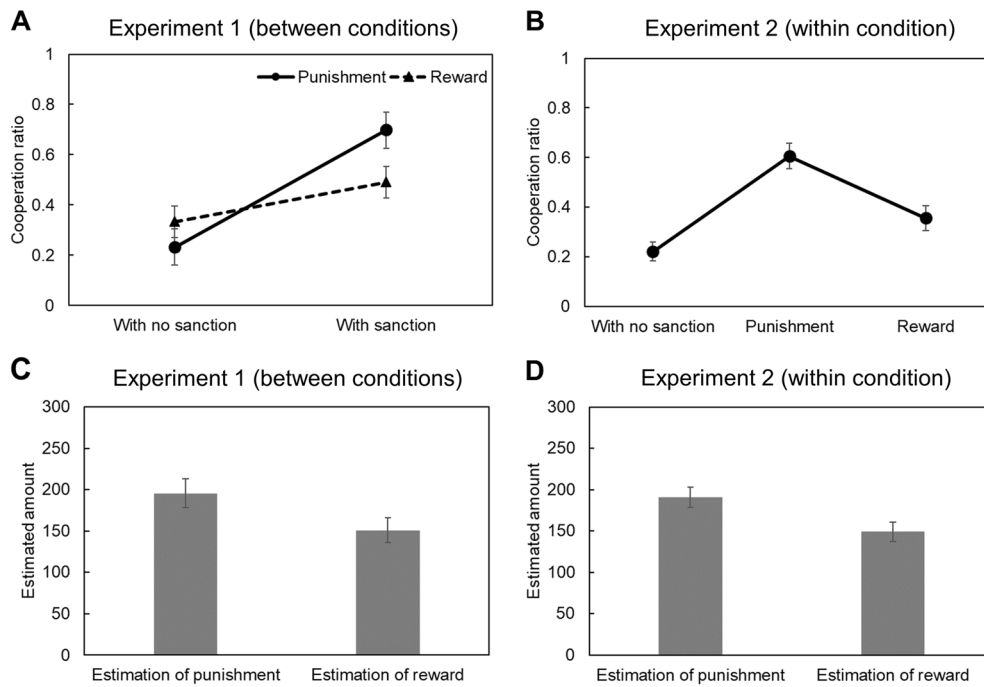


Fig. 1 Cooperation ratio and estimation of sanctions. The upper panels show the cooperation ratio with no sanction and with sanctions. The cooperation ratio represents the ratio of the number of participants who chose to contribute to public goods. Panel **a** shows the results for the between-subjects design (Experiment 1). Panel **b** shows the results for the within-subjects design (Experiment 2). The cooperation ratio for the punishment condition was high. The lower panels show the estimated punishments and rewards. Panel **c** shows the results of Experiment 1, and panel **d** shows those of Experiment 2. Both sets of results show that the estimated sanction is higher for punishments than for rewards. The error bars in each panel show the standard error

Table 1 Predictors of cooperation in logistic regression

	Experiment 1		Experiment 2	
	Estimate	SE	Estimate	SE
Type of sanctions (punishment = 0, reward = 1)	-1.464**	.535	-1.131**	.345
Estimated amount of sanction	.004 ⁺	.002	.005***	.001
Intercept	1.092	.962	-7.709*	.337
N	71		167	
AIC	88.718		208.470	

⁺p < .1. *p < .05. **p < .01. ***p < .001

sanction in both experiments: Experiment 1 ($F(1, 98) = 3.703, p = .057, \eta^2 = .036$) and Experiment 2 ($F(1, 95) = 6.671, p = .011, \eta^2 = .030$).

Effects of beliefs about sanctions. Most people who cooperated under the sanctionless situations also cooperate under conditions of sanctions. This situation is unsuitable for verifying the effects of beliefs. Therefore, we verified hypotheses limited to non-cooperating people in the N-PGGs. The stronger their beliefs (i.e., they guessed a higher amount for sanctions imposed by others), the more non-cooperating people in N-PGGs showed intent to cooperate in the S-PGGs.

A logistics regression analysis was conducted to verify H.1 (Table 1). Dummy variables and the amounts of sanctions imposed were the inputs for the types of sanctions to predict intention to cooperate in the S-PGGs. Experiment 2 used robust standard errors because it was a within-subjects design. The results showed that the effects of both types of sanctions were significant, and that punishment brought about greater intention

to cooperate. Concerning the effects of beliefs, the tendency was significant in Experiments 1 and 2, and the greater a player's estimation of the sanctions imposed by others, the greater was his or her intention to cooperate. Although the sample size in Experiment 1 was small, and the effect was not significant at the 5% level, it had an estimated value similar to that of Experiment 2. Thus, it can be said that H.1 was supported.

To test H.2, generalised trust and reciprocity were introduced into the regression. To measure generalised trust, subjects were asked to rate the following two items: "Most people are trustworthy" and "Most people will respond in kind when they are trusted by others". A subject's score on generalised trust was obtained by simple addition of the two items' scores ($\alpha = .571$, avg. = 5.330, SD = 1.897 for Experiment 1; $\alpha = .311$, avg. = 5.302, SD = 1.724 for Experiment 2).

We treat reciprocity as a concept similar to Putnam's "generalised reciprocity", which he describes as "I'll do this for you without expecting anything specific back from you in the confident expectation that someone will do something for me down the road" (Putnam, 2000). This idea can be interpreted as "The good you do for others is the good you do yourself". After completing the experiments, two statements on a questionnaire were used to measure the subjects' level of reciprocity. Their overall score was obtained by a simple addition of the scores of the responses to the two statements. The two items used to calculate reciprocity were: "If I'm helped by someone, I'll help myself or someone else", and "If I'm kind to people, eventually something good will happen to me" ($\alpha = .526$, avg. = 7.290, SD = 2.114 for Experiment 1; $\alpha = .700$, avg. = 6.979, SD = 2.262 for Experiment 2).

A logistic regression analysis that introduced generalised trust and reciprocity was conducted (Table 2). When generalised trust and reciprocity were controlled for, the effect of belief was still significant and H.2 was also supported.

Table 2 Predictors of cooperation controlled for trust and reciprocity in logistic regression

	Experiment 1		Experiment 2	
	Estimate	SE	Estimate	SE
Type of sanctions (punishment = 0, reward = 1)	-1.361*	.587	-1.136**	.352
Estimated amount of sanction	.005+	.002	.005***	.001
Trust	.206	.186	-.078	.108
Reciprocity	.436*	.193	.138	.087
Intercept	-3.583+	2.028	-1.292+	.700
N	71		167	
AIC	82.634		209.830	

+*p* < .1. **p* < .05. ***p* < .01. ****p* < .001

However, only the effects of reciprocity were observed in Experiment 1. Because the subjects of these analysis were limited to non-cooperating participants in the N-PGGs, it is possible that the effects of reciprocity and trust themselves were not significant.

Discussion

Much research has focused on the effects of sanction systems in resolving social dilemmas. In particular, a variety of analyses have been conducted on the differences between the effects of punishments and rewards. However, there has been little discussion on the effects of beliefs about sanctions.

We conducted a study using the framework of a PGG to focus on the existence of sanction systems of punishment and reward and the effects that players' beliefs about each type of sanction have on cooperation. The results showed first that in a one-shot PGG, punishment resulted in a greater intention to cooperate compared with reward. We believe the reason to be that beliefs about punishment were greater than those about reward. Even after psychological attitudes such as reciprocity and trust were controlled for, beliefs about sanctions influenced the intention to cooperate and we learned that the effect was independent of psychological factors such as reciprocity and trust. These results reveal that opportunists who do not cooperate under sanctionless conditions but do when sanctions are in place tend to have strong beliefs about sanctions.

A belief about a sanction is a different factor from the preference for cooperative behaviour and the amount the subject pays to exercise a sanction. It is necessary to consider players' beliefs about the sanction when studying its effectiveness.

Beliefs that sanctions are effective suggest that simply signalling the existence of a sanction system can maintain cooperative behaviour. For example, the findings of an experiment in which simply displaying a face photo increases cooperative behaviour (Bateson et al., 2006) is probably a side effect of this phenomenon. Because exercising sanctions incurs costs, it is difficult to monitor all non-cooperative behaviour and reward all cooperative behaviour. Under such conditions, even if the possibility that a sanction will be imposed is low, it is possible to increase the intention to cooperate by priming the belief about the sanction. This can be done by signalling the existence of the sanction system. For example, train conductors in Europe do not check the tickets of all passengers. However, a large fine is imposed if a passenger does not have the correct ticket, even though the probability of being checked is small. This system can be considered as a social measure using beliefs about sanctions. The results of our study can be said to verify the effectiveness of such a system with experiments.

In this study, we did not consider mechanisms formed by beliefs. For example, a study about changes in beliefs in repeated

games that do not consider sanctions has shown that beliefs change because of the amounts contributed by others in previous games (Fischbacher and Gächter, 2010). Researchers have also shown that different beliefs are formed depending on the frame of the game (Dufwenberg et al. 2011). How beliefs are formed by different mechanisms will be a critical topic in future. It will be necessary to study what differences exist between beliefs about punishment and beliefs about reward. Additionally, the experiments used one-shot PGGs to study only the effects of beliefs. Feedback about the results of the games was not given to players. However, critical questions remain about how beliefs about sanctions change as the games are repeated, and what the effects of such changes on cooperative behaviour are. As a future extension of this study, we will analyse belief changes in repeated games.

Data availability

The data supporting the findings of this study are stored in a Dryad data package titled 'Data of the effects of beliefs about sanctions on promoting cooperation in a public goods game' (Yamamoto, 2018), which can be accessed at: <https://osf.io/z78tj/>.

Received: 21 May 2018 Accepted: 23 November 2018

Published online: 11 December 2018

References

- Andreoni J, Harbaugh W, Vesterlund L (2003) The carrot or the stick: rewards, punishments, and cooperation. *Am Econ Rev* 93(3):893–902. <https://doi.org/10.1257/000282803322157142>
- Axelrod RM (1986) An evolutionary approach to norms. *Am Political Sci Rev* 80(4):1095–1111. <https://doi.org/10.1017/S0003055400185016>
- Bakos Y, Dellarocas C (2011) Cooperation without enforcement? A comparative analysis of litigation and online reputation as quality assurance mechanisms. *Manag Sci* 57(11):1944–1962. <https://doi.org/10.1287/mnsc.1110.1390>
- Balliet D, Mulder LB, Van Lange PAM (2011) Reward, punishment, and cooperation: a meta-analysis. *Psychol Bull* 137(4):594–615. <https://doi.org/10.1037/a0023489>
- Bateson M, Nettle D, Roberts G (2006) Cues of being watched enhance cooperation in a real-world setting. *Biol Lett* 2(3):412–414. <https://doi.org/10.1098/rsbl.2006.0509>
- Bowles S, Gintis H (2004) The evolution of strong reciprocity: cooperation in heterogeneous populations. *Theor Popul Biol* 65(1):17–28. <https://doi.org/10.1016/j.tpb.2003.07.001>
- Dellarocas C (2003) The digitization of word of mouth: promise and challenges of online feedback mechanisms. *Manag Sci* 49(10):1407–1424. <https://doi.org/10.1287/mnsc.49.10.1407.17308>
- Dong Y, Zhang B, Tao Y (2016) The dynamics of human behavior in the public goods game with institutional incentives. *Sci Rep* 6:28809. <https://doi.org/10.1038/srep28809>
- Dufwenberg M, Gächter S, Hennig-Schmidt H (2011) The framing of games and the psychology of play. *Games Econ Behav* 73(2):459–478. <https://doi.org/10.1016/j.geb.2011.02.003>
- Fehr E, Fischbacher U, Gächter S (2002) Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nat* 13(1):1–25. <https://doi.org/10.1007/s12110-002-1012-7>
- Fehr E, Gächter S (2002) Altruistic punishment in humans. *Nature* 415(6868):137–140. <https://doi.org/10.1038/415137a>
- Fischbacher U, Gächter S (2010) Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *Am Econ Rev* 100(1):541–56. <https://doi.org/10.1257/aer.100.1.541>
- Galán JM, Izquierdo LR (2005) Appearances can be deceiving: lessons learned re-implementing Axelrod's 'evolutionary approach to norms'. *J Arti Societ Soc Simulat* 8(3):2
- Gürerk Ö, Birlenbusch, Rockenbach B (2006) The competitive advantage of sanctioning institutions *Science* 312(5770):108–111. <https://doi.org/10.1126/science.1123633>
- Kiyonari T, Barclay P (2008) Cooperation in social dilemmas: Free riding may be thwarted by second-order reward rather than by punishment. *J Personal Social Psychol* 95(4):826–42. <https://doi.org/10.1037/a0011381>
- McCusker C, Carnevale PJ (1995) Framing in resource dilemmas: Loss aversion and the moderating effects of sanctions. *Organ Behav Human Decis Process* 61(2):190–201. <https://doi.org/10.1006/obhd.1995.1015>

- Milinski M, Rockenbach B (2011) On the interaction of the stick and the carrot in social dilemmas. *J Theor Biol* 299:139–143. <https://doi.org/10.1016/j.jtbi.2011.03.014>
- Nowak M (2006) Five rules for the evolution of cooperation. *Science* 314(5805):1560–1563. <https://doi.org/10.1126/science.1133755>
- Nowak M, Sigmund K (2005) Evolution of indirect reciprocity. *Nature* 437(7063):1291–1298. <https://doi.org/10.1038/nature04131>
- Okada I, Yamamoto H, Toriumi F, Sasaki T (2015) The effect of incentives and meta-incentives on the evolution of cooperation. *PLoS Comput Biol* 11(5): e1004232. <https://doi.org/10.1371/journal.pcbi.1004232>
- Onoda R, Matsumoto Y, Jin N (2009) Overestimation of the norm as a factor for promoting cooperation in social dilemmas. Center for Experimental Research in Social Science Working Paper Series 92
- Putnam R (2000) *Bowling alone*. Simon & Schuster, New York
- Sefton M, Shupp R, Walker JM (2007) The effect of rewards and sanctions in provision of public goods. *Econ Inq* 45(4):671–690. <https://doi.org/10.1111/j.1465-7295.2007.00051.x>
- Sigmund K (2011) Moral assessment in indirect reciprocity. *J Theor Biol* 299(2):25–30. <https://doi.org/10.1016/j.jtbi.2011.03.024>
- Sutter M, Haigner S, Kocher M (2010) Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations. *Rev Econ Stud* 77(4):1540–1566. <https://doi.org/10.1111/j.1467-937X.2010.00608.x>
- Ule A, Schram A, Riedl A, Cason TN (2009) Indirect punishment and generosity toward strangers. *Science* 326(5960):1701–1704. <https://doi.org/10.1126/science.1178883>
- Yamagishi T, Yamagishi M (1994) Trust and commitment in the United States and Japan. *Motiv Emot* 18(2):129–166. <https://doi.org/10.1007/BF02249397>
- Yamagishi T, Kiyonari T (2000) The group as the container of generalized reciprocity. *Social Psychol Q* 63(2):116–132. <https://doi.org/10.2307/2695887>
- Yamamoto H, Okada I (2016) How to keep punishment to maintain cooperation: Introducing social vaccine. *Phys A: Stat Mech its Appl* 443:526–536. <https://doi.org/10.1016/j.physa.2015.08.053>
- Yamamoto H (2018) Data of effects of beliefs about sanctions on promoting cooperation in a public goods game [WWW Document]. <https://osf.io/z78tj/> [Accessed 22 Nov 2018]

Acknowledgements

HY acknowledges Grants-in-Aid for Scientific Research 15KT0133, 16H03120, 17H02044, and 18H03498.

Author contributions

HY and TS developed the concept and discussed the results; HY and TS analysed the data; and HY wrote the paper.

Additional information

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://www.nature.com/reprints>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018