

1 *Journal of Experimental Social Psychology, 2020*

2 **DOI: 10.1016/j.jesp.2020.104021**

3

4

5

6

7

8 **Altruism does not always lead to a good reputation: A normative explanation**

9 (author version)

10

11 Yuta Kawamura<sup>1, 2</sup>, Takashi Kusumi<sup>3</sup>

12

13

14 <sup>1</sup> Department of Psychology, Kobe University, Japan.

15 <sup>2</sup> Japan Society for the Promotion of Science, Tokyo, Japan

16 <sup>3</sup> Graduate School of Education, Kyoto University, Japan.

17

18 Correspondence should be addressed to:

19 Yuta Kawamura

20 Department of Psychology, Kobe University

21 1-1 Rokkodai-cho, Nada-ku, Kobe 657-8501, Japan

22 E-mail address: kawamura@lit.kobe-u.ac.jp

23

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14

### **Abstract**

Individuals who engage in altruistic behaviors generally acquire a good reputation. However, recent studies have suggested that altruists are not always welcomed by others. We examined the possibility that norm-deviant altruism leads to unfavorable evaluations; distributing quite large amounts of one's resources could be less favored because the behavior deviates from social norms. In four studies, participants rated their feelings (i.e., liking and respect) toward a person who distributed his/her resources to others. We found that altruistic behavior that deviates from social norms was less favorably regarded than modestly altruistic behavior (Study 1–4), specifically in a culture with low tolerance for norm deviation (Japan; Study 3) and especially when the degree of the deviance was high (Study 2). These findings suggest that altruistic behavior is less favored when the behavior deviates from social norms and norm deviation is evaluated negatively.

*Keywords:* altruism, reputation, antisocial punishment, social norm, generosity, culture

## 1 Altruism Does Not Always Lead to a Good Reputation: A Normative Explanation

2 Altruistic behavior, such as distributing money to others, is generally favored by others.  
3 However, some studies have reported that altruists are not always welcomed but can be evaluated  
4 negatively (Parks & Stone, 2010) and even punished by others (Herrmann, Thoni, & Gächter,  
5 2008; Kuběna, Houdek, Lindova, Priplatova, & Flegr, 2014; Pleasant & Barclay, 2018).  
6 Recently, growing attention has been paid to this counterintuitive phenomenon (Herrmann et al.,  
7 2008; Irwin & Horne, 2013; Kuběna et al., 2014; Parks & Stone, 2010; Pleasant & Barclay,  
8 2018). There are two possible mechanisms: One is that altruistic behavior is unfavored when the  
9 behavior threatens the observers' reputation (e.g., Pleasant & Barclay, 2018), and the other is that  
10 evaluation of altruism is influenced by the extent to which the behavior deviates from others'  
11 typical behaviors (i.e., social norms; e.g., Herrmann et al., 2008). The present study  
12 discriminated between these two to focus on the latter. Specifically, we examined whether an  
13 actor who gives away a large amount of money (i.e., altruists who deviate from social norms) is  
14 less favored by others despite performing a generous act.

### 15 **Unfavorable Evaluations of Altruism**

16 Altruism often leads to a good reputation. People typically praise others' altruistic acts  
17 and like altruistic individuals (Bereczkei, Birkas, & Kerekes, 2007, 2010; Hardy & Van Vugt,  
18 2006), while they punish and ostracize those who commit selfish deeds (Fehr & Gächter, 2002;  
19 Feinberg, Willer, & Schultz, 2014). Furthermore, generous individuals are preferred as social and  
20 romantic partners (e.g., Arnocky, Piche, Albert, Ouellette, & Barclay, 2017; Barclay & Willer,  
21 2007). However, generous behavior does not always lead to a good reputation. In some instances,  
22 people have been shown to dislike those who seem to be too moralistic (Minson & Monin, 2012;  
23 Monin, Sawyer, & Marquez, 2008). People may not only dislike such altruists but are also

1 willing to incur costs to punish them. Some studies have reported evidence for the existence of  
2 “antisocial punishment,” namely punishment for altruistic behavior (e.g., Herrmann et al., 2008).  
3 Antisocial punishment has typically been found in the public goods game paradigm. The public  
4 goods game, which has been commonly used across disciplines, involves the distribution of  
5 resources within a small group. In each trial, participants are given some seed money or tokens  
6 and are asked to decide how much they wish to contribute to public goods and how much they  
7 want to keep for themselves. When they contribute to public goods, their contribution is typically  
8 doubled and distributed among all members; the total gain of members is increased by each  
9 contribution, whereas the individual gain of the contributing member is reduced. Generally, if  
10 each member has a choice to punish others, they punish more selfish members, namely members  
11 who contribute to the public goods less than the punisher. However, people have also been  
12 shown to punish altruistic individuals in this task (e.g., Herrmann et al., 2008).

13         One plausible explanation as to why some forms of altruism are viewed unfavorably is  
14 because these behaviors deviate from social norms (e.g., Herrmann et al., 2008). There is indirect  
15 evidence to support this explanation. For example, Parks and Stone (2010) reported that some  
16 participants cited norm deviations as a reason for their negative evaluations of others’ large  
17 contributions in the public good game. Irwin and Horne (2013) showed that when intragroup  
18 behavioral variance was small in the public goods game (i.e., descriptive norms were strong),  
19 altruistic behavior that deviated from norms was more likely to be punished. Additionally,  
20 Kuběna et al. (2014) showed that antisocial punishment in the public goods game tended to be  
21 carried out against especially high contributors.

22         These studies suggest the possibility that unfavorable evaluations of altruism depend on  
23 the extent to which the behavior deviates from social norms; however, there are two limitations.

1 First, with few exceptions (Irwin & Horne, 2013), prior studies did not investigate a potential  
2 causal relationship between norm deviations and negative evaluations of altruism. Second, and  
3 more importantly, in the public goods game, used in the majority of research on antisocial  
4 punishment, the threat to others' reputations could contribute to the unfavorable evaluations of  
5 altruism. More concretely, an important feature of the public goods game is that participants are  
6 evaluated by each other; thus, the existence of an extreme altruist undermines the perceived  
7 altruism of other members. That is, negative attitudes toward extreme altruism in the public  
8 goods game could be caused by the threat to the evaluator's own reputation. Some studies  
9 support this notion. For example, Parks and Stone (2010) found that the most frequent  
10 justification participants cited for negative evaluations of extremely altruistic individuals was  
11 concern for their own reputation, although participants also cited normative reasons.

12 Additionally, Pleasant and Barclay (2018) showed that when people were competing with others  
13 about their reputation, antisocial punishment increased. Thus, negative evaluations of altruism in  
14 the public goods game can be explained by the threat to one's own reputation, and it is not clear  
15 whether the norm deviation itself matters.

16 To investigate whether unfavorable evaluations of altruistic behavior depend on the  
17 extent to which the behavior deviates from social norms, we considered a situation where  
18 evaluators were not concerned about how they were evaluated. Specifically, instead of the public  
19 goods game, we adopted another common paradigm, the dictator game (Kahneman, Knetsch, &  
20 Thaler, 1986), in which one person distributes money to the other. Participants in the present  
21 study read vignettes about people involved in such distribution and were asked to rate their  
22 feelings toward various distributors from a third-party perspective. Thus, their own evaluation of  
23 altruism was not threatened by others' extremely altruistic behavior. The distributor could choose

1 one of three actions: keep all the money (i.e., keeping-everything), share the money equally with  
2 the recipients (i.e., sharing-50/50), or give the full amount to the recipients (i.e., giving-away-  
3 everything). Giving-away-everything has not been observed as frequently as sharing-50/50 in  
4 past studies that used the dictator game (for a meta-analytic review, see Engel, 2011). Therefore,  
5 while giving away all the money is more altruistic than sharing the money equally based on the  
6 definition of altruistic behavior (i.e., more cost for oneself and more benefit for others), it can  
7 also be regarded as an action that deviates from the typical behavior of other people: that is,  
8 giving-away-everything deviates from social norms. If more positive evaluations are given to  
9 more altruistic behavior, it is reasonable to speculate that giving away all the money would earn  
10 better evaluations than would sharing the money equally. In contrast, if participants are  
11 uncomfortable with norm violation, the former behavior may be less favored than the latter, even  
12 though evaluators might not be concerned about their own reputation.

### 13 **Cultural Differences in Tolerance for Deviation**

14 To the best of our knowledge, only Klein and colleagues (Klein & Epley, 2014; Klein,  
15 Grossmann, Uskul, Kraus, & Epley, 2015) have examined the evaluations of various  
16 distributions in the dictator game. Although they did not suggest that sharing the money equally  
17 is favored more than giving away all the money, Klein et al. (2015) implied the possibility of  
18 cultural differences in evaluations, which could be due to the presence of cultural differences in  
19 social norms. A previous study revealed cultural differences in tolerance for norm deviation  
20 (Gelfand et al., 2011); some cultures are “tight” (i.e., low tolerance for deviation), whereas others  
21 are “loose” (i.e., high tolerance for deviation). With regards to evaluations of altruism, the  
22 frequency of antisocial punishment in a culture is related to its social norms regarding altruism  
23 (Herrmann et al., 2008). These cross-societal differences in antisocial punishment might also be

1 related to attitudes toward extremely generous distributors (Klein et al., 2015). To investigate the  
2 possibility that deviations from social norms regarding altruism causes unfavorable evaluations,  
3 we first examined the evaluations of altruism in Japan, which is known as a tight society (see  
4 Gelfand et al., 2011). Then, in Study 3, we investigated cultural differences in evaluations of  
5 altruism in by comparing a Japanese sample with a U.S. sample.

## 6 **Two Dimensions of Interpersonal Attitudes: Liking and Respect**

7 In the present study, we focused on liking and respect as evaluations of altruism. Liking  
8 and respect are regarded as two dimensions of interpersonal attitudes (e.g., Baryła, 2014;  
9 DeBono & Muraven, 2014; Wojciszke, Abele, & Baryła, 2009). Previous studies have suggested  
10 that liking and respect are associated with different aspects of the target person. While liking  
11 depends on the target's communal traits (i.e., aspects related to other-oriented attitudes), respect  
12 is influenced by the target's agency (i.e., aspects related to ability and social success; Wojciszke  
13 et al., 2009). Although deviant behavior may be viewed as uncoordinated, deviating from the  
14 norm is not always seen as an indication of a lack of competence or low status (Bellezza, Gino,  
15 & Keinan, 2014). Therefore, liking may be influenced by perceptions of deviance, and respect is  
16 less likely to be influenced by such perceptions.

## 17 **The Current Study**

18 The present study consisted of four experiments to investigate evaluations of altruism  
19 that deviate from social norms. The first study investigated liking and respect toward the three  
20 types of distribution: keeping-everything, sharing-50/50, and giving-away-everything. Study 2  
21 manipulated social norms in different forms and investigated whether the evaluations of altruism  
22 were moderated by social norms. Study 3 investigated whether unfavorable evaluations of  
23 altruism are moderated by cultural differences in tolerance for norm deviation; if unfavorable

1 evaluations are caused by norm deviations, they would be observed particularly in tight cultures.  
2 The final study examined alternative explanations. We report all measures, manipulations, and  
3 exclusions in each of these studies. The data and R code can be found at <https://osf.io/4qr9n/>

#### 4 **Study 1**

5 In Study 1, we examined feelings of liking and respect toward three types of  
6 distributors. The participants read one vignette about person A, who shared 10,000 JPY with a  
7 stranger. Then, they evaluated their feelings for each A, who distributed all, half, or none of  
8 his/her money. We compared the liking and respect ratings among the three conditions.

#### 9 **Methods**

10 **Participants and design.** We recruited 682 participants (357 men, 325 women) through  
11 an online research system (Cross Marketing Inc., Tokyo, Japan). The average age of the  
12 participants was 24.2 years ( $SD = 3.40$ ; range: 18–29). Half of the participants were university  
13 students, whereas the remainder were employed. Sample size was determined before any data  
14 analysis. The study used a one factor within-participants design (distribution type: keeping-  
15 everything, sharing-50/50, or giving-away-everything). A power analysis using G\*Power (Faul,  
16 Erdfelder, Lang, & Buchner, 2007) indicated that the required sample size was 526 to detect a  
17 small main effect (i.e.,  $f = .10$ ) in a one-way analysis of variance (ANOVA) with three repeated  
18 measures factors, with  $\alpha = .05$  and  $\beta = .80$ . The participants completed other questionnaires (e.g.,  
19 questionnaires about critical thinking, literacy, moral identity, and learning motivation) in  
20 addition to that used for this study, but we do not report those results here because the  
21 questionnaires served a different purpose<sup>1</sup>. This study was approved by the ethics committee of

---

<sup>1</sup> Answering these questionnaires might have affected participants' responses and thus our results. However, the effect does not differ between conditions because the distribution type was manipulated within participants. Furthermore, Studies 2–4, in which participants did not answer these questionnaires, replicated the results of Study 1.



1 the authors' institution.

2           **Procedure.** Participants read one short vignette about person A, who shared 10,000 JPY  
3 (100 JPY = approximately 1 USD) with one stranger. The vignette read as follows:

4           “Person A won a lottery and was supposed to receive 10,000 JPY. But later, the  
5 management of the lottery told A that since there were mistakenly two winners in the  
6 lottery, the management wished that A would help decide how to distribute the reward  
7 with the other winner. Since A happened to contact the management beforehand, A can  
8 decide how to divide the remuneration as A wants.

9           The following are examples of how to divide the 10,000 JPY:

10           1: A receives 10,000 JPY, and he/she does not give even 1 JPY to the other  
11 person

12           2: A gives 10,000 JPY to the other person without receiving even 1 JPY

13           3: A gets 5,000 JPY and equally gives 5,000 JPY to the other person

14           The other winner cannot refuse A's decision and there is no room for any negotiation  
15 following A's decision. Also, the other winner and A are strangers and they cannot know  
16 each other's personal information.”

17           This vignette was a modified version of the dictator game situation and better reflected a  
18 daily life situation. After reading the vignette, participants first answered two manipulation check  
19 questions; they were asked if one person obtained 6,000 JPY or 3,000 JPY, how much the other  
20 would receive (0 to 10,000 JPY; in units of 1,000 JPY). Then, they rated their liking (two items:  
21 liking and friendliness) and respect (respect and regard) for each version of A who adopted one  
22 of the three choices, i.e., did not share the 10,000 JPY, shared it equally, or gave the entire  
23 amount to the other person. All items were rated on a 7-point scale (1 = *not at all*, 7 = *extremely*).

1 For each condition, the ratings of the two “liking” items (i.e., liking and friendliness; inter-item  
2 correlations:  $r(465)s > .73, ps < .001$ ) and “respect” items (i.e., respect and regard; inter-item  
3 correlations:  $r(465)s > .82, ps < .001$ ) were averaged to obtain the dependent variables of liking  
4 and respect, respectively. In addition, as a further way to measure liking and respect, participants  
5 were asked to select the one person out of the three with whom they could become friends and  
6 would respect (1 = *a person who takes the entire amount*, 2 = *a person who gives the entire*  
7 *amount*, 3 = *a person who shares it equally*). We hypothesized that a person who gives the entire  
8 amount is less likely to be chosen as a friend than a person who shares it equally because  
9 friendship is regarded as a communal relationship.

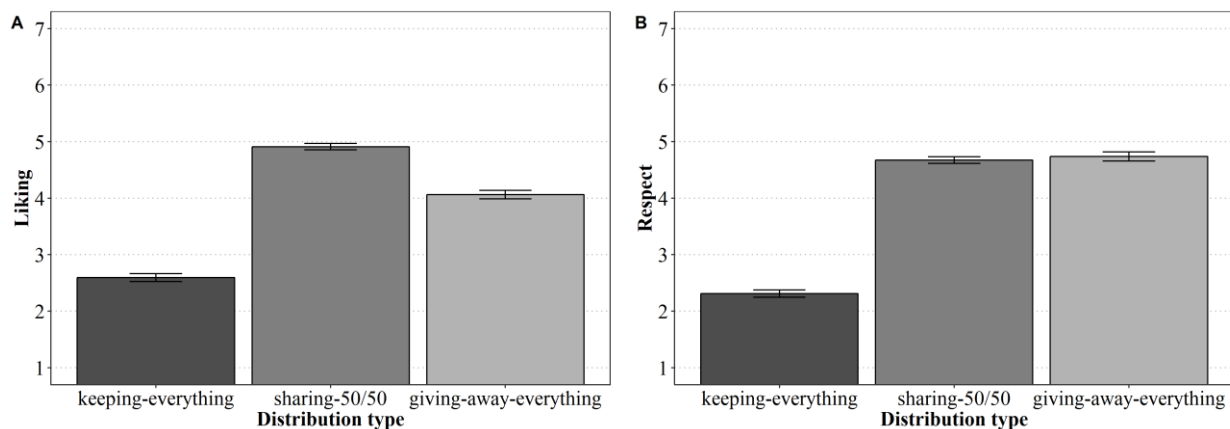
## 10 **Results and Discussion**

11 A total of 215 participants incorrectly answered the manipulation check question. Thus,  
12 data from these participants were excluded from the following analyses (we conducted the  
13 comparable analyses including these participants and confirmed that the exclusion of these  
14 participants did not alter the reported patterns of significance; see Supplementary Materials). The  
15 final sample comprised 467 Japanese individuals aged 18–29 years ( $M = 23.9, SD = 3.40$ ; 230  
16 men, 237 women). A sensitivity power analysis using G\*Power (Faul et al., 2007) indicated that  
17 our final sample size had 80% power to detect a small effect of  $f = .11$ , with  $\alpha = .05$ .

18 The correlations between the liking and respect scales in each condition were  $r(465)s$   
19 = .74–.82. First, we conducted a (distribution type: keeping-everything, sharing-50/50, or giving-  
20 away-everything) one-way repeated measures ANOVA on liking (Figure 1A). Since Mendoza’s  
21 test of sphericity was significant, the Greenhouse–Geisser correction was used for the ANOVA,  
22 resulting in fractional degrees of freedom. As shown in Figure 1A, there was a significant main  
23 effect of distribution,  $F(1.78, 831.64) = 302.74, p < .001, \eta_p^2 = .39$ . Multiple comparisons with

1 Holm's sequentially rejective Bonferroni procedure revealed that liking in the keeping-  
 2 everything condition was lower than the observed liking in the sharing-50/50 condition,  $t(466) =$   
 3  $24.25, p < .001, g = 1.71$  (Hedges'  $g$  statistic; Hedges, 1981), and in the giving-away-everything  
 4 condition,  $t(466) = 13.51, p < .001, g = 0.95$ . Furthermore, as predicted, liking in the giving-  
 5 away-everything condition was lower than liking in the sharing-50/50 condition,  $t(466) = 10.70,$   
 6  $p < .001, g = 0.59$ .

7 We also conducted a one-way repeated measures ANOVA on respect (Figure 1B). There  
 8 was a significant main effect of distribution type,  $F(1.80, 841.05) = 416.43, p < .001, \eta_p^2 = .47$ .  
 9 Multiple comparisons with Holm's sequentially rejective Bonferroni procedure revealed that  
 10 respect ratings in the keeping-everything condition were lower than those in the sharing-50/50  
 11 condition,  $t(466) = 26.11, p < .001, g = 1.79$ , and in the giving-away-everything condition,  $t(466)$   
 12  $= 22.00, p < .001, g = 1.55$ . However, respect in the giving-away-everything condition did not  
 13 differ from that in the sharing-50/50 condition,  $t(466) = 0.72, p = .473, g = 0.04$ .



14  
 15 *Figure 1.* Mean ratings (error bars represent *SEs*) of dependent variables in each condition in Study  
 16 1. (A) shows the ratings of liking, (B) shows the ratings of respect.

17

18 Next, we analyzed the responses concerning which person the participants felt they

1 could most likely become friends with or would respect (Table 1). A chi-square test on friend and  
 2 respect choices showed that there were significant differences among groups,  $\chi^2$ s (2) = 643.33,  
 3 194.57;  $ps < .001$ ;  $ws = 1.17, 0.65$ , respectively. Multiple comparisons with the Bonferroni  
 4 method revealed that the keeping-everything and giving-away-everything actors were less likely  
 5 to be chosen as friends than was the actor who shared equally ( $Zs = 18.11, 18.73, ps < .001, gs =$   
 6  $0.43, 0.45$ , respectively). There were no significant differences between the selection of the  
 7 keeping-everything and giving-away-everything actors as a friend ( $Z = 1.10, p = .815, g = 0.09$ ).  
 8 Regarding respect, the actor who kept everything was chosen as a person that participants would  
 9 respect less than the actors who shared equally or gave away everything ( $Zs = 14.28, 12.02, ps$   
 10  $< .001, gs = 0.43, 0.42$ , respectively). Additionally, the actor who gave away everything was  
 11 chosen as a person they would respect less than the actor who shared equally ( $Z = 2.83, p = .014,$   
 12  $g = 0.07$ ).

13  
 14 Table 1. *Number of Participants who Selected each Distributor as Being a Person They Were*  
 15 *Most Likely to Become Friends with or Would Respect (N = 467).*

Variable	Keeping-everything	Sharing-50/50	Giving-away-everything
	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)
Friend	31 (6.64%)	414 (88.65%)	22 (4.71%)
Respect	18 (3.85%)	255 (54.60%)	194 (41.54%)

16

17 In Study 1, we examined the evaluations of the person who selected the three types of  
 18 distribution (i.e., keeping-everything, sharing-50/50, and giving-away-everything). Clearly,  
 19 distributing the full 10,000 JPY (i.e., giving-away-everything) was more altruistic than

1 distributing 5,000 JPY (i.e., sharing-50/50), in that more self-sacrifice was exhibited and more  
2 benefits were given to others. Nonetheless, the person who distributed all the money did not  
3 acquire a better reputation than the person who equally shared the money. Although the  
4 difference in liking between the giving-away-everything and sharing-50/50 conditions was  
5 smaller than those in other comparisons (keeping-away-everything vs. sharing-50/50:  $g = 1.71$ ;  
6 keeping-away-everything vs. giving-away-everything:  $g = 0.95$ ), the effect was still moderate ( $g$   
7  $= 0.59$ ). Since the vignette was not a situation in which the reputation of participants was  
8 threatened by the extreme altruists, this finding may have been caused by the fact that the  
9 altruists deviated from the social norm.

10         Although norm-deviant altruism was associated with reduced evaluations of liking,  
11 evaluations of respect were not significantly different between the giving-away-everything and  
12 sharing-50/50 conditions. This finding is consistent with a previous study in which deviance did  
13 not always lead to low estimates of ability (Bellezza et al., 2014), as assessments of ability are  
14 related to feelings of respect (Wojciszke et al., 2009).

15         In the dictator game, giving away all the money rarely occurs compared with sharing  
16 equally (e.g., Engel, 2011). However, differences other than in terms of social norms exist  
17 between the two because the distribution ratio itself is different. For example, the actor who  
18 shared the money equally might be more likely to be seen as being fair compared to those who  
19 gave away all of the money, in that the former means both people get an equal amount of the  
20 money, whereas the latter means one person gets more money than the other. To eliminate this  
21 possibility, in Study 2, we manipulated the degree of deviation from social norms by varying  
22 stake size (i.e., the maximum distribution amount).

23

## 1 Study 2

2 In Study 2, we manipulated the stake size of the decisions. Specifically, in addition to  
3 the condition in which the person distributed 10,000 JPY (i.e., high stake size [HSS] condition),  
4 we added a condition in which the person distributed 400 JPY (i.e., low stake size [LSS]  
5 condition). Previous studies have suggested that people distribute more when the stake size is  
6 low (e.g., Larney, Rotella, & Barclay, 2019; Yamagishi, Li, Matsumoto, & Kiyonari, 2016).  
7 Therefore, in the LSS condition, giving away all the money is less likely to be regarded as an act  
8 of norm deviation than in the HSS condition. If the unfavorable evaluation of giving away all the  
9 money is caused by its norm-deviant nature, this phenomenon is less likely to be observed when  
10 the stake size is low.

## 11 Methods

12 **Participants and design.** We recruited 600 participants through the Japanese  
13 crowdsourcing service, CrowdWorks. The average age of participants was 36.8 years ( $SD = 9.66$ ;  
14 range = 20–60; 298 men, 302 women). Sample size was determined before any data analysis.  
15 This experiment used a 3 (distribution type: keeping-everything, sharing-50/50, or giving-away-  
16 everything)  $\times$  2 (stake size: LSS vs. HSS) mixed design. The former factor was manipulated  
17 within subjects. A power analysis using G\*Power (Faul et al., 2007) indicated that the required  
18 sample size was 526 to detect a small interaction effect (i.e.,  $f = .10$ ) in a 3  $\times$  2 mixed-design  
19 ANOVA, with  $\alpha = .05$  and  $\beta = .80$ . This study was approved by the ethics committee of the  
20 authors' institution.

21 **Procedure.** As in Study 1, participants read short vignettes and answered some  
22 questions. Most parts of the vignettes were the same as those of Study 1 (see Supplemental  
23 Material). However, in the LSS condition, the maximum amount that the actor had available for

1 distribution was 400 JPY. After reading the vignettes, participants first answered manipulation  
2 check questions; in the HSS condition, participants were asked to state the amount of money the  
3 distributor or recipient would obtain if one of these individuals received 3,000 or 6,000 JPY (0 –  
4 10,000 JPY; in units of 1,000 JPY), whereas in the LSS condition, participants were asked to  
5 state the amount of money the distributor or recipient would obtain if one of these individuals  
6 received 200 or 300 JPY (0–400 JPY; in units of 100 JPY). Then, they provided ratings of liking  
7 and respect, as in Study 1. Participants also provided demographic information.

## 8 **Results and Discussion**

9 Twelve participants incorrectly answered the manipulation check question. Thus, their  
10 data were excluded from the following analyses. The final sample comprised 588 Japanese  
11 individuals, aged 20–60 years ( $M = 36.9$ ,  $SD = 9.61$ ; 287 men, 301 women). The breakdown by  
12 condition was as follows: (a) LSS condition (146 men, 152 women), (b) HSS condition (141  
13 men, 149 women). A sensitivity power analysis using G\*Power (Faul et al., 2007) indicated that  
14 our final sample size had 80% power to detect a small effect of  $f = .09$ , with  $\alpha = .05$ .

15 For each condition, the two “liking” items (inter-item correlations in LSS condition:  
16  $r(296)s > .64$ ,  $ps < .001$ ; inter-item correlations in HSS condition:  $r(288)s > .67$ ,  $ps < .001$ ) and  
17 “respect” items (inter-item correlations in LSS condition:  $r(296)s > .75$ ,  $ps < .001$ ; inter-item  
18 correlations in HSS condition:  $r(288)s > .72$ ,  $ps < .001$ ) were averaged to obtain two dependent  
19 variables, as in Study 1. The correlations between the liking and respect scales were  $r(288)s$   
20  $= .71$ – $.72$  in the HSS condition and  $r(296)s = .68$ – $.77$  in the LSS condition.

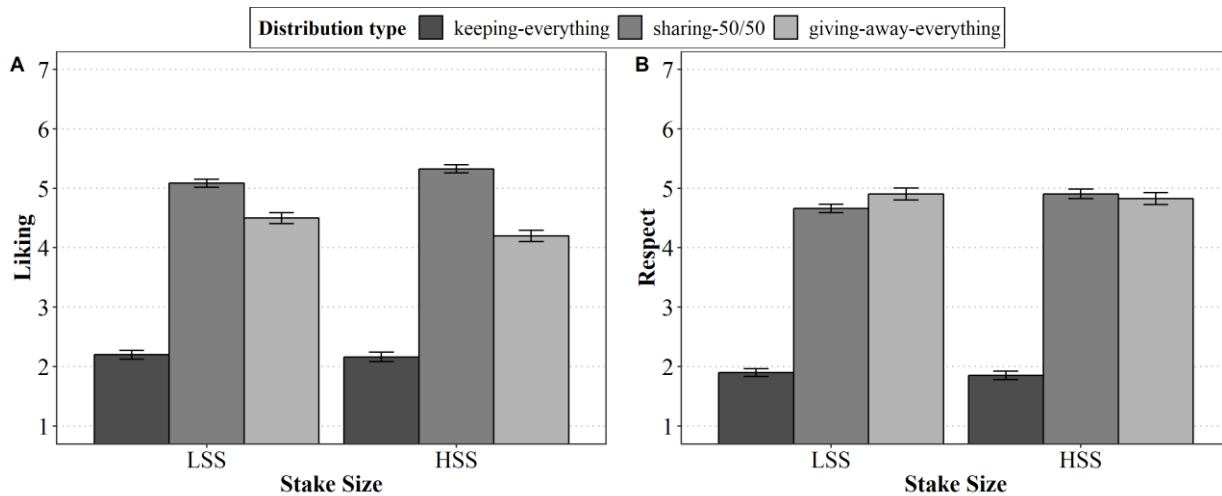
21 A 3 (distribution type: keeping-everything, sharing-50/50, or giving-away-everything)  $\times$   
22 2 (stake size: LSS vs. HSS) mixed-design ANOVA was conducted on liking ratings (see Figure  
23 2A). Since Mendoza’s test of sphericity was significant, the Greenhouse–Geisser correction was

1 used for the ANOVA, resulting in fractional degrees of freedom. A main effect of stake size was  
2 not significant,  $F(1, 586) = 0.25, p = .617, \eta_p^2 < .01$ . Additionally, a main effect of distribution  
3 type was significant,  $F(1.83, 1071.04) = 738.63, p < .001, \eta_p^2 = .56$ . However, this effect was  
4 moderated by stake size: The interaction between stake size and distribution type was significant,  
5  $F(1.83, 1071.04) = 5.55, p = .005, \eta_p^2 = .01$ . Irrespective of the HSS or LSS condition, liking in  
6 the keeping-everything condition was lower than that in the sharing-50/50 condition (HSS:  
7  $t(289) = 28.37, p < .001, g = 2.48$ ; LSS:  $t(297) = 28.90, p < .001, g = 2.35$ ) and in the giving-  
8 away-everything condition (HSS:  $t(289) = 14.93, p < .001, g = 1.35$ ; LSS:  $t(297) = 18.25, p$   
9  $< .001, g = 1.59$ ). Additionally, liking in the giving-away-everything condition was lower than  
10 that in the sharing-50/50 condition (HSS:  $t(289) = 10.72, p < .001, g = 0.79$ ; LSS:  $t(297) = 5.57,$   
11  $p < .001, g = 0.42$ ). In the keeping-everything condition, there were no significant differences  
12 between stake size conditions,  $F(1, 586) = 0.12, p = .732, \eta_p^2 < .01$ , whereas in the sharing-  
13 50/50 condition, liking was higher in the HSS condition than in the LSS condition,  $F(1, 586) =$   
14  $6.25, p = .013, \eta_p^2 = .01$ . Furthermore, as predicted, in the giving-away-everything condition,  
15 liking was higher in the LSS condition than in the HSS condition,  $F(1, 586) = 5.09, p = .024, \eta_p^2$   
16  $= .01$ .

17         Next, the same mixed-design ANOVA was conducted on respect ratings (see Figure 2B).  
18 The main effect of stake size was not significant,  $F(1, 586) = 0.36, p = .551, \eta_p^2 < .01$ . However,  
19 the main effect of distribution type was significant,  $F(1.85, 1087) = 864.20, p < .001, \eta_p^2 = .60$ .  
20 Multiple comparisons with Holm's sequentially rejective Bonferroni procedure revealed that the  
21 evaluation of respect in the keeping-everything condition was lower than the evaluation in the  
22 sharing-50/50 condition,  $t(586) = 40.34, p < .001, g = 2.34$ , and in the giving-away-everything  
23 condition,  $t(586) = 32.56, p < .001, g = 2.01$ . Respect ratings did not differ between giving-away-



1 everything and sharing-50/50 conditions,  $t(586) = 1.01$ ,  $p = .315$ ,  $g = 0.05$ . Furthermore, the  
 2 interaction between distribution type and stake size was not significant,  $F(1.85, 1087) = 2.35$ ,  $p$   
 3  $= .100$ ,  $\eta_p^2 < .01$ .



5  
 6 *Figure 2.* Mean ratings (error bars represent *SEs*) of dependent variables in each condition in Study  
 7 2. (A) shows the ratings of liking, (B) shows the ratings of respect. LSS = low stake size, HSS =  
 8 high stake size.

9  
 10 In Study 2, we manipulated stake size to determine the influence of social norms. The  
 11 main results of Study 1 were replicated; liking ratings for the actor who distributed all the money  
 12 were lower than those for the actor who equally shared the money (HSS:  $g = 0.79$ ; LSS:  $g =$   
 13  $0.42$ ). In addition, consistent with our hypothesis, a moderation effect was found; the liking for  
 14 the actor who distributed all the money in the HSS condition was lower than that in the LSS  
 15 condition. This was probably because the giving-away-everything in the HSS condition was  
 16 regarded as even more non-normative than in the LSS condition.

17 There are two limitations in Study 2. First, it should be noted that the difference between

1 the liking ratings of giving-away-everything in the HSS and LSS conditions was small: The  
2 effect size was  $\eta_p^2 = .01$ . The between-participants manipulation of stake size (400 vs. 10,000  
3 JPY) might not be salient. Second, it is unclear whether the manipulation of stake size was akin  
4 to manipulating social norms. We assumed that giving away all the money in the HSS condition  
5 would be more likely to be regarded as norm-deviant behavior than in the LSS condition because  
6 people distribute less when the stake size is high (e.g., Larney et al., 2019). However, this  
7 assumption was not empirically confirmed in Study 2. Therefore, we conducted an additional  
8 experiment to confirm the possibility that giving-away-everything in the HSS condition was seen  
9 as more norm deviant than that in the LSS condition. Two hundred eight participants, who did  
10 not participate in Study 2, read the same vignette as Study 2 and rated their perception of norm  
11 violation. In accordance with our assumption, participants considered giving-away-everything in  
12 the HSS condition a more prominent form of social norm deviance than that in the LSS condition  
13 (see Study S2 in the Supplementary Materials for details).

14 The results of Study 2 support the notion that unfavorable evaluations of altruism  
15 depend on the extent to which the behavior deviates from social norms. In Study 3, instead of  
16 experimentally manipulating social norms, we examined the effect of tolerance for norm  
17 deviations in the real world.

### 18 Study 3

19 Studies 1 and 2 suggest that if altruistic behavior deviates from social norms, the  
20 reputation of the actor is damaged. Previous studies have noted that cultural differences exist in  
21 the tendency to blame others for their deviant behaviors (Gelfand et al., 2011). Additionally,  
22 there exist cultural differences in unfavorable evaluations of altruism, which are related to norms  
23 existing in society (Herrmann et al., 2008; Klein et al., 2015). Therefore, if an unfavorable

1 reputation for altruistic behavior stems from the extent to which the behavior deviates from  
2 social norms, the unfavorable evaluations may be more prominently observed in societies with a  
3 strong expectation of adherence to social norms.

4         Studies 1 and 2 were conducted in Japan, which has low tolerance for deviant behavior  
5 (Gelfand et al., 2011); thus, if less positive evaluations of those who give away a large amount of  
6 money are caused by the behavior's deviant nature, the differences in liking between sharing-  
7 50/50 and giving-away-everything would be smaller in societies tolerant of norm deviation.  
8 Therefore, in Study 3, we conducted the experiment not only in Japan but also in the U.S., which  
9 has a high tolerance for deviance. The tendency to provide less favorable evaluations of those  
10 who distribute large amounts of money may be more evident in Japan and less so in the U.S.  
11 Additionally, we asked participants to rate perceptions of norm deviation to examine whether  
12 such perceptions mediate the relationship between distribution type and participants' evaluations,  
13 particularly in Japan.

## 14 **Methods**

15         **Participants and design.** We recruited 816 participants. The subsample of 416 Japanese  
16 individuals (208 men, 208 women;  $M_{\text{age}} = 39.9$ ,  $SD_{\text{age}} = 10.98$ , age range = 20–59) were recruited  
17 through the Macromill online research system (Macromill, Inc., Tokyo, Japan). The subsample of  
18 400 Americans (210 men, 190 women;  $M_{\text{age}} = 34.5$ ,  $SD_{\text{age}} = 9.50$ , age range = 20–59) were  
19 recruited through Amazon M-Turk. Sample size was determined before any data analysis. This  
20 experiment used a 2 (distribution type: sharing-50/50 vs. giving-away-everything)  $\times$  2 (culture:  
21 Japanese vs. American) mixed design; the keeping-everything condition used in Studies 1 and 2  
22 was omitted from Study 3. A power analysis using G\*Power (Faul et al., 2007) indicated that the  
23 required sample size was 396 to detect a small interaction effect (i.e.,  $f = .10$ ) in a 2  $\times$  2 mixed-

1 design ANOVA, with  $\alpha = .05$  and  $\beta = .80$ . This study was approved by the ethics committee of  
2 the authors' institution.

3         **Procedure.** Almost all procedures were the same as in the HSS condition of Study 2. In  
4 addition, participants responded to two questions addressing the perception of norm violation:  
5 “Most people would not give half of the \$100 (or 10,000 JPY) prize to the other person in this  
6 situation” and “Few people would give half of the \$100 (or 10,000 JPY) prize to the other person  
7 in this situation” in the sharing-50/50 condition. In the giving-away-everything condition, the  
8 same questions were used, but the word “half” was replaced with “all.” These items were rated  
9 on a 7-point-scale (1 = *strongly disagree*, 7 = *strongly agree*). One procedural detail differed  
10 between countries: in Japan, all choices on each scale were labeled (1 = *strongly disagree*, 2 =  
11 *disagree*, 3 = *disagree a little*, 4 = *neither agree nor disagree*, and so on). However, the scales  
12 used for the U.S. sample were labeled only at the poles (i.e., at the ratings of 1 and 7 only). We  
13 considered this to be a minor difference and therefore analyzed the data with the assumption that  
14 these differences did not add unnecessary bias to our design.

## 15 **Results and Discussion**

16         Sixty-six participants incorrectly answered the manipulation check question; thus, their  
17 data were excluded from the following analyses. The final sample consisted of 374 Japanese  
18 individuals (182 men, 192 women) and 376 Americans (195 men, 181 women). A sensitivity  
19 power analysis using G\*Power (Faul et al., 2007) indicated that our final sample size had 80%  
20 power to detect a small effect of  $f = .07$ , with  $\alpha = .05$ .

21         For each condition, the two “liking” items (inter-item correlations in American  
22 participants:  $r(374)s > .82$ ,  $ps < .001$ ; inter-item correlations in Japanese participants:  $r(372)s$   
23  $> .80$ ,  $ps < .001$ ), “respect” items (inter-item correlations in American participants:  $r(374)s > .79$ ,

1  $ps < .001$ ; inter-item correlations in Japanese participants:  $r(372)s > .83, ps < .001$ ), and norm  
2 perception items (inter-item correlations in American participants:  $r(374)s > .31, ps < .001$ ; inter-  
3 item correlations in Japanese participants:  $r(372)s > .79, ps < .001$ ) were averaged to obtain the  
4 dependent variables, as in Studies 1 and 2. The correlations between the liking and respect scales  
5 in each condition were  $r(374)s = .71-.83$  in Japanese participants and  $r(372)s = .90-.91$  in  
6 American participants.

7 We conducted a 2 (distribution type: sharing-50/50 vs. giving-away-everything)  $\times$  2  
8 (culture: Japanese vs. American) mixed-design ANOVA on the liking ratings (see Figure 3A).  
9 There were significant main effects of culture,  $F(1, 748) = 294.10, p < .001, \eta_p^2 = .28$ , and  
10 distribution type,  $F(1, 748) = 71.19, p < .001, \eta_p^2 = .09$ . In addition, an interaction effect was  
11 found,  $F(1, 748) = 62.77, p < .001, \eta_p^2 = .08$ . Among the Japanese participants, liking was  
12 higher in the sharing-50/50 condition than in the giving-away-everything condition,  $F(1, 373) =$   
13  $99.32, p < .001, \eta_p^2 = .21$ , whereas among the American participants, there were no significant  
14 differences in liking,  $F(1, 375) = 0.20, p = .654, \eta_p^2 < .01$ . For respect ratings (see Figure 3B),  
15 there was a main effect of culture,  $F(1, 748) = 247.46, p < .001, \eta_p^2 = .25$ , such that the respect  
16 ratings from Americans were higher than those from the Japanese. The main effect of distribution  
17 type,  $F(1, 748) = 2.42, p = .120, \eta_p^2 < .01$ , and the interaction,  $F(1, 748) = 1.42, p = .234, \eta_p^2$   
18  $< .01$ , were not significant.

19

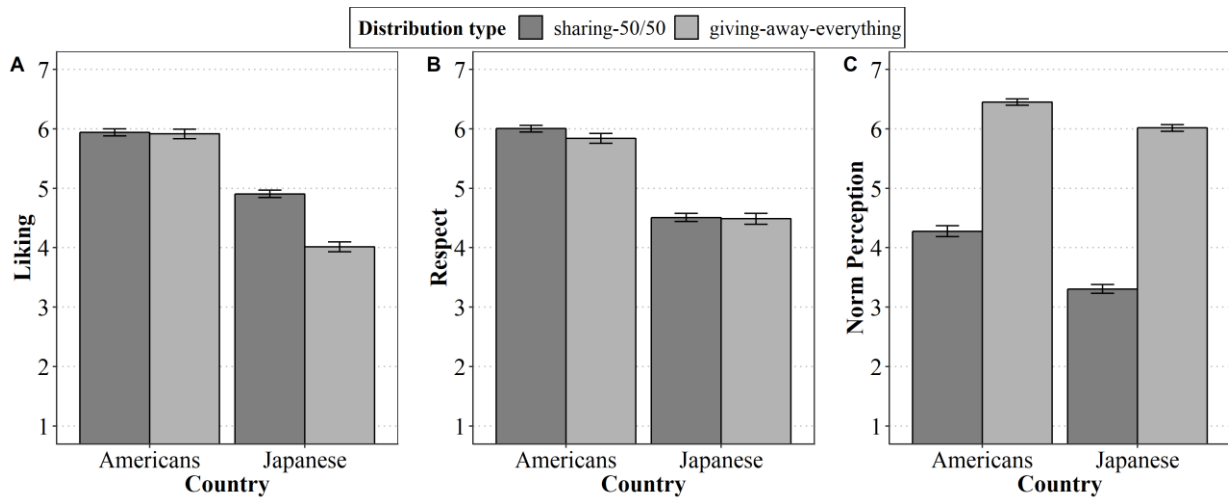


Figure 3. Mean ratings (*SEs*) of dependent variables in each condition in Study 3. (A) shows the ratings of liking, (B) shows the ratings of respect, (C) shows the ratings of norm perception.

We also conducted the same mixed-design ANOVA on the perception of norm deviation (see Figure 3C). Higher scores reflect greater perception of deviance. There were significant main effects of culture,  $F(1, 748) = 87.62, p < .001, \eta_p^2 = .10$ , and distribution type,  $F(1, 748) = 1311.91, p < .001, \eta_p^2 = .64$ . The interaction effect was also significant,  $F(1, 748) = 15.73, p < .001, \eta_p^2 = .02$ ; however, among both American and Japanese participants, giving-away-everything was regarded as being more norm deviant behavior than was sharing-50/50 (for Americans,  $F(1, 375) = 452.63, p < .001, \eta_p^2 = .55$ ; for Japanese,  $F(1, 373) = 950.92, p < .001, \eta_p^2 = .72$ ).

Next, we conducted mediation analyses to investigate whether the effects of distribution type on liking and respect ratings were mediated by the perception of norm violation. We used the SPSS macro MEMORE (Montoya & Hayes, 2017) version 2.1. To estimate the confidence interval (CI) for indirect effects, we used the bootstrap method (5,000 samples). First, with respect to liking for Japanese participants, the unfavorable evaluation of giving-away-everything

1 was mediated by the perception of norm violation (indirect effect = -0.54; bootstrap 95% CI [-  
2 0.85, -0.24]). In contrast, among Americans, this mediation effect was not found (indirect effect  
3 = -0.09; bootstrap 95% CI [-0.28, 0.11]). For respect ratings, no mediation effect was found in  
4 Japanese participants (indirect effect = -0.11; bootstrap 95% CI [-0.40, 0.18]) nor in Americans  
5 (indirect effect = -0.13; bootstrap 95% CI [-0.33, 0.06]).

6 Since the inter-item correlation of the two norm perception items was not high in  
7 American participants ( $r(374)s > .31$ ), we also conducted comparable analyses, treating the two  
8 items as independent mediators instead of using the averaged score. The reported pattern of  
9 significance did not change (see Supplementary Materials).

10 Study 3 investigated the reputation of extreme altruists in Japan and the U.S. Although  
11 respect did not differ between the sharing-50/50 and giving-away-everything conditions in terms  
12 of culture, the liking of giving-away-everything was lower than that of sharing-50/50 for  
13 Japanese participants only. Furthermore, the unfavorable evaluation by Japanese participants was  
14 mediated by the perception of norm deviation. In contrast, although American participants also  
15 viewed giving-away-everything as a form of norm violation, their liking of the actor did not  
16 decrease. These results are consistent with previous findings that Japanese people have stricter  
17 social norms than people in the U.S. (see Gelfand et al., 2011). More importantly, these results  
18 support our hypothesis that norm deviations are a factor in unfavorable evaluations of extremely  
19 altruistic behavior.

20 Two limitations of Study 3 should be noted. First, we found unexpected main effects of  
21 culture: Across the three scales, ratings of Japanese participants were lower than those of  
22 American participants. This may be caused by the cultural difference in response style: Japanese  
23 people are more likely to choose a midpoint response than Americans (Tasaki & Shin, 2017).

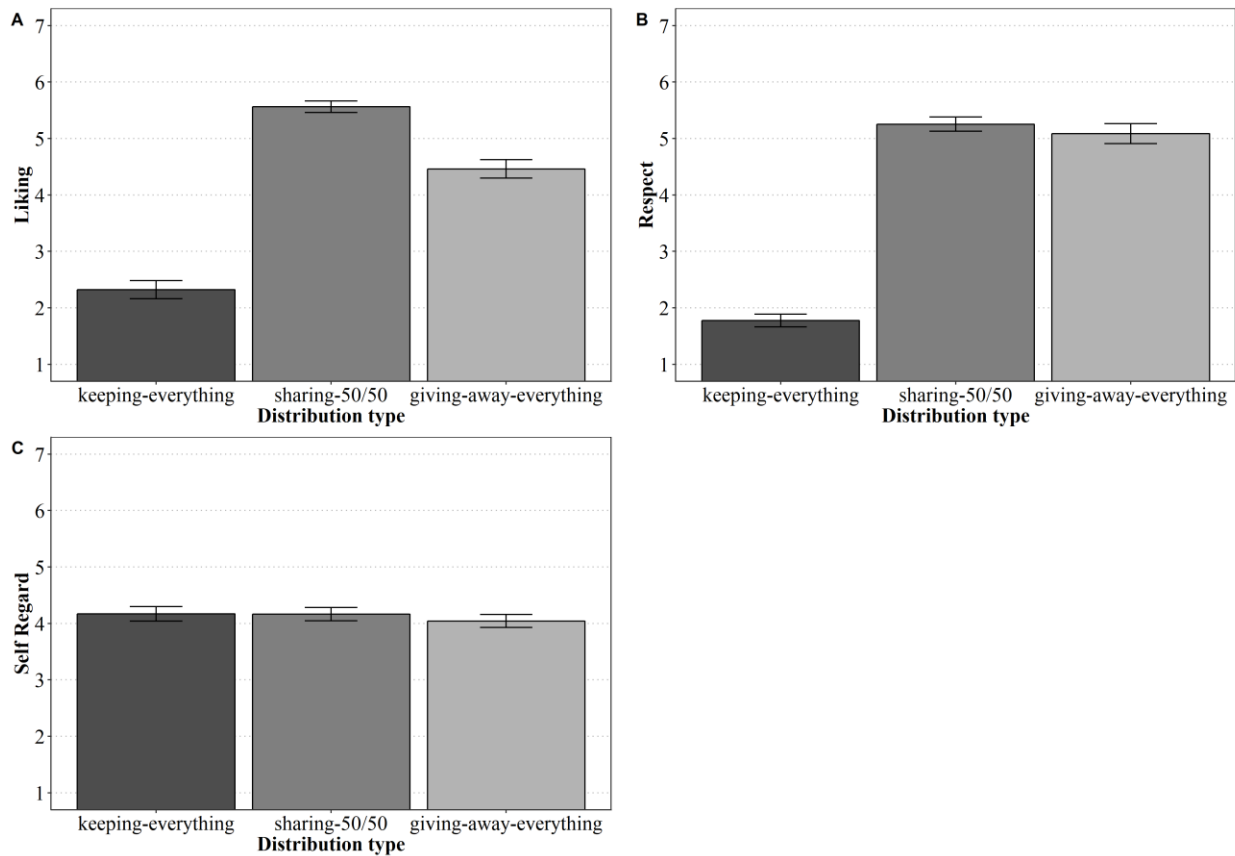
1 Second, as mentioned in the Method section, there were procedural differences between cultures:  
2 scales for U.S. participants were labeled only at the poles. Although the labels were identical  
3 between the giving-away-everything and sharing-50/50 conditions, this might have led to a  
4 different interpretation of midpoints between the two cultural groups.

#### 5 **Study 4**

6 The three presented studies have shown that altruists who deviate from social norms are  
7 less favored in Japan, which is characterized as having strong norms. However, three alternative  
8 explanations for this low favorability are still possible. First, all three studies applied a within-  
9 participant design; thus, participants could compare different distribution behaviors explicitly,  
10 which might have strengthened the differences among the three distribution types. Second, in the  
11 vignettes used in Studies 1–3, giving away all the money might be considered somewhat strange  
12 because participants might have assumed that the actor bought lottery tickets to get the money.  
13 Therefore, a different setting that makes giving away all the money seem more natural should be  
14 adopted. Third, although the present experiments eliminated the possibility that the altruist's  
15 behavior harmed participants' reputations, i.e., evaluation of others, the altruistic behavior might  
16 affect participants' self-concept, i.e., evaluation of oneself. Bolderdijk, Brouwer, and Cornelissen  
17 (2018) showed that a morally motivated innovator (e.g., a person who deviates from social  
18 norms to support others' welfare) is not preferred by observers because the innovator threatens  
19 the observers' moral self-concept. Bolderdijk et al. (2018) also showed that the morally  
20 motivated innovator threatens observers' self-concept and was unfavored by observers only when  
21 the observers are self-involved in the situation (e.g., before observers evaluated the innovator,  
22 they decided not to endorse the related ethical innovation). In the present experiments,  
23 participants are not self-involved in the situation; therefore, altruists who deviate from social



1 norms may not harm participants' self-evaluations. Nevertheless, this idea should be empirically  
 2 examined. To test these three alternative possibilities, Study 4 a) adopted a between-participants  
 3 design, b) modified the vignettes, and c) assessed participants' self-concept after they evaluated  
 4 the actors.



5  
 6 *Figure 4.* Mean ratings (*SEs*) of dependent variables in each condition in Study 4. (A) shows the  
 7 ratings of liking, (B) shows the ratings of respect, (C) shows the ratings of self-regard.

8

## 9 **Methods**

10 **Participants and design.** We recruited 270 participants (134 men, 136 women;  $M_{\text{age}} =$   
 11 40.9,  $SD_{\text{age}} = 9.41$ ; age range: 20–59) through CrowdWorks. None of the participants did not  
 12 participate in Study 2. Sample size was determined before any data analysis. The experiment

1 used a between-participants design (distribution type: keeping-everything, sharing-50/50, or  
2 giving-away-everything). A power analysis using G\*Power (Faul et al., 2007) indicated that the  
3 required sample size was 159 to detect a medium main effect (i.e.,  $f = .25$ ) in a  $1 \times 3$  between-  
4 participants ANOVA, with  $\alpha = .05$  and  $\beta = .80$ . This study was approved by the ethics committee  
5 of the authors' institution.

6 **Procedure.** Almost all procedures were identical with those in Study 1. However, the  
7 vignettes were slightly modified: The lottery tickets were described as given by a friend rather  
8 than the actor having bought them him/herself. Additionally, after rating the actor, participants  
9 rated their self-regard (Bolderdijk et al., 2018). Specifically, participants rated the extent they felt  
10 happy with themselves, satisfied with themselves, good, happy, comfortable, confident, and  
11 determined, as well as the reversed items, namely disappointed with themselves, annoyed with  
12 themselves, disgusted with themselves, angry with themselves, dissatisfied with themselves, self-  
13 critical, and guilty ( $\alpha = .93$ ). All items were rated on a 7-point scale (1 = *totally not applicable*, 7  
14 = *totally applicable*). Seven items were reverse coded such that higher scores reflected higher  
15 levels of self-regard. These items were the same as those used in Bolderdijk et al. (2018).

## 16 **Results and Discussion**

17 Since five participants incorrectly answered the manipulation check question, they were  
18 excluded from the following analyses. The final sample consisted of 265 Japanese individuals,  
19 aged 20–59 years ( $M = 40.9$ ,  $SD = 9.34$ ; 132 men, 133 women). The breakdown by condition  
20 was as follows: (a) keeping-everything (45 men, 44 women), (b) sharing-50/50 (44 men, 45  
21 women), (c) giving-away-everything (43 men, 44 women). A sensitivity power analysis using  
22 G\*Power (Faul et al., 2007) indicated that our final sample size had 80% power to detect a  
23 medium effect of  $f = .19$ , with  $\alpha = .05$ .

1           For each condition, the two “liking” items (inter-item correlations:  $r(87) = .79$  in the  
2 keeping-everything condition,  $r(87) = .58$  in the sharing-50/50 condition, and  $r(85) = .79$  in the  
3 giving-away-everything condition,  $ps < .001$ ) and “respect” items (inter-item correlations:  $r(87)$   
4  $= .75$  [keeping-everything],  $r(87) = .79$  [sharing-50/50], and  $r(85) = .82$  [giving-away-  
5 everything],  $ps < .001$ ) were averaged to obtain two dependent variables. The correlations  
6 between the liking and respect scales were  $r(87) = .76$  (keeping-everything),  $r(87) = .52$   
7 (sharing-50/50), and  $r(85) = .70$  (giving-away-everything).

8           A 3 (distribution type: keeping-everything, sharing-50/50, or giving-away-everything)  
9 between-participants ANOVA was conducted on liking ratings (see Figure 4A). A main effect of  
10 distribution type was found,  $F(2, 262) = 132.70$ ,  $p < .001$ ,  $\eta_p^2 = .50$ . Multiple comparisons with  
11 Holm’s sequentially rejective Bonferroni procedure revealed that liking in the keeping-  
12 everything condition was lower than liking in the sharing-50/50 condition,  $t(262) = 16.02$ ,  $p$   
13  $< .001$ ,  $g = 2.57$ , and in the giving-away-everything condition,  $t(262) = 10.52$ ,  $p < .001$ ,  $g = 1.41$ .  
14 Furthermore, as predicted, liking in the giving-away-everything condition was lower than liking  
15 in the sharing-50/50 condition,  $t(262) = 5.42$ ,  $p < .001$ ,  $g = 0.87$ .

16           Next, a 3 (distribution type: keeping-everything, sharing-50/50, or giving-away-  
17 everything) between-participants ANOVA was conducted on respect ratings (see Figure 4B). A  
18 main effect of distribution type was found,  $F(2, 262) = 196.88$ ,  $p < .001$ ,  $\eta_p^2 = .60$ . Multiple  
19 comparisons with Holm’s sequentially rejective Bonferroni procedure revealed that respect  
20 ratings in the keeping-everything condition were lower than those in the sharing-50/50 condition,  
21  $t(262) = 17.62$ ,  $p < .001$ ,  $g = 3.11$ , and in the giving-away-everything condition,  $t(262) = 16.68$ ,  $p$   
22  $< .001$ ,  $g = 2.39$ . As in Studies 1–3, respect in the giving-away-everything condition was not  
23 different from that in the sharing-50/50 condition,  $t(262) = 0.84$ ,  $p = .402$ ,  $g = 0.12$ .

1           Finally, a 3 (distribution type: keeping-everything, sharing-50/50, or giving-away-  
2 everything) between-participants ANOVA was conducted on self-regard ratings (see Figure 4C).  
3 The main effect of distribution type was not significant,  $F(2, 262) = 0.34, p = .709, \eta_p^2 < .01$ .

4           Study 4 demonstrated that, in Japan, extreme altruists (i.e., those who more extremely  
5 deviated from social norms) were less favored than modestly altruistic actors even when the  
6 distribution type was manipulated between participants and the act of giving away all the money  
7 was set up to seem more plausible. Furthermore, the difference in liking between the giving-  
8 away-everything and sharing-50/50 conditions was not smaller than that found in Study 1 (Study  
9 1:  $g = 0.59$  vs. Study 4:  $g = 0.87$ ). Additionally, being exposed to extremely altruistic actors did  
10 not threaten participants' self-concept: Participants' self-regard after reading the vignettes was  
11 not different regardless of which vignette they read. These results eliminated the aforementioned  
12 alternative explanations and confirmed that unfavorable evaluations toward extreme altruists in  
13 Japan are derived from the behavior's norm-deviant nature.

14

15

### General Discussion

16

17

18

19

20

21

22

23

We investigated whether norm violation would cause unfavorable evaluations of  
altruistic behavior across four studies. For Japanese participants, Study 1 showed that norm-  
deviant altruism (i.e., giving-away-everything) was regarded less favorably relative to  
moderately altruistic behavior (i.e., sharing-50/50). Study 2 found that giving away all the money  
with a high stake size, that is, giving away everything of a less normative nature, was regarded  
less favorably than was giving away everything with a low stake size, which is relatively  
normative. Furthermore, in Study 3, the tendency that norm-deviant altruism is less favored was  
only found in Japan, where norm deviation is likely to be punished. In addition, this unfavorable

1 evaluation was mediated by the perception of norm deviation. Finally, Study 4 successfully  
2 replicated the findings for a between-participants design and modified vignettes. Study 4 also  
3 demonstrated that even though the norm-deviant altruism did not threaten observers' self-  
4 concepts, altruists who deviate from social norms are less favored in Japan. These results suggest  
5 that norm-deviant altruism is less favored when norm deviation is regarded as undesirable.

6 In Study 3, the actor who distributed all the money was not evaluated more negatively  
7 than was the actor who shared equally in the American sample. Thus, altruists who deviate from  
8 social norms are not always unfavored by observers. Nonetheless, a critical point is that the non-  
9 negative evaluations in the U.S. could still be interpreted from the perspective of social norms as  
10 the U.S. has not been considered to have strong norms (see Gelfand et al., 2011). This result is  
11 also consistent with research regarding cultural differences in antisocial punishment (Herrmann  
12 et al., 2008; see also Klein et al., 2015). Thus, the Studies 1–4 consistently showed that the  
13 deviant nature of some types of altruism could cause unfavorable evaluations of altruism when  
14 norm deviation is evaluated negatively.

15 Extant studies have already suggested that altruistic behavior that deviates from social  
16 norms is evaluated negatively or punished (Herrmann et al., 2008; Irwin & Horne, 2013; Parks &  
17 Stone, 2010). However, these studies placed observers in situations where they would also be  
18 evaluated by others; in those situations, the existence of extreme altruists can threaten observers'  
19 reputations. Therefore, observers might be concerned about their own reputations, which leads to  
20 their negative evaluations of altruists (e.g., Pleasant & Barclay, 2018). In the present research,  
21 however, participants evaluated altruists from a third-person perspective. Thus, they were not  
22 likely to be concerned with their own relative reputation. Nevertheless, unfavorable evaluations  
23 were assigned to altruism that deviated from social norms in Japan; this constitutes evidence for

1 the hypothesis that norm deviation itself can cause unfavorable evaluations of altruism.

2         The present study does not deny the notion that threats to the evaluator's own reputation  
3 may also contribute to unfavorable attitudes toward altruism. Antisocial punishment has been  
4 universally observed in many societies, despite cross-cultural variance in its frequency (e.g.,  
5 Herrmann et al., 2008); this might be because negative evaluations of altruism can also be driven  
6 by concern for one's own reputation. The theoretical contribution of the present study is the  
7 finding that deviating from social norms itself can harm the altruist's reputation, independent of  
8 threats to the evaluator's own reputation. Future studies on antisocial punishment should  
9 consider the occurrence of these two processes.

10         In addition, our findings suggest that even when there is low tolerance for norm  
11 deviation, extreme altruism is not negatively evaluated in all domains. Regarding evaluations of  
12 respect, the present findings suggest that there are no differences between the giving-away-  
13 everything and sharing-50/50 conditions. Bellezza et al. (2014) suggested that norm violation is  
14 not always related to the perceived weakness of the violators. Additionally, perceptions of the  
15 target's ability are associated with respect (Wojciszke et al., 2009). Then, even when altruists  
16 violate social norms in societies intolerant of such deviation, they do not receive decreased  
17 respect from observers. Future studies are needed to explore how such domain-specific attitudes  
18 toward altruism affect interpersonal relations in real life.

19         It should be noted that the present study referred to descriptive norms, that is, norms  
20 defined by what others do (Cialdini, Reno, & Kallgren, 1990). However, injunctive norms, what  
21 people ought to do, should also be considered. Although giving away a large amount of money  
22 deviates from descriptive norms, it is not obvious whether the behavior violates injunctive  
23 norms. In the present study, extreme altruism acquired respect even in places and cultures where

1 deviation from descriptive norms is unfavored (i.e., Japan); this might be because the altruism  
2 does not violate injunctive norms. Future studies should focus on what types of norms extreme  
3 altruistic behavior violates.

4         The present study may also have implications for how to promote altruism. Previous  
5 studies have indicated that opportunities to obtain a good reputation may be a useful means of  
6 promoting altruistic behavior in the real world (Barclay, 2012). Indeed, the presence of others  
7 can generally increase altruism (for a meta-analytic review, see Bradley, Lawrence, & Ferguson,  
8 2018). This may hinge on the assumption that altruism can lead to the acquisition of a good  
9 reputation. However, as the present study showed, extreme altruism is sometimes unfavored.  
10 Therefore, opportunities to be evaluated by others will not always promote altruism and may  
11 even inhibit it. In fact, in a Japanese sample, Kawamura and Kusumi (2018) showed that people  
12 who tend to avoid a bad reputation inhibit altruistic behavior when others do not behave  
13 altruistically. Similarly, people who donate extremely large amounts of money to charitable  
14 causes tend to hide their donation from others (Raihani, 2014). This might be because donors are  
15 concerned about receiving a bad reputation. Thus, reputational concern can inhibit altruism. In  
16 the future, it would be useful to empirically investigate whether people avoid donating money  
17 when extreme altruism could lead to a bad reputation.

18         Finally, some limitations should be noted. First, all four studies did not assess real  
19 behavior. Second, the mediation analysis in Study 3 did not negate the existence of other possible  
20 mediators (Fiedler, Harris, & Schott, 2018); there could also be other cultural differences  
21 between Japan and the U.S. at play, such as cultural construal of self (Markus & Kitayama, 1991)  
22 and relational mobility (Yuki et al., 2007). Third, the lasting effect of an unfavorable rating is  
23 unknown: Although giving away large amounts of money is less favored in Japan, it is unclear

1 how long the effect remains. Finally, the present study did not consider the possibility that people  
2 have different norms and different propensities to openly express dislike of a norm violation.  
3 These issues remain to be considered in the future.

#### 4 **Conclusions**

5         Although altruism is generally praised by others (e.g., Barclay, 2012), it is sometimes  
6 evaluated negatively (e.g., Herrmann et al., 2008). The present study focused on the possibility  
7 that one reason for unfavorable evaluations is the behavior's norm-deviant nature; thus, the  
8 unfavorable evaluations of altruism depend on the extent to which the behavior deviates from  
9 social norms. Four experiments consistently showed that altruistic behavior that deviates from  
10 social norms was less preferred than modestly altruistic behavior—particularly in a tight society  
11 like Japan (Study 3), and especially when the degree of deviance is high (Study 2). These  
12 findings suggest that when norm deviation is unfavored, discomfort with norm deviation can  
13 result in unfavorable evaluations of altruism even though the deviation benefits others.

14

15



## References

- 1  
2
- 3 Arnocky, S., Piche, T., Albert, G., Ouellette, D., & Barclay, P. (2017). Altruism predicts mating  
4 success in humans. *British Journal of Psychology*, *108*, 416–435. doi:10.1111/bjop.12208
- 5 Barclay, P. (2012). Harnessing the power of reputation: strengths and limits for promoting  
6 cooperative behaviors. *Evolutionary Psychology*, *10*, 868-883.  
7 doi:10.1177/147470491201000509
- 8 Barclay, P., & Willer, R. (2007). Partner choice creates competitive altruism in humans.  
9 *Proceedings of the Royal Society B: Biological Sciences*, *274*, 749-753.  
10 doi:10.1098/rspb.2006.0209
- 11 Baryła, W. (2014). Liking goes from the perceiver's self-interest, but respect is socially shared.  
12 *Polish Psychological Bulletin*, *45*, 402-410. doi:10.2478/ppb-2014-0049
- 13 Bellezza, S., Gino, F., & Keinan, A. (2014). The Red Sneakers Effect: Inferring Status and  
14 Competence from Signals of Nonconformity. *Journal of Consumer Research*, *41*, 35-54.  
15 doi:10.1086/674870
- 16 Bereczkei, T., Birkas, B., & Kerekes, Z. (2007). Public charity offer as a proximate factor of  
17 evolved reputation-building strategy: an experimental analysis of a real-life situation.  
18 *Evolution and Human Behavior*, *28*, 277-284. doi:10.1016/j.evolhumbehav.2007.04.002
- 19 Bereczkei, T., Birkas, B., & Kerekes, Z. (2010). Altruism towards strangers in need: costly  
20 signaling in an industrial society. *Evolution and Human Behavior*, *31*, 95-103.  
21 doi:10.1016/j.evolhumbehav.2009.07.004
- 22 Bolderdijk, J. W., Brouwer, C., & Cornelissen, G. (2018). When Do Morally Motivated Innovators  
23 Elicit Inspiration Instead of Irritation? *Frontiers in Psychology*, *8*, 2362.

- 1           doi:10.3389/fpsyg.2017.02362
- 2   Bradley, A., Lawrence, C., & Ferguson, E. (2018). Does observability affect prosociality?
- 3           *Proceedings of the Royal Society B: Biological Sciences*, 285, 20180116.
- 4           doi:10.1098/rspb.2018.0116
- 5   Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A Focus Theory of Normative Conduct:
- 6           Recycling the Concept of Norms to Reduce Littering in Public Places. *Journal of*
- 7           *Personality and Social Psychology*, 58, 1015-1026. doi:10.1037/0022-3514.58.6.1015
- 8   DeBono, A., & Muraven, M. (2014). Rejection perceptions: feeling disrespected leads to greater
- 9           aggression than feeling disliked. *Journal of Experimental Social Psychology*, 55, 43-52.
- 10          doi:10.1016/j.jesp.2014.05.014
- 11   Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, 14, 583-610.
- 12          doi:10.1007/s10683-011-9283-7
- 13   Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power
- 14           analysis program for the social, behavioral, and biomedical sciences. *Behavior Research*
- 15           *Methods*, 39, 175-191. doi:10.3758/BF03193146
- 16   Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415, 137-140.
- 17          doi:10.1038/415137a
- 18   Feinberg, M., Willer, R., & Schultz, M. (2014). Gossip and ostracism promote cooperation in
- 19           groups. *Psychological Science*, 25, 656-664. doi:10.1177/0956797613510184
- 20   Fiedler, K., Harris, C., & Schott, M. (2018). Unwarranted inferences from statistical mediation
- 21           tests – An analysis of articles published in 2015. *Journal of Experimental Social*
- 22           *Psychology*, 75, 95-102. doi:10.1016/j.jesp.2017.11.008
- 23   Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., . . . Yamaguchi, S. (2011).

- 1 Differences Between Tight and Loose Cultures: A 33-Nation Study. *Science*, 332, 1100-  
2 1104. doi:10.1126/science.1197754
- 3 Hardy, C. L., & Van Vugt, M. (2006). Nice guys finish first: the competitive altruism hypothesis.  
4 *Personality and Social Psychology Bulletin*, 32, 1402-1413.  
5 doi:10.1177/0146167206291006
- 6 Hedges, L. V. (1981). Distribution theory for glass's estimator of effect size and related estimators.  
7 *Journal of Educational Statistics*, 6, 107-128. doi:10.2307/1164588
- 8 Herrmann, B., Thoni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science*,  
9 319, 1362-1367. doi:10.1126/science.1153808
- 10 Irwin, K., & Horne, C. (2013). A normative explanation of antisocial punishment. *Social Science*  
11 *Research*, 42, 562-570. doi:10.1016/j.ssresearch.2012.10.004
- 12 Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the assumptions of economics.  
13 *The Journal of Business*, 59, 285-300. doi:10.1086/296367
- 14 Kawamura, Y., & Kusumi, T. (2018). The relationship between rejection avoidance and altruism  
15 is moderated by social norms. *Personality and Individual Differences*, 129, 24-27.  
16 doi:10.1016/j.paid.2018.02.038
- 17 Klein, N., & Epley, N. (2014). The topography of generosity: asymmetric evaluations of prosocial  
18 actions. *Journal of Experimental Psychology: General*, 143, 2366-2379.  
19 doi:10.1037/xge0000025
- 20 Klein, N., Grossmann, I., Uskul, A. K., Kraus, A. A., & Epley, N. (2015). It pays to be nice, but  
21 not really nice: Asymmetric reputations from prosociality across 7 countries. *Judgment and*  
22 *Decision Making*, 10, 355-364.
- 23 Kuběna, A. A., Houdek, P., Lindova, J., Priplatova, L., & Flegr, J. (2014). Justine effect:

- 1           punishment of the unduly self-sacrificing cooperative individuals. *PLoS One*, *9*, e92336.  
2           doi:10.1371/journal.pone.0092336
- 3   Larney, A., Rotella, A., & Barclay, P. (2019). Stake size effects in ultimatum game and dictator  
4           game offers: A meta-analysis. *Organizational Behavior and Human Decision Processes*,  
5           *151*, 61-72. doi:10.1016/j.obhdp.2019.01.002
- 6   Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion,  
7           and motivation. *Psychological Review*, *98*, 224. doi:10.1037/0033-295X.98.2.224
- 8   Minson, J. A., & Monin, B. (2012). Do-Gooder Derogation: Disparaging Morally Motivated  
9           Minorities to Defuse Anticipated Reproach. *Social Psychological and Personality Science*,  
10           *3*, 200-207. doi:10.1177/1948550611415695
- 11   Monin, B., Sawyer, P. J., & Marquez, M. J. (2008). The rejection of moral rebels: resenting those  
12           who do the right thing. *Journal of Personality and Social Psychology*, *95*, 76-93.  
13           doi:10.1037/0022-3514.95.1.76
- 14   Montoya, A. K., & Hayes, A. F. (2017). Two-condition within-participant statistical mediation  
15           analysis: A path-analytic framework. *Psychological Methods*, *22*, 6-27.  
16           doi:10.1037/met0000086
- 17   Parks, C. D., & Stone, A. B. (2010). The desire to expel unselfish members from the group. *Journal*  
18           *of Personality and Social Psychology*, *99*, 303-310. doi:10.1037/a0018403
- 19   Pleasant, A., & Barclay, P. (2018). Why Hate the Good Guy? Antisocial Punishment of High  
20           Cooperators Is Greater When People Compete To Be Chosen. *Psychological Science*, *29*,  
21           868-876. doi:10.1177/0956797617752642
- 22   Raihani, N. J. (2014). Hidden altruism in a real-world setting. *Biology Letters*, *10*, 20130884.  
23           doi:10.1098/rsbl.2013.0884

- 1 Tasaki, K., & Shin, J. (2017). Japanese response bias: Cross-level and cross-national comparisons  
2 on response styles. *The Japanese journal of psychology*, 88, 32-42.  
3 doi:10.4992/jjpsy.88.15065
- 4 Wojciszke, B., Abele, A. E., & Baryla, W. (2009). Two dimensions of interpersonal attitudes:  
5 Liking depends on communion, respect depends on agency. *European Journal of Social*  
6 *Psychology*, 39, 973-990. doi:10.1002/ejsp.595
- 7 Yamagishi, T., Li, Y., Matsumoto, Y., & Kiyonari, T. (2016). Moral Bargain Hunters Purchase  
8 Moral Righteousness When it is Cheap: Within-Individual Effect of Stake Size in  
9 Economic Games. *Scientific Reports*, 6, 27824. doi:10.1038/srep27824
- 10 Yuki, M., Schug, J., Horikawa, H., Takemura, K., Sato, K., Yokota, K., & Kamaya, K. (2007).  
11 Development of a scale to measure perceptions of relational mobility in society. *CERSS*  
12 *Working Paper 75*, Center for Experimental Research in Social Sciences, Hokkaido  
13 University, Sapporo, Japan.  
14