

MOR-16-056R2  
13 September 2017

## **Cultural Universals and Cultural Differences in Meta-Norms about Peer Punishment**

Kimmo Eriksson<sup>1,2</sup>, Pontus Strimling<sup>3,1</sup>, Per A. Andersson<sup>4</sup>, Mark Aveyard<sup>5</sup>, Markus Brauer<sup>6</sup>, Vladimir Gritskov<sup>7</sup>, Toko Kiyonari<sup>8</sup>, David M. Kuhlman<sup>9</sup>, Angela T. Maitner<sup>5</sup>, Zoi Manesi<sup>10</sup>, Catherine Molho<sup>10</sup>, Leonard S. Peperkoorn<sup>10</sup>, Muhammad Rizwan<sup>11</sup>, Adam W. Stivers<sup>12</sup>, Qirui Tian<sup>13</sup>, Paul A. M. Van Lange<sup>10</sup>, Irina Vartanova<sup>3,14</sup>, Junhui Wu<sup>15</sup>, Toshio Yamagishi<sup>16</sup>

<sup>1</sup>Stockholm University, Sweden, <sup>2</sup>Mälardalen University, Sweden, <sup>3</sup>Institute for Futures Studies, Sweden, <sup>4</sup>Linköping University, Sweden, <sup>5</sup>American University of Sharjah, United Arab Emirates, <sup>6</sup>University of Wisconsin-Madison, USA, <sup>7</sup>Saint Petersburg State University, Russian Federation, <sup>8</sup>Aoyama Gakuin University, Japan, <sup>9</sup>University of Delaware, USA, <sup>10</sup>Vrije Universiteit Amsterdam, Netherlands, <sup>11</sup>University of Karachi, Pakistan, <sup>12</sup>Gonzaga University, USA, <sup>13</sup>Shandong Normal University, China, <sup>14</sup>National Research University Higher School of Economics, Russian Federation, <sup>15</sup>Beijing Normal University, China, and <sup>16</sup>Hitotsubashi University, Japan

### **Abstract**

Violators of cooperation norms may be informally punished by their peers. How such norm enforcement is judged by others can be regarded as a meta-norm (i.e., a second-order norm). We examined whether meta-norms about peer punishment vary across cultures by having students in eight countries judge animations in which an agent who over-harvested a common resource was punished either by a single peer or by the entire peer group. Whether the punishment was retributive or restorative varied between two studies, and findings were largely consistent across these two types of punishment. Across all countries, punishment was judged as more appropriate when implemented by the entire peer group than by an individual. Differences between countries were revealed in judgments of punishers vs. non-punishers. Specifically, appraisals of punishers were relatively negative in three Western countries and Japan, and more neutral in Pakistan, UAE, Russia, and China, consistent with the influence of individualism, power distance, and/or indulgence. Our studies constitute a first step in

mapping how meta-norms vary around the globe, demonstrating both cultural universals and cultural differences.

**Keywords:** cross-cultural research, individualism-collectivism, meta-norms, norm enforcement peer punishment

Running title: Culture and Meta-Norms about Peer Punishment

Corresponding author: Kimmo Eriksson ([kimmo.eriksson@mdh.se](mailto:kimmo.eriksson@mdh.se))

## INTRODUCTION

A manager chides his colleague for arriving 15 minutes late to a meeting. A construction worker yells at a team member for not pulling his weight. A child smacks his sibling for taking more than his share of the pizza. Researchers refer to such actions as peer punishment and the underlying situations as social dilemmas, characterized by the temptation for individuals to maximize selfish benefits instead of maximizing collective benefits. Research on social dilemmas demonstrates that peer punishment of free-riders can solve the problem of sustaining cooperation by eliminating the temptation to free-ride (e.g., Yamagishi, 1986). To the extent that peer punishment is often beneficial for the group but costly for the punisher, it has been considered as ‘altruistic’ (e.g., Fehr & Gächter, 2000). However, as punishment is also an aggressive act, people may not always regard such actions as altruistic. According to the interactionist approach to aggression (Felson, 1981; Tedeschi, Gaes, & Rivera, 1977; Tedeschi, Smith, & Brown, 1974), perceived legitimacy of an aggressive act depends on both the context and the audience. Deviation from a norm may legitimize punishment, but such punishment may conflict with rules of politeness that preclude certain kinds of admonishment (Felson, 1981). In other words, social norms may also regulate how people treat norm violations. Such second-order norms are sometimes referred to as meta-norms (Axelrod,

1986). We seek to examine how meta-norms on peer punishment in social dilemmas depend on cultural factors.

Building on the interactionist approach, Bond, Wan, Leung, and Giacalone (1985) suggested that the degree to which aggressive acts are illegitimate modes of social control should vary across cultures, in particular along the two most important dimensions of national culture variation, power distance and individualism-collectivism (Hofstede, 1980). To examine the role of culture, Bond et al. (1985) conducted a study with students from the United States (individualistic and low power distance) and Hong Kong China (collectivistic and high power distance). Participants were presented with a scenario depicting a company board meeting chaired by the company president and involving four key agents: the manager and the assistant manager from each of two departments. As tension rose in this meeting, one of the managers insulted one of the assistant managers or vice versa. Participants rated the acceptability of the way the insulter spoke. An interaction effect between the status of the insulter (manager or assistant) and whether the target was within the same department (ingroup or outgroup) suggested that the legitimacy of an insult was highest when it came from a manager to his own assistant. This effect existed in both cultures but was stronger among Hong Kong Chinese participants than among US participants.

The findings of Bond et al. (1985) suggest that meta-norms may differ only in degree rather than in kind across different cultures. However, to examine social dilemmas the scenario used by Bond et al. (1985) is not ideal. First, the insult was triggered by some disagreement rather than by selfish behavior. Second, the situation involved vertical relations (i.e., managers and their assistants), whereas most social dilemmas involve situations with peers. The aim of the present paper is to examine the cultural influence on meta-norms in a situation where peers are involved in a social dilemma.

## THEORETICAL BACKGROUND AND HYPOTHESES

The main idea behind the notion of ‘altruistic punishment’ is that peer punishment may be beneficial for the group but costly for the punisher. As such, the provision of punishment would entail a second-order social dilemma, the solution to which would require that non-punishers are also punished (e.g., Henrich & Boyd, 2001). From this theoretical perspective that norm enforcement is itself enforced, it seems that meta-norms should prescribe the use of peer punishment of non-cooperators. However, the fact that punishment may be needed by the group to sustain cooperation, does not preclude the possibility that it will be negatively appraised – a well-known example is whistleblowing, which generally is thought of as the morally right thing to do but nonetheless tends to elicit harsh social sanctions from the ingroup (e.g., Dasgupta & Kesharwani, 2010). Employees may question the motives of whistleblowers.

Several studies have examined social responses to peer punishment in economic games. In such games, punishment entails reducing the target’s payoff at a smaller cost to oneself (e.g., reduce other’s payoff by three units by giving up one unit). The typical finding in such studies is more negative responses to peer punishers than to non-punishers (Cinyabuguma et al., 2006; Kiyonari & Barclay, 2008; Strimling & Eriksson, 2014). This finding was replicated in a study that used animations of retributive and restorative punishment, instead of economic games (Eriksson, Andersson, & Strimling, 2016). These findings indicate that meta-norms in social dilemmas do *not* tend to prescribe the use of peer punishment, contrary to the abovementioned theoretical perspective (e.g., Henrich & Boyd, 2001).

The situational feature that all actors are peers may explain why punishment is not approved of. Without any vertical relations it is not clear which particular individual should take the lead to punish, so the person who punishes may be seen as overstepping his or her

authority. This problem may be mitigated if punishment is administered by the entire peer group instead of an individual. Indeed, three studies have found more positive ratings on appropriateness when punishment was implemented by the group instead of an individual (Eriksson et al., 2016; Mathew, 2017; Strimling & Eriksson, 2014).

Although most studies were conducted in modern Western societies, exceptions included online surveys with participants from India (Strimling & Eriksson, 2014), and a face-to-face survey in a pastoralist tribe in Kenya (Mathew, 2017). Overall, meta-norms against peer punishers have been found across several cultures. However, because studies have used different methods and measures, results are not directly comparable and do not allow for meaningful investigation of cultural influence on meta-norms of peer punishment. Given the theoretical importance of meta-norms in solving social dilemmas, we set out to conduct a systematic investigation of meta-norms across cultures. Below we outline our questions and hypotheses.

### **How Does Culture Influence the Appropriateness of Peer Punishment?**

To theoretically analyze how culture may influence the appropriateness of peer punishment in social dilemmas, we followed Bond et al. (1985). They focused on the cultural dimensions of power distance and individualism-collectivism. The situation they studied involved explicit power differences. In contrast, we examine situations where power differences are absent. Power distance should therefore be less relevant for the situations we examine. Accordingly, we focused our theorizing on the possible influence of individualism-collectivism. In the general discussion we will return to other cultural dimensions and consider how they may also influence meta-norms of peer punishment.

Individualism-collectivism refers to the degree to which a society emphasizes individual autonomy over group embeddedness. The relation between individualism-collectivism and

justice was discussed by Leung and Stephan (2001, p. 393) in an essay on culture and justice. One of their arguments is that people from collectivistic societies are more comfortable with the use of social sanctions against counter-normative behavior than people in individualistic societies. Specifically, due to their prioritization of autonomy, individualists are less likely than collectivists societies to consider social sanctions as “a tool to ensure that individuals act in accordance with the norms of the group” (Leung & Stephan, 2001, p. 393). In support of this argument, several cross-cultural studies find that norm violators are less likely to be punished in more individualistic cultures than in more collectivistic cultures (Brauer & Chaurand, 2010; Gelfand et al., 2011; Wang & Leung, 2010). In line with these arguments and data, we therefore expect punishment of social norm violations to be appraised as less appropriate in more individualistic cultures.

### **Is Group Punishment Judged More Appropriate than Individual Punishment?**

As noted earlier, research suggests that the exact same punishment is rated as more appropriate when implemented by the entire group instead of an individual (Eriksson et al., 2016; Strimling & Eriksson, 2014). Can this effect be accounted for by individualism-collectivism? An established line of research has examined how participants from individualistic and collectivistic societies differ in their perceptions of scenarios that involve an individual or a group (e.g., Chiu, Morris, Hong, & Menon, 2000; Menon, Morris, Chiu, & Hong, 1999; Morris, Menon, & Ames, 2001). However, they focused on causal attribution and not on relative preferences for individual vs. group action. We are not aware of any research that directly answers the question of such preferences. The cited research has established a greater tendency to attribute agency and dispositions to groups in collectivistic cultures than in individualistic cultures. For instance, Menon et al. (1999) found that negative group and team outcomes were attributed to group-level disposition more by Hong Kong

Chinese than American students. As this example shows, collectivistic cultures do not simply perceive group agents as good. Differences in preferences are likely to depend on the context. In the context of punishment in a social dilemma, it does not seem obvious how individualism-collectivism would influence the preference for group punishment.

### **Are Meta-Norms the Same for Retributive and Restorative Punishment?**

Theories of justice distinguish between two forms of punishment: (a) *retributive punishment* that makes the offender suffer for the harm they have done and (b) *restorative punishment* that repair just relations among victim, offender, and community (e.g., Cohen, 2016; Wenzel & Okimoto, 2016). They are part of different philosophical approaches to justice that are much more complex than this simple distinction in punishment. However, in the social dilemma context of a group sharing a common resource (e.g., Ostrom, 1990), the simple distinction between two forms of punishment is clearly meaningful. If one group member takes too much of the common resource, restorative punishment would amount to the rightful return of these resources. Retributive punishment, on the other hand, would be imposing pain of some sort on the selfish individual for taking too much. Eriksson et al. (2016) found lower approval of punishers than of non-punishers, and higher approval of group punishment than of individual punishment, both for restorative and retributive punishment. Here we examine whether the same pattern holds cross-culturally.

### **Addressing the Research Questions Using Animations**

We have outlined three research questions regarding meta-norms about peer punishment in social dilemmas: (1) Are positive judgments of non-punishment culturally universal or particularly pronounced in individualistic societies? (2) Are positive judgments of group punishment culturally universal or particularly pronounced in collectivistic societies? (3) Are meta-norms the same across retributive and restorative punishment? To address these

questions, we asked participants to judge geometric animations that show a group of triangles jointly harvesting a common resource (see also Eriksson et al., 2016). As one triangle takes more than its share of resources, it is punished either by a single peer or by the entire group, and either by retributive punishment or by restorative punishment.

In a pioneering animation study, Heider and Simmel (1944) demonstrated that animated displays of abstract moving shapes are often automatically interpreted in terms of human actions, intentions, and emotions. As the animated display method does not rely on verbal stimuli, it has been useful to assess the attribution of intentions in abnormal populations (e.g., Abell, Happe, & Frith, 2000). Animations may similarly be suited for cross-cultural research, as they avoid the problem that linguistic stimuli in different languages can carry different connotations even when the literal meaning has been translated correctly. Importantly, the impressions of interpretations of emotions evoked by animated displays are fairly consistent across cultures (Rimé, Boulanger, Laubin, Richir, & Stroobants, 1985).

Drawing on the classic work of Michotte (1963), cognitive psychologists have used animations to investigate perceptions of causality from trajectory cues (Scholl & Tremoulet, 2000). Cross-cultural studies have adapted this method as a way of capturing cultural influences on causal judgments. For instance, Morris and Peng (1994) found that causal perceptions of social interactions, but not mechanical events, differed between American and Chinese students. Thus, prior research indicates that animations can successfully assess cultural differences in how social events are perceived. In the present work, we build on previous research by using animations of geometric shapes to measure social perceptions of peer punishment in a resource sharing situation, with the aim of assessing cultural differences in such social perceptions. In the next section (Study 1), we will focus on retributive punishment and examine appraisals of individual punishment, group punishment, and non-punishment, in eight different countries.



## STUDY 1: RETRIBUTIVE PUNISHMENT

### Method

*Selection of countries.* Study 1 was initiated during a workshop involving researchers from the United States, the Netherlands, Sweden, and Japan. These countries score high to medium on individualism (91, 80, 71, and 46, respectively) according to a well-known measure (Hofstede, Hofstede, & Minkov, 2010). In order to also include the lower range, we further invited researchers in Russia (39), United Arab Emirates (25, unofficial score obtained from <https://geert-hofstede.com>), China (20), and Pakistan (14).

*Participants.* Participants were recruited from student populations at universities in the selected countries. In the UAE, most participants came from other countries, mainly from countries across the Middle East and North Africa, which is consistent with the demographic makeup of the country. In other samples, only a very small minority of participants had a nationality different from the sampling country. Exclusion of participants whose nationality differ from the sampling country do not qualitatively change any results, either in the cross-national analyses or for the UAE in particular. Thus, we reported results based on the entire samples.

We excluded 28 participants who failed the manipulation check (see below), leaving 1,420 participants for data analysis (37% male; mean age 21.1 years,  $SD = 4.50$ , age range = 16 to 63; see Table 1).

---

INSERT TABLE 1 ABOUT HERE

---

*Design, procedure, and measures.* As an earlier study of animations showed no difference in results between online and laboratory conditions (Eriksson et al., 2016), participants took the study either online or in more controlled conditions at their universities. They were instructed to watch two animations of triangles and report how they felt about the behavior of different triangles. The two animations were displayed on a computer screen in a counterbalanced order.

Both animations showed a white stage where the action took place, with a collection of small circles in the center of the stage and four triangles of different colors (Blue, Green, Pink, or Purple) in four respective corners. The triangles took turns at harvesting the circles by moving one circle to their own corner (see Figure 1). After a while, Purple violated the norm by harvesting all the remaining circles in one go. The animations differed from this point on.

---

INSERT FIGURE 1 ABOUT HERE

---

In the *group punishment* animation, Green then moved to the center. Finding no circles there, Green went to convene Pink and Blue from their corners. Subsequently, they moved in synchrony to face the norm violator in its corner. Blue, who was then in the center, punished Purple by making a quick move toward Purple who lurched backward, while Green and Pink acted as bystanders. Thus, the animation represented retributive punishment in the form of aggression that stopped with no physical contact. Finally, all triangles returned to their own corners.

In the *individual punishment* animation, Green returned to its own corner after finding no circles in the center. The next to find the circles gone was Blue, who then went alone to confront the norm violator and delivered the same punishment as in the previous animation.

The animations can be accessed online (<https://pontustrimling.com/animations-for-cross-cultural-study/>).

After each animation, participants rated their approval of each of the four triangles' behavior using three items: (1) 'I think the BLUE triangle's behavior was appropriate'. (2) 'I would like to spend time with a person who behaves like the BLUE triangle'. and (3) 'If a person who behaves like the BLUE triangle belonged to my group I would consider that person to be a problem (rather than an asset) for the group'. (see also Eriksson et al., 2016). Responses were given on seven-point Likert scales (1 = *Strongly Disagree*, 7 = *Strongly Agree*), with the third item reverse-coded. The same three items were used for each triangle (with the color name changed to GREEN, PINK, and PURPLE). The three items about the Blue triangle were averaged to yield a *Punisher rating* for that animation. Similarly, the items about the Purple triangle were averaged into a *Deviant rating*. Finally, the six items about the Green and Pink triangles were averaged into an *Others rating*. A total of six rating measures were obtained across the two animations. Our main interest lies in the Punisher and Others ratings, both of which had good internal consistency ( $\alpha > 0.8$ ). Possibly due to floor effects, the Deviant ratings had weaker internal consistency ( $\alpha > 0.6$ ).

As a manipulation check, participants were asked to describe the difference between the two animations they had watched in free-text format. An example of a typical response was 'Only the blue triangle went up to the purple triangle instead of the blue, pink, and green like last time'. Participants were excluded if they stated no difference at all or made false claims about the difference between the animations (e.g., 'Each triangle only takes one circle', 'the purple one was looking around to see if anyone would catch it').

We also asked participants to what extent they found the triangles to look alive and goal-directed/intentional on 7-point scales (1 = *Strongly Disagree*, 7 = *Strongly Agree*). Overall, they perceived the triangles to look both alive ( $M = 5.25$ ,  $SD = 1.50$ ) and goal-

directed/intentional ( $M = 5.68$ ,  $SD = 1.25$ ). No participants were excluded for their responses on these items.

Finally, participants reported their demographic information including gender, age, country of residence, nationality, academic focus or major, each parent's educational level (university or not), and an estimate of family income relative to the average income. All questions were presented in either English (Pakistan, the United Arab Emirates, and the United States) or the dominant language (other countries). When the questions were translated from English to the dominant language, it was also back-translated by a separate translator, following standard practice to identify and resolve translation problems.

## Results and Discussion

Table 2 presents descriptive statistics of the ratings of three targets (Punisher, Deviant, and Others) in each of two animations (individual and group punishment) in eight countries (USA, Netherlands, Sweden, Japan, Russia, UAE, China, and Pakistan). Our research questions focus on two particular contrasts: individual-versus-group punishment in Punisher ratings, and individual Punisher *versus* non-punisher rating. In line with the design and aims of the studies, data were analyzed (in SPSS) using three-way mixed ANOVAs, followed by the two focused contrasts.

---

INSERT TABLE 2 ABOUT HERE

---

A  $3 \times 2 \times 8$  (Target [punisher, deviant, others]  $\times$  Animation [group punishment, individual punishment]  $\times$  Country [USA, Netherlands, Sweden, Japan, Russia, UAE, China, Pakistan]) mixed analysis of variance (ANOVA) on target ratings revealed a large main effect of target,  $F(2, 2804) = 1315.98$ ,  $p < .001$ ,  $\eta_p^2 = .48$ , reflecting much lower ratings for Deviant

than for Punisher, as well as somewhat lower ratings for Punisher than for. There was a weak main effect of animation,  $F(1, 1402) = 20.39, p < 0.001, \eta_p^2 = 0.01$ , and a somewhat larger interaction between target and animation,  $F(2, 2804) = 29.60, p < 0.001, \eta_p^2 = 0.02$ , reflecting higher Punisher ratings in the animation with group, compared to individual, punishment.

There was a medium-sized main effect of country,  $F(7, 1402) = 11.93, p < 0.001, \eta_p^2 = .06$ , a significant Country  $\times$  Target interaction,  $F(14, 2804) = 8.28, p < 0.001, \eta_p^2 = 0.04$ , a Country  $\times$  Animation interaction,  $F(7, 1402) = 5.64, p < 0.001, \eta_p^2 = 0.03$ , as well as a three-way interaction,  $F(14, 2804) = 6.46, p < 0.001, \eta_p^2 = 0.03$ . In particular, in the animation with individual punishment, the Punisher ratings were especially low and the Others ratings especially high in Sweden, the Netherlands, and the United States.

*Punishers vs. non-punishers.* Our first research question was how appropriate punishment is compared to non-punishment across countries. To answer this question, we analyzed the Punisher rating and the Others rating in the animation with individual punishment. We performed a  $2 \times 8$  (Target [punisher, others]  $\times$  Country [USA, Netherlands, Sweden, Japan, Russia, UAE, China, Pakistan]) mixed ANOVA on ratings. This analysis revealed a small main effect of country,  $F(7, 1408) = 7.49, p < 0.001, \eta_p^2 = 0.04$ , reflecting country variation in the absolute levels of Punisher and Others ratings. However, our main focus is the effect of target. The ANOVA revealed a medium effect of target,  $F(1, 1408) = 118.98, p < 0.001, \eta_p^2 = 0.08$ , which was moderated by country,  $F(7, 1408) = 14.05, p < 0.001, \eta_p^2 = 0.07$ . To illustrate this result, the left panel of Figure 2 shows 95% confidence intervals (CIs) of the difference between Others rating and Punisher rating in the animation with individual punishment in each country. Note that the mean difference was clearly positive in Sweden, the Netherlands, and the United States, whereas it did not differ from zero in China, Pakistan, Russia, and the United Arab Emirates. The result for Japan was at midway in between. This

shows that participants from our collectivistic cultures showed no clear preference for non-punishers over punishers. Replicating previous research, participants from our individualistic cultures did show such preference.

---

INSERT FIGURE 2 ABOUT HERE

---

*Individual vs. group punishment.* Our second research question was whether there is a universal preference for group punishment over individual punishment. We performed a  $2 \times 8$  (Animation [group punishment, individual punishment]  $\times$  Country [USA, Netherlands, Sweden, Japan, Russia, UAE, China, Pakistan]) mixed ANOVA on Punisher ratings. This analysis revealed a medium-sized main effect of country,  $F(7, 1408) = 11.74$ ,  $p < 0.001$ ,  $\eta_p^2 = .06$ , reflecting country variation in the absolute levels of Punisher ratings. Importantly, there was a medium main effect of animation,  $F(1, 1408) = 77.46$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.05$ , which was only weakly moderated by country,  $F(7, 1408) = 2.76$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.01$ . To illustrate this result, the left panel of Figure 3 shows 95% CIs of the difference in Punisher ratings between group punishment and individual punishment in each country. Note that the mean difference was positive in every country, although only marginally significant in some countries. In other words, the preference for group-based versus individually-based peer punishment seems to be shared across cultures.

---

INSERT FIGURE 3 ABOUT HERE

---

Study 1 suggests a culturally universal preference for group punishment, as well as cultural variation in relative appraisals of punishers (vs. non-punishers). To test the robustness of these findings, we further conducted Study 2 using restorative instead of retributive

punishment. Study 2 also addressed two procedural concerns. First, the individual punishment animation showed Green choosing not to punish before Blue chose to punish. It is theoretically possible that Green's behavior could make participants see the deviant behavior as less serious. Evidence against this possibility is the lack of a difference in the ratings of the deviant between the two conditions (see Table 2). Nonetheless, we altered the animation to avoid this concern in Study 2. Second, the approval items were somewhat ambiguous on how participants should relate to the animations, so we rephrased the items in Study 2.

## STUDY 2: RESTORATIVE PUNISHMENT

### Method

*Participants.* We recruited student samples from the same eight countries as in Study 1. After excluding 26 participants who failed the manipulation check, there remained 1,183 participants for analysis (34% male; mean age 21.8 years,  $SD = 5.27$ , age range = 17 to 59; see Table 1).

*Design, procedure, and measures.* The procedure was like Study 1 except for the two animations, which showed restorative instead of retributive punishment. Specifically, the group punishment animation was similar to the group punishment animation in Study 1, except that the meeting ended with all three triangles entering Purple's corner, where they jointly removed all the circles that were present, and brought them back to the center. In other words, the common resource was restored to what it would have been if the Purple triangle had not engaged in harvesting at all. Another possible version of restorative punishment would have been to restore only that part of the common resource that Purple took in its last move, which would be a slightly weaker punishment than the version we used. A previous study with US participants found that both versions of restorative punishment were appraised more negatively than no punishment, and the difference in appraisal between the two versions of restorative punishment was smaller than the difference in appraisal between the weaker restorative punishment and no punishment (Eriksson et al., 2016, Study 3). Thus, the exact level of restoration may not matter much.

In the *individual punishment* animation, Green never left its corner before Blue found the circles gone and went alone to carry out the same act of restorative punishment as the group did in the previous animation. Again, the animations can be accessed online (<https://pontusstrimling.com/animations-for-cross-cultural-study/>).



Study 2 used the same measures as employed in Study 1 with one important difference: the approval items were rephrased to clarify that we wanted participants to rate the animations as if the situation occurred in their own community. The revised items employed in Study 2 all started with ‘If I were part of a group and a person in that group behaved as the [BLUE / GREEN / PINK / PURPLE] triangle...’, followed by (1) ‘I would consider it appropriate’, (2) ‘I would like to spend time with that person’, and (3) ‘I would consider that person to be a problem for the group (rather than an asset)’. As in Study 1, Punisher and Others ratings had good internal consistency ( $\alpha > 0.85$ ) and Deviant ratings somewhat weaker internal consistency ( $\alpha > 0.65$ ).

## Results and Discussion

Descriptive statistics of the rating measures are presented in Table 3. A  $3 \times 2 \times 8$  (Target [deviant, punisher, others]  $\times$  Animation [group punishment, individual punishment]  $\times$  Country [USA, Netherlands, Sweden, Japan, Russia, UAE, China, Pakistan]) mixed ANOVA revealed the same pattern of effects as in Study 1, only of somewhat different strength. There was a large main effect of target,  $F(2, 2350) = 2666.69, p < 0.001, \eta_p^2 = 0.69$ , again reflecting that Deviant ratings were much lower than Punisher ratings and that Punisher ratings were somewhat lower than Others ratings. There was a medium main effect of animation,  $F(1, 1175) = 92.45, p < 0.001, \eta_p^2 = 0.07$ , as well as a medium Target  $\times$  Animation interaction,  $F(2, 2350) = 90.12, p < 0.001, \eta_p^2 = 0.07$ . As in Study 1, this reflected higher Punisher ratings in the animation involving group, compared to individual, punishment.

Moreover, there was a medium main effect of country,  $F(7, 1175) = 10.36, p < 0.001, \eta_p^2 = 0.06$ , a significant Country  $\times$  Target interaction,  $F(14, 2350) = 4.40, p < 0.001, \eta_p^2 = 0.03$ , a significant Country  $\times$  Animation interaction,  $F(7, 1175) = 4.12, p < 0.001, \eta_p^2 = 0.02$ , as well as a significant three-way interaction,  $F(14, 2350) = 2.88, p < 0.001, \eta_p^2 = 0.02$ . These

findings reflected that for individual punishment, countries with relatively low Other ratings (Russia, China, UAE, Pakistan) tended to have relatively high Punisher ratings.

---

INSERT TABLE 3 ABOUT HERE

---

*Punishers vs. non-punishers.* We then analyzed how appropriate punishment was compared to non-punishment across countries using a  $2 \times 8$  (Target [punisher, others]  $\times$  Country [USA, Netherlands, Sweden, Japan, Russia, UAE, China, Pakistan]) mixed ANOVA for the animation involving individual punishment. There was a small main effect of country,  $F(7, 1175) = 4.40$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.03$ , reflecting country variation in the absolute levels of Punisher and Others ratings. We found a medium-to-large effect of target,  $F(1, 1175) = 141.67$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.11$ , which was moderated by country,  $F(7, 1408) = 4.96$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.03$ . The right panel of Figure 2 illustrates the favorable ratings of non-punishers over punishers in all countries. The results are not identical to Study 1 but the overall pattern is quite similar: USA, Sweden, Netherlands, and Japan showed a significant preference for non-punishers in both studies, whereas Pakistan and UAE were neutral between non-punishers and punishers in both studies.

*Individual vs. group punishment.* We then performed a  $2 \times 8$  (Animation [group punishment, individual punishment]  $\times$  Country [USA, Netherlands, Sweden, Japan, Russia, UAE, China, Pakistan]) mixed ANOVA on Punisher ratings. Similar to Study 1, there was a medium-sized main effect of country,  $F(7, 1175) = 9.40$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.05$ . Our focus here is the effect of animation. The ANOVA revealed a large main effect of animation,  $F(1, 1175) = 192.81$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.14$ , which was not significantly moderated by country,  $F(7, 1175) = 0.56$ ,  $p = 0.83$ ,  $\eta_p^2 = 0.00$ . This is illustrated in the right panel of Figure 3, which shows that, as in Study 1, all samples rated punishment higher when implemented by the group than by an individual. Thus it appears that, regardless of whether punishment takes a retributive or restorative form, people everywhere prefer that punishment for defecting in a social dilemma be enacted by the group.

## **FURTHER ANALYSES ACROSS THE TWO STUDIES**

### **Comparing Retributive and Restorative Punishment**

Our third research question was whether meta-norms are similar for retributive and restorative punishment. Findings from two studies were largely consistent across retributive and restorative punishment. Specifically, punishers were consistently rated negatively in Western countries and Japan, and consistently more neutrally in Pakistan, UAE, Russia, and, to some extent, China (see Figure 2). However, it should be noted that Japan and China exhibited more negative appraisals of punishers for restorative punishment than for retributive punishment. Finally, there was a universal preference for punishment by a group of peers instead of an individual peer across cultures as well as across retributive and restorative punishment (see Figure 3).

### Meta-Norms and Cultural Dimensions

Finally, we examined the correlations between meta-norms and country scores on Hofstede's six cultural dimensions. In the introduction we mentioned power distance and individualism. The other four are masculinity (a cultural preference for achievement and competition), uncertainty avoidance (the degree to which the members of a society feel uncomfortable with uncertainty and ambiguity), long term orientation (prioritizing preparations for the future), and indulgence (relatively free gratification of human drives related to enjoying life and having fun). Hofstede et al. (2010) provides official scores for seven of eight countries. The exception is UAE, for which unofficial scores are available for the first four dimensions (<https://geert-hofstede.com/>); for long term orientation and indulgence in UAE we used the official scores on 'Arab countries' (Hofstede et al., 2010).

To obtain a single estimate of the meta-norm against peer punishment in each country, we pooled the data from Studies 1 and 2, and calculated the mean difference in ratings of the non-punisher and the punisher (as in Figure 2). These meta-norm scores were very strongly correlated with individualism,  $r(6) = 0.84$ ,  $p = 0.010$ , power distance,  $r(6) = -0.86$ ,  $p = 0.005$ , and indulgence,  $r(6) = 0.89$ ,  $p = 0.003$  (see the scatterplots in Figure 4). Individualism was also strongly correlated with power distance,  $r(6) = -0.72$ , and indulgence,  $r(6) = 0.93$ . In other words, these three cultural dimensions cannot really be teased apart for this particular sample of countries. The correlations between meta-norm scores and the remaining three dimensions were modest and non-significant,  $r_s < 0.5$ ,  $p_s > 0.2$ .

---

INSERT FIGURE 4 ABOUT HERE

---

## DISCUSSION

The purpose of this research was to make a first attempt to systematically investigate appraisals of peer punishment across cultures. One major finding was a culturally universal tendency to judge punishment as more appropriate when implemented by a group than an individual. This effect was found when the group designated one member to deliver retributive punishment to the norm-breaker while the others were watching (Study 1), as well as when the entire group was actively involved in restorative punishment (Study 2). We expect the preference for group punishment to arise specifically in situations involving only peers, as such situations make it unclear that any specific peer has the authority to punish. However, other mechanisms may also be involved. For instance, a recent study found that punishing as a group decreases the individuals' feelings of responsibility for the harm done to the punished party (Molenmaker, De Kwaadsteniet, & Van Dijk, 2016), and it is plausible that a similar mechanism could operate among observers. More research is needed to better understand the causes and scope of the preference for group punishment.

Another major finding concerned systematic differences in the appraisal of punishers (vs. non-punishers) across countries. Economic experiments on peer punishment in social dilemmas have yielded a paradox. Although punishment can effectively promote cooperation (for a meta-analysis, see Balliet, Mulder, & Van Lange, 2011), an individual who punishes will often receive very little approval, sometimes even disapproval (Cinyabuguma et al., 2006; Eriksson et al., 2016; Kiyonari & Barclay, 2008; Mathew, 2017; Strimling & Eriksson, 2014). Here we replicated the finding that a peer who punishes often tends to be rated less favorably than another peer who refrains from punishing. However, this lack of encouragement to use punishment does not necessarily have negative consequences for the group. Outside the laboratory it remains challenging to reliably measure the costs and benefits associated with peer punishment (Guala, 2012). However, it is clear that peer punishment may engender

social costs and escalate conflicts within the group (Nikiforakis & Engelmann, 2011; Nikiforakis & Normann, 2008), and that it is often used with little regard for whether it benefits the group (Eriksson, Cownden, Ehn, & Strimling, 2014). Therefore, social discouragement of peer punishment might overall be a good thing (Eriksson, Strimling, & Ehn, 2013). After all, non-punitive alternatives such as gossip may be equally effective (Wu, Balliet, & Van Lange, 2016).

The most novel aspect of our findings is that the negative appraisal of peer punishers was moderated by culture. In particular, the appraisals of punishers tended to be more negative in more individualistic societies, consistent with the notion that an emphasis on autonomy leads to stronger rights for individuals to behave counter-normatively (Leung & Stephan, 2001). The same basic pattern of cross-cultural variation was observed across retributive punishment (Study 1) and restorative punishment (Study 2). This finding was obtained despite the fact that our study was limited to student samples, which may underestimate cultural differences (Henrich et al., 2010). Thus, representative non-student samples would likely exhibit greater cultural differences in meta-norms than those demonstrated here.

Based on our theory and data, we tentatively conclude that individualism-collectivism may influence appraisals of punishment in social dilemmas. However, we were unable to disentangle the influence of individualism from possible influences of power distance and indulgence with this sample of countries. Power distance and indulgence could be relevant too. Given that power distance reveals the degree to which inequality in power is culturally accepted, it is conceivable that an individual who steps up to punish may be perceived as more powerful than his/her peers, and that peer punishment may be more condoned in cultures with higher power distance. Or consider indulgence, which quantifies the extent to which a society allows relatively free gratification of enjoying life and having fun. Peer

punishers may be perceived as killjoys, which would explain why peer punishment was less condoned in cultures that score high on indulgence. A larger number of country-level measures of meta-norms are required to confirm or disconfirm the specific role of each of these cultural dimensions.

Our work on cultural variation in meta-norms is distinct from previous research on cultural variation in the *effectiveness* of peer punishment at promoting cooperation in economic experiments (for a meta-analysis, see Balliet & Van Lange, 2013; Herrmann, Thöni, & Gächter, 2008). The distinction between these questions is underscored by our finding that Western countries exhibit especially negative appraisals of peer punishers, although previous economic experiments suggest that punishment is especially effective in Western countries.

## CONCLUSION

Starting with Axelrod (1986), a large body of literature – drawing on game theory, economic experiments, and evolutionary arguments – has been devoted to meta-norms or second-order punishment in social dilemmas (e.g., Binmore, 1998; Fehr & Gächter, 2000; Gintis, 2009; Henrich & Boyd, 2001). The present research breaks new ground by closely examining how punishers and different forms of punishment are evaluated across cultures. Two studies revealed consistent yet novel evidence that meta-norms about peer punishment exhibit culturally universal features as well as systematic cultural differences. These findings help us understand the relation between culture and the enforcement of norms, and ultimately provide insight about how societal and organizational trust may be promoted and sustained. However, cultures are not static. Studies of single countries indicate an ongoing move toward a greater emphasis on individualism in United States, Japan, and the United Kingdom (Greenfield, 2013; Matsumoto et al., 1996; Twenge et al., 2013). Moreover, Inglehart's (1997) theory of a large-scale change in values toward autonomy and self-expression also indicate a general

pattern of change toward greater individualism. Based on our findings, we might therefore expect a corresponding trend in meta-norms toward decreasing approval of peer punishment. To track such trends, it would be interesting to see how the same measures of meta-norms change at regular intervals.

## NOTES

This research was funded by the Swedish Research Council [grant number 2009-2390]; the Knut and Alice Wallenberg Foundation [grant number 2015.0005]; Japan Society for the Promotion of Science [KAKENHI grant numbers 15H05730 and 25118004]; the Russian Academic Excellence Project '5-100'; and a fellowship from China Scholarship Council (201206040030) awarded to Junhui Wu.

KE and PS designed the study. PA designed the animations and managed the data. KE analyzed the data and wrote the manuscript. Other authors collected data and/or provided critical feedback on the design and/or the manuscript.

## REFERENCES

- Abell, F., Happe, F., & Frith, U. 2000. Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development*, 15(1): 1–16.
- Archer, J. 2004. Sex differences in aggression in real-world settings: A meta-analytic review. *Review of General Psychology*, 8(4): 291–322.
- Axelrod, R. M. 1986. An evolutionary approach to norms. *American Political Science Review*, 80(4): 1095–1111.
- Balliet, D., Mulder, L. B., & Van Lange, P. A.M. 2011. Reward, punishment, and cooperation: A meta-analysis. *Psychological Bulletin*, 137(4): 594–615.



- Balliet, D., & Van Lange, P. A. M. 2013. Trust, punishment, and cooperation across 18 societies: A meta-analysis. *Perspectives on Psychological Science*, 8: 363-379.
- Binmore, K. G. 1998. *Game theory and the social contract: Just playing* (Vol. 2). MIT press.
- Bond, M. H., Wan, K. C., Leung, K., & Giacalone, R. A. 1985. How are responses to verbal insult related to cultural collectivism and power distance? *Journal of Cross-Cultural Psychology*, 16(1): 111–127.
- Brauer, M., & Chaurand, N. 2010. Descriptive norms, prescriptive norms, and social control: An intercultural comparison of people’s reactions to uncivil behaviors. *European Journal of Social Psychology*, 40(3): 490–499.
- Chiu, C. Y., Morris, M. W., Hong, Y. Y., & Menon, T. (2000). Motivated cultural cognition: The impact of implicit theories on dispositional attribution varies as a function of need for closure. *Journal of Personality and Social Psychology*, 78: 247–259.
- Cinyabuguma, M., Page, T., & Putterman, L. 2006. Can second-order punishment deter perverse punishment? *Experimental Economics*, 9(3): 265–279.
- Cohen, R. L. 2016. Restorative justice. In C. Sabbagh & M. Schmitt (Eds.), *Handbook of social justice theory and research* (pp. 257-272). Springer New York.
- Eriksson, K., Andersson, P.A., & Strimling, P. 2016. Moderators of the disapproval of peer punishment. *Group Processes & Intergroup Relations*, 19(2): 152–168.
- Eriksson, K., Cownden, D., Ehn, M., & Strimling, P. 2014. ‘Altruistic’ and ‘antisocial’ punishers are one and the same. *Review of Behavioral Economics*, 1(3): 209-221.
- Eriksson, K., Strimling, P., & Coultas, J. C. 2015. Bidirectional associations between descriptive and injunctive norms. *Organizational Behavior and Human Decision Processes*, 129: 59–69.
- Eriksson, K., Strimling, P., & Ehn, M. 2013. Ubiquity and efficiency of restrictions on informal punishment rights. *Journal of Evolutionary Psychology*, 11(1): 17–34.

Fehr, E. & Gächter, S. 2002. Altruistic punishment in humans. *Nature*, 415(6868): 137–140.

Felson, R. B. (1981). An interactionist approach to aggression. In J. T. Tedeschi (Ed.)

*Impression management theory and social psychological research*: 181–199. New York: Academic Press.

Gelfand, M. J., Nishii, L. H., & Raver, J. L. 2006. On the nature and importance of cultural tightness-looseness. *Journal of Applied Psychology*, 91(6): 1225–1244.

Gintis, H. 2009. *The bounds of reason: Game theory and the unification of the behavioral sciences*. Princeton University Press.

Greenfield, P. M. 2013. The changing psychology of culture from 1800 through 2000. *Psychological Science*, 24(9): 1722–1731.

Guala, F. 2012. Reciprocity: Weak or strong? What punishment experiments do (and do not) demonstrate. *Behavioral and Brain Sciences*, 35, 1–15.

Heider, F., & Simmel, M. 1944. An experimental study of apparent behavior. *American Journal of Psychology*, 57(2): 243–259.

Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. 2002. What's wrong with cross-cultural comparisons of subjective Likert scales?: The reference-group effect. *Journal of Personality and Social Psychology*, 82(6): 903–918.

Henrich, J., & Boyd, R. 2001. Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of Theoretical Biology*, 208(1): 79–89.

Henrich, J., Heine, S. J., & Norenzayan, A. 2010. The weirdest people in the world?. *Behavioral and Brain Sciences*, 33(2–3): 61–83.

Herrmann, B., Thöni, C., & Gächter, S. 2008. Antisocial punishment across societies. *Science*, 319(5868): 1362–1367.

- Hjörne, E., & Larsson, P. 2012. Beyond teaching and learning: Disciplining boys in remedial classes during the 1960s in Sweden. In *Learning, social interaction and diversity—Exploring identities in school practices*: 139–154. Sense Publishers.
- Hofstede, G. 1980. *Culture's consequences: International differences in work-related values*. Beverly Hills, CA: Sage.
- Hofstede, G., Hofstede, G.J., & Minkov, M. 2010. *Cultures and organizations: Software of the mind*. Revised and expanded 3rd edition. New York: McGraw-Hill.
- Inglehart, R. 1997. *Modernization and postmodernization: Cultural, economic, and political change in 43 societies* (Vol. 19). Princeton, NJ: Princeton University Press.
- Kelley, H. H. 1971. Moral evaluation. *American Psychologist*, 26(3): 293–300.
- Kiyonari, T. and Barclay, P. 2008. Cooperation in social dilemmas: Free riding may be thwarted by second-order reward rather than by punishment. *Journal of Personality and Social Psychology*, 95(4): 826–842.
- Kleinsmith, A., De Silva, P. R., & Bianchi-Berthouze, N. 2006. Cross-cultural differences in recognizing affect from body posture. *Interacting with Computers*, 18(6): 1371–1389.
- Kuwabara, K., Yu, S., Lee, A. J., & Galinsky, A. D. 2016. Status decreases dominance in the West but increases dominance in the East. *Psychological Science*, 27(2): 127-137.
- Leung, A. K. Y., & Cohen, D. 2011. Within-and between-culture variation: Individual differences and the cultural logics of honor, face, and dignity cultures. *Journal of Personality and Social Psychology*, 100(3): 507–526.
- Leung, K., & Stephan, W. G. 2001. Social justice from a cultural perspective. In D. Matsumoto (Ed.), *The handbook of culture and psychology* (pp. 375–410). New York, NY: Oxford University Press.
- Mathew, S. 2017. How the second-order free rider problem is solved in a small-scale society. *American Economic Review*, 107(5): 578–581.

- Matsumoto, D., Kudoh, T., & Takeuchi, S. 1996. Changing patterns of individualism and collectivism in the United States and Japan. *Culture & Psychology*, 2(1): 77–107.
- Menon, T., Morris, M. W., Chiu, C. Y., & Hong, Y. Y. 1999. Culture and the construal of agency: Attribution to individual versus group dispositions. *Journal of Personality and Social Psychology*, 76: 701–717.
- Michotte, A. 1963. *The perception of causality*. Routledge.
- Molenmaker W.E., De Kwaadsteniet E.W. & Van Dijk E. 2016. The impact of personal responsibility on the (un)willingness to punish non-cooperation and rewards cooperation, *Organizational Behavior and Human Decision Processes*, 134: 1–15.
- Morris, M. W., Menon, T., & Ames, D. R. 2001. Culturally conferred conceptions of agency: A key to social perception of persons, groups, and other actors. *Personality and Social Psychology Review*, 5(2): 169–182.
- Morris, M. W., & Peng, K. 1994. Culture and cause: American and Chinese attributions for social and physical events. *Journal of Personality and Social Psychology*, 67(6): 949–971.
- Nikiforakis, N., & Engelmann, D. 2011. Altruistic punishment and the threat of feuds. *Journal of Economic Behavior & Organization*, 78: 319-332.
- Nikiforakis, N., & Normann, H. T. 2008. A comparative statics analysis of punishment in public-good experiments. *Experimental Economics*, 11(4): 358-369.
- Ostrom, E. 1990. *Governing the commons*. Cambridge University Press.
- Rimé, B., Boulanger, B., Laubin, P., Richir, M., & Stroobants, K. 1985. The perception of interpersonal emotions originated by patterns of movement. *Motivation and Emotion*, 9(3): 241–260.
- Roos, P., Gelfand, M., Nau, D., & Carr, R. 2014. High strength-of-ties and low mobility enable the evolution of third-party punishment. *Proceedings of the Royal Society of London B: Biological Sciences*, 281(1776): 20132661.

- Roos, P., Gelfand, M., Nau, D., & Lun, J. 2015. Societal threat and cultural variation in the strength of social norms: An evolutionary basis. *Organizational Behavior and Human Decision Processes*, 129: 14–23.
- Scholl, B. J., & Tremoulet, P. D. 2000. Perceptual causality and animacy. *Trends in Cognitive Sciences*, 4(8): 299–309.
- Strimling, P., & Eriksson, K. 2014. Regulating the regulation: Norms about punishment. In P. A. M. Van Lange, T. Yamagishi, & B. Rockenbach (Eds.) *Reward and punishment in social dilemmas*: 52–69. Oxford: Oxford University Press.
- Tedeschi, J. T., Gaes, G. G., & Rivera, A. N. 1977. Aggression and the use of coercive power. *Journal of Social Issues*, 33(1): 101–125.
- Tedeschi, J. T., Smith, R. B., & Brown, R. C. 1974. A reinterpretation of research on aggression. *Psychological Bulletin*, 81(9): 540–562.
- Triandis, H. C., & Gelfand, M. J. 1998. Converging measurement of horizontal and vertical individualism and collectivism. *Journal of Personality and Social Psychology*, 74(1): 118–128.
- Twenge, J. M., Campbell, W. K., & Gentile, B. 2013. Changes in pronoun use in American books and the rise of individualism, 1960-2008. *Journal of Cross-Cultural Psychology*, 44(3): 406–415.
- Van Lange, P. A. M., Rockenbach, B., & Yamagishi, T. 2014. Reward and punishment in social dilemmas: An introduction. In P.A.M. Van Lange, B., Rockenbach, & T. Yamagishi, T. (Eds.) *Reward and punishment in social dilemmas* (pp. 1–14). New York: Oxford University Press.
- Wang, C. S., & Leung, A. K. Y. 2010. The cultural dynamics of rewarding honesty and punishing deception. *Personality and Social Psychology Bulletin*, 36(1): 1529–1542.
- Wenzel, M., & Okimoto, T. G. 2016. Retributive justice. In C. Sabbagh & M. Schmitt (Eds.), *Handbook of social justice theory and research* (pp. 237-256). Springer New York.

- Wu, J., Balliet, D., & Van Lange, P. A. M. 2016. Gossip versus punishment: The efficiency of reputation to promote and maintain cooperation. *Scientific Reports*, 6: 23919.
- Yamagishi, T. 1986. The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, 51(1): 110–116.

Table 1. Sample sizes and sample characteristics

| Country     | Study 1 – retributive punishment |        |                         | Study 2 – restorative punishment |        |                         |
|-------------|----------------------------------|--------|-------------------------|----------------------------------|--------|-------------------------|
|             | <i>N</i>                         | Male % | <i>M</i> <sub>age</sub> | <i>N</i>                         | Male % | <i>M</i> <sub>age</sub> |
| USA         | 163 (4)                          | 33%    | 18.8                    | 131 (4)                          | 43%    | 19.0                    |
| Netherlands | 120 (0)                          | 27%    | 20.7                    | 121 (13)                         | 27%    | 21.5                    |
| Sweden      | 176 (4)                          | 22%    | 27.9                    | 202 (3)                          | 26%    | 27.6                    |
| Japan       | 178 (4)                          | 60%    | 20.6                    | 128 (1)                          | 55%    | 20.7                    |
| Russia      | 165 (5)                          | 23%    | 20.7                    | 113 (1)                          | 29%    | 21.7                    |
| UAE         | 190 (9)                          | 41%    | 19.9                    | 204 (2)                          | 37%    | 20.2                    |
| China       | 150 (0)                          | 13%    | 19.4                    | 179 (1)                          | 13%    | 19.0                    |
| Pakistan    | 278 (2)                          | 57%    | 20.5                    | 105 (1)                          | 56%    | 23.3                    |
| Total       | 1420<br>(28)                     | 37%    | 21.1                    | 1183<br>(26)                     | 34%    | 21.8                    |

*Note.* *N* = Sample size after exclusions (number of exclusions in parentheses). Male % = the percentage of male participants in the sample. *M*<sub>age</sub> = Mean age.

Table 2. *Descriptive statistics of ratings in Study 1 (Retributive Punishment)*

| Country     | Individual punishment |                |                | Group punishment |                |                |
|-------------|-----------------------|----------------|----------------|------------------|----------------|----------------|
|             | Punisher              | Deviant        | Others         | Punisher         | Deviant        | Others         |
| USA         | 3.65<br>(1.50)        | 2.12<br>(1.15) | 4.88<br>(1.18) | 4.05<br>(1.50)   | 1.96<br>(1.03) | 4.83<br>(1.30) |
| Netherlands | 3.72<br>(1.46)        | 2.27<br>(1.12) | 4.88<br>(0.97) | 3.93<br>(1.40)   | 2.29<br>(1.15) | 4.34<br>(1.33) |
| Sweden      | 3.26<br>(1.75)        | 2.13<br>(1.20) | 4.69<br>(1.21) | 3.41<br>(1.68)   | 2.24<br>(1.21) | 3.99<br>(1.54) |
| Japan       | 4.02<br>(1.29)        | 2.42<br>(1.05) | 4.66<br>(0.91) | 4.63<br>(1.15)   | 2.38<br>(1.00) | 4.52<br>(0.99) |
| Russia      | 4.01<br>(1.39)        | 2.44<br>(1.25) | 3.95<br>(0.99) | 4.18<br>(1.42)   | 2.49<br>(1.24) | 4.21<br>(1.20) |
| UAE         | 3.97<br>(1.65)        | 2.20<br>(1.21) | 4.14<br>(1.09) | 4.18<br>(1.58)   | 2.34<br>(1.28) | 4.37<br>(1.38) |
| China       | 3.85<br>(1.44)        | 2.40<br>(1.33) | 3.98<br>(0.97) | 4.38<br>(1.45)   | 2.29<br>(1.26) | 4.34<br>(1.28) |
| Pakistan    | 4.24<br>(1.69)        | 2.49<br>(1.45) | 4.41<br>(1.31) | 4.56<br>(1.61)   | 2.49<br>(1.45) | 4.71<br>(1.33) |
| Total       | 3.87<br>(1.57)        | 2.32<br>(1.25) | 4.43<br>(1.16) | 4.20<br>(1.54)   | 2.32<br>(1.24) | 4.44<br>(1.33) |

*Note.* Cell entries are mean values with standard deviations within parentheses. Sample sizes are reported in the methods section.



Table 3. Descriptive statistics of ratings in Study 2 (Restorative Punishment)

| Country     | Individual punishment |                |                | Group punishment |                |                |
|-------------|-----------------------|----------------|----------------|------------------|----------------|----------------|
|             | Punisher              | Deviant        | Others         | Punisher         | Deviant        | Others         |
| USA         | 4.73<br>(1.52)        | 2.27<br>(1.23) | 5.32<br>(1.19) | 5.43<br>(1.35)   | 2.22<br>(1.21) | 5.49<br>(1.26) |
| Netherlands | 4.58<br>(1.58)        | 2.10<br>(0.96) | 5.55<br>(1.09) | 5.44<br>(1.26)   | 2.23<br>(1.09) | 5.41<br>(1.17) |
| Sweden      | 4.21<br>(1.59)        | 2.14<br>(1.29) | 5.07<br>(1.34) | 4.75<br>(1.59)   | 2.10<br>(1.09) | 4.76<br>(1.48) |
| Japan       | 4.01<br>(1.50)        | 2.17<br>(1.13) | 5.29<br>(1.04) | 4.67<br>(1.46)   | 2.11<br>(1.07) | 4.73<br>(1.31) |
| Russia      | 4.55<br>(1.40)        | 2.29<br>(1.19) | 5.00<br>(1.20) | 5.34<br>(1.14)   | 2.31<br>(1.09) | 5.22<br>(1.14) |
| UAE         | 4.85<br>(1.55)        | 1.96<br>(1.06) | 5.09<br>(1.19) | 5.50<br>(1.34)   | 1.94<br>(1.11) | 5.49<br>(1.26) |
| China       | 4.35<br>(1.56)        | 1.90<br>(1.02) | 5.05<br>(1.19) | 5.05<br>(1.41)   | 1.92<br>(1.04) | 4.96<br>(1.43) |
| Pakistan    | 4.52<br>(1.73)        | 2.23<br>(1.34) | 4.84<br>(1.42) | 5.27<br>(1.52)   | 2.31<br>(1.35) | 5.30<br>(1.27) |
| Total       | 4.48<br>(1.57)        | 2.11<br>(1.16) | 5.14<br>(1.23) | 5.16<br>(1.43)   | 2.11<br>(1.13) | 5.15<br>(1.34) |

*Note.* Cell entries are mean values with standard deviations within parentheses. Sample sizes are reported in the methods section.

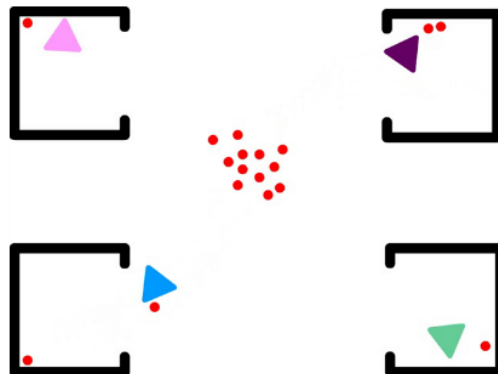


Figure 1. The triangles at their respective corners, with the blue triangle moving back from collecting a circle from the center

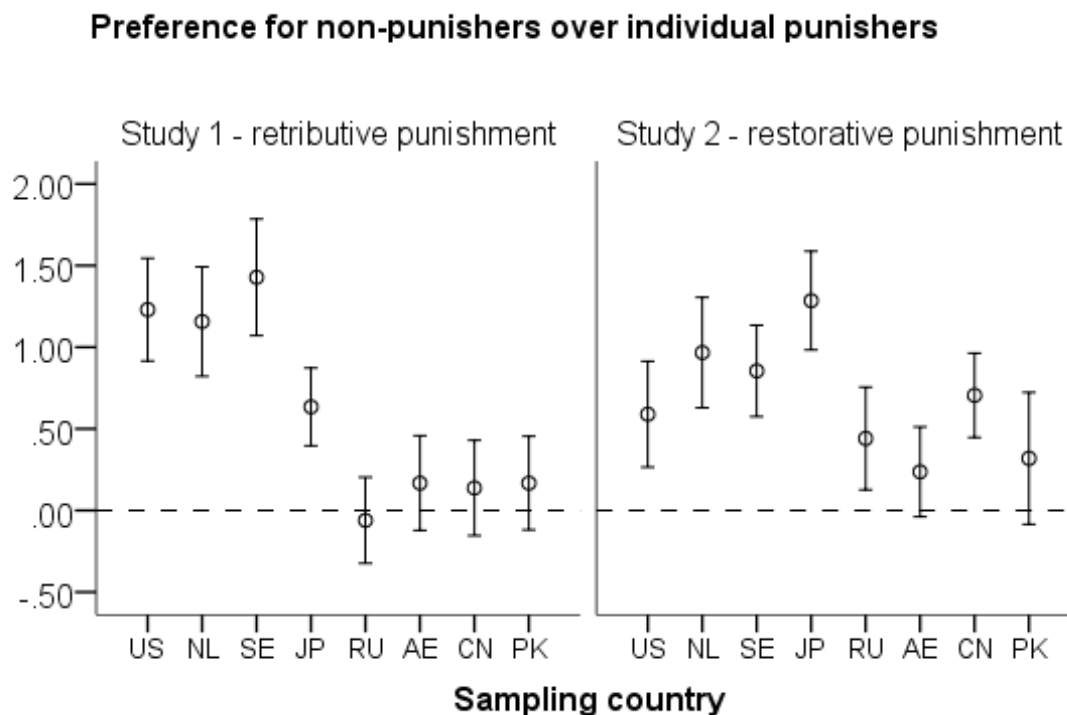


Figure 2. The preference for non-punishers over punishers in the individual punishment animation across countries in Study 1 (left) and Study 2 (right). Error bars represent 95% confidence intervals. The dashed line of the zero level indicates the same ratings for non-punishers and punishers. Countries are ordered from the most individualistic (United States) to the least individualistic (Pakistan).

### Preference for group punishment over individual punishment

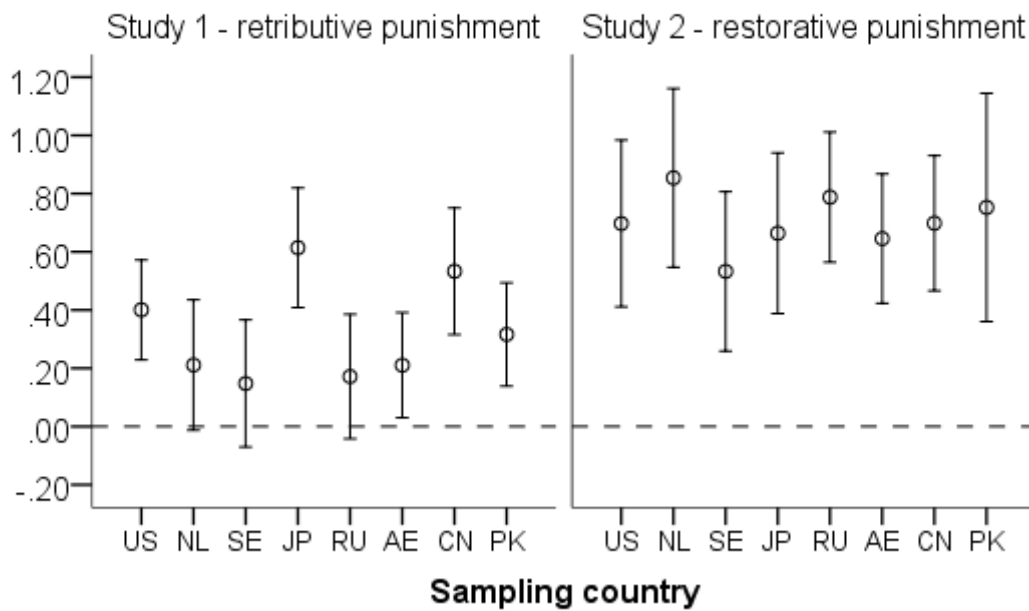


Figure 3. The preference for group punishment over individual punishment across countries in Study 1 (left) and Study 2 (right). Error bars represent 95% confidence intervals. The dashed line of the zero level indicates the same punisher ratings for group punishment and individual punishment. Countries are ordered as in Figure 2.

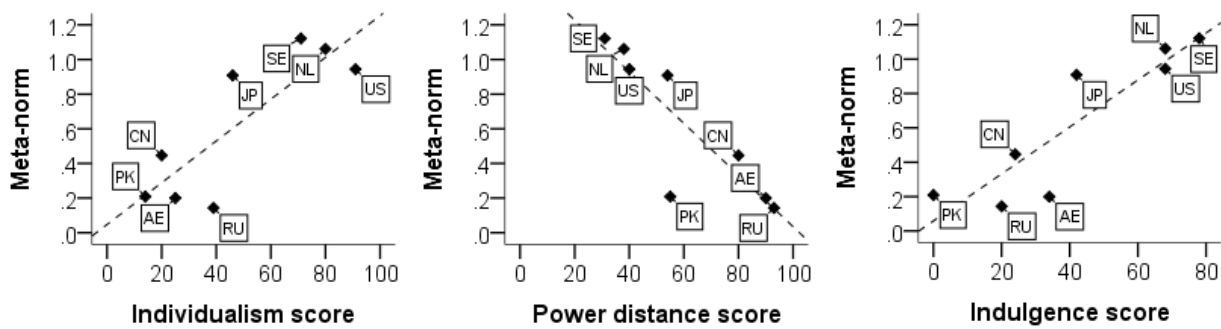


Figure 4. Scatterplots of estimates of the meta-norm about peer punishment as a function of country scores of individualism (left), power distance (middle), and indulgence (right).