Preprint

This is the submitted version of a paper published in Management and Organization Review.

Citation for the original published paper (version of record):


Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:mdh:diva-38010
Cultural Universals and Cultural Variation in Meta-Norms about Peer Punishment

Kimmo Eriksson¹,², Pontus Strimling³, Per A. Andersson⁴, Mark Aveyard⁵, Markus Brauer⁶, Vladimir Gritskov⁷, Toko Kiyonari⁸, David M. Kuhlman⁹, Angela T. Maitner⁵, Zoi Manesi¹⁰, Catherine Molho¹⁰, Leonard S. Peperkoorn¹⁰, Muhammad Rizwan¹¹, Adam W. Stivers¹², Qirui Tian¹³, Paul A. M. Van Lange¹⁰, Irina Vartanova³,¹⁴, Junhui Wu¹⁵, Toshio Yamagishi¹⁶

¹Centre for the Study of Cultural Evolution, Stockholm University, Sweden
²School of Education, Culture and Communication, Mälardalen University, Sweden
³Institute for Futures Studies, Sweden
⁴Department of Management and Engineering, Linköping University, Sweden
⁵Department of International Studies, American University of Sharjah, United Arab Emirates
⁶Department of Psychology, University of Wisconsin-Madison, USA
⁷Department of Psychology, Saint Petersburg State University, Russian Federation
⁸School of Social Informatics, Aoyama Gakuin University, Japan
⁹Department of Psychological and Brain Sciences, University of Delaware, USA
¹⁰Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam, Netherlands
¹¹Institute of Clinical Psychology, University of Karachi, Pakistan
¹²Psychology Department, Gonzaga University, USA
¹³Department of Psychology, Shandong Normal University, China
¹⁴Laboratory for Comparative Social Research, National Research University Higher School of Economics, Russian Federation
¹⁵Institute of Developmental Psychology, Beijing Normal University, China
¹⁶Hitotsubashi University, Japan

Corresponding author: Kimmo Eriksson, Mälardalen University, School of Education, Culture and Communication, Box 883, SE-72123 Västerås, Sweden. Email:

kimmo.eriksson@mdh.se Phone: +46 21101533.

Author contributions: 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.
Title: Cultural Universals and Cultural Differences in Meta-Norms about Peer Punishment

Running title: Culture and Meta-Norms about Peer Punishment

Abstract

Violators of cooperation norms often experience informal sanctions such as aggressive confrontations or informal fines. How such norm enforcement is itself judged by others can be regarded as meta-norms. We examined whether meta-norms vary between cultures by having students in eight countries judge animations in which an agent who over-harvested a common resource was sanctioned either by a single peer or by a group of peers. The type of sanction, aggressive confrontation or informal fine, varied between two studies. We find evidence for cultural universals and differences in meta-norms. Across the countries, peer punishment was consistently judged as less appropriate when it was managed by an individual rather than a group. Across the two studies, differences between countries were revealed in judgments of punisher versus non-punishers. In individualistic countries (Sweden, the Netherlands, and the United States) non-punishers were rated more favorably than an individual punisher, whereas in some collectivistic countries (Pakistan, Russia, and UAE) they were rated equally. Japan revealed an intermediate pattern (Study 1), or a pattern like the individualistic countries (Study 2). Our studies constitute a first step in mapping how meta-norms vary around the globe.

Keywords. norm enforcement; peer punishment; individualism-collectivism; cross-cultural research; meta-norms
INTRODUCTION

Consider the generic situation of someone informally sanctioning another for behaving selfishly. In the research literature, this is referred to by terms such as peer punishment, social control, and norm enforcement. Informal sanctioning has been given particular attention in the literature on social dilemmas, as research demonstrates that the problem of sustaining individual contributions to a public good can be solved if group members punish free-riders (e.g., Yamagishi, 1986). Thus, punishment of free-riders may be seen 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 (Tedeschi et al., 1974, 1977; Felson, 1981), the same act of aggression may be viewed as legitimate or illegitimate depending on both context and audience. Felson (1981) noted that deviation from a norm may legitimize punishment, but also that such punishment may conflict with rules of politeness that preclude certain kinds of admonishment. In other words, social norms may also regulate how norm violations can be dealt with. Such norms are sometimes referred to as meta-norms (Axelrod, 1986). The question we seek to answer here is whether meta-norms depend partly on cultural factors.

Building on the interactionist approach, Bond, Wan, Leung, and Giacalone (1985) suggested that there should be systematic cultural variation in which aggressive acts fall outside legitimate modes of social control. Specifically, they predicted that this should vary with power distance and individualism-collectivism, two general dimensions that are central to understanding differences in culture (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 managers and assistant managers of two
departments. As tension rose in this meeting, either one of the managers insulted one of the assistant managers or vice versa. Participants rated the legitimacy of the insult and the personality of the insulter. An interaction was found between the status of the insulter (manager or assistant) and whether the target was within the same department (in-group or outgroup) such that the legitimacy of an insult was highest when it came from a manager to his own assistant. Moreover, this interaction was moderated by culture such that insults from a manager to his own assistant were particularly legitimate among Hong Kong Chinese participants.

The finding that only superiors could legitimately use punishment, and only against their own assistant, is consistent with recent research by Eriksson, Strimling, and Ehn (2013, Study 2). They conducted an international on-line survey with a broader set of scenarios, all of which involved several agents of varying roles. Across countries there was a stable pattern that only one role in each scenario was perceived as having the right to punish the norm-breaker. This cross-cultural consistency suggests that meta-norms may have some underlying basic structure that is culturally universal. At the same time, the finding that insults from a manager to his own assistant were particularly legitimate among Chinese participants (Bond of peer punishment. Previous research (e.g., Eriksson et al., 2016) has demonstrated good internal consistency of a scale that employs both judgments of the appropriateness of punishment behavior and approval ratings of the punisher as a person. Thus, we consider these judgments of punishment itself and judgments of the person who punishes as tapping into the same meta-norm construct.

**Is Peer Punishment Universally Judged as More Appropriate when Group Managed?**

Strimling and Eriksson (2014) conducted a series of online surveys on judgments of informal punishment of selfish behaviors. Participants came from the US and India. Several key findings emerged. First, punishment was most condoned when it was minimally severe
(i.e., when the punisher only explained that the selfish behavior was wrong). Second, when compared to a non-punisher, a peer who yelled at someone who behaved selfishly was often judged as an angrier and less trustworthy person whom one would be less willing to spend time with. Third, punishment of a selfish behavior was more condoned when the punishment was administered by a group rather than an individual.

All these findings held in samples from both the US and India, suggesting that there might be human universals in norms regulating punishment. Specifically, three properties of a universal structure of meta-norms were suggested: (1) when a superior is present, it is not appropriate for anyone else to punish a norm-breaker; (2) when punishment occurs in a situation involving peers but no superiors, it is more appropriate the less severe the punishment is; (3) group managed peer punishment is more appropriate than individually managed peer punishment.

The latter two properties were further examined using computer animations instead of vignettes (Eriksson, Andersson, & Strimling, 2016). The animations showed one agent punishing a peer for over-harvesting a common resource. As expected, the punisher was more disapproved of when acting alone, and more disapproved of when using more severe punishment. These results held in samples from both the United States and Sweden, and for two different types of punishment (aggressive confrontations and informal fines).

The abovementioned studies provide some evidence in support of a basic universal structure of meta-norms. However, an obvious weakness is that the cross-cultural comparisons have involved only pairs of countries. In the present paper, we address this limitation by collecting data in a culturally diverse set of eight countries (see details below).

The first research question we address is whether peer punishment is universally judged as more appropriate when group managed than when individually managed, as hypothesized by Strimling and Eriksson (2014). This hypothesis is based on the assumption that group
involvement in an act is universally seen as an indication that the group endorses the appropriateness of the act. Thereby the group provides observers with an external standard to guide their own judgments (Eriksson, Strimling, & Coultas, 2015; Kelley, 1971).

**How Appropriate is Peer Punishment Compared to Not Punishing at All?**

Our second research question is how judgments of peer punishers compare to judgments of non-punishers. In other words, does a peer improve or worsen his/her reputation by punishing a norm-breaker instead of refraining from doing so? Social reactions to peer punishers have been examined in a number of previous studies. In studies on adults from modern western societies, reactions have typically been more negative to peer punishers than to non-punishers (Cinyabuguma et al., 2006; Kiyonari & Barclay, 2008; Eriksson et al., 2016). An impact of culture was indicated in a study with online samples from the USA and India, in which vignettes presented someone behaving selfishly and two peers reacting in different ways: one peer said nothing whereas the other peer yelled at the selfish person. Respondents compared the non-punisher and the punisher on a number of traits (e.g., trustworthiness, anger, etc.). Both samples judged punishers more negatively than non-punishers, but the effect was stronger in the US than in India (Strimling & Eriksson, 2014).

We shall now discuss some factors that could lead to cultural variation in how appropriate peer punishment is judged to be. For instance, societal threat (e.g., food deprivation, natural disasters, territorial threats, and pathogens) is a factor that has been both empirically and theoretically linked to the development of stronger rules and punishments (Gelfand et al., 2011; Roos et al., 2015). For this reason, we would expect peer punishment to be more condoned in countries with higher levels of threat. Another factor that could play a role is power distance, the degree to which inequality in power is accepted (Hofstede, 1980). The use of punishment by the powerful was recently found to be greater among experiment participants from China and India (countries that score high on power distance) than among
participants from the United States (Kuwabara et al., 2016). Even though we here consider punishment between peers it is possible that those who punish are seen as inherently more powerful, in which case we would expect peer punishment to be more condoned in cultures that score higher on power distance. There may also be cultural variation in the overall legitimacy of the specific form of punishment that is used. For instance, how appropriate it is to angrily confront a norm-breaker may partly depend on the general legitimacy of aggression, whereas the appropriateness of an informal fine may partly depend on the general level of respect for individual property rights. In sum, a multitude of factors might cause cultural variation in how appropriate it is to punish a peer for violating an informal norm (compared to not punishing). Acknowledging this complexity, our argument will focus on the role that individualism-collectivism may play.

According to Brewer and Chen (2007), individualism refers to the degree to which a society emphasizes individual autonomy over group embeddedness with respect to identity, agency, and obligations. Individualism with respect to obligations implies assigning greater weight to the rights of the individual in relation to the group. In particular, this should imply a stronger belief in the individual’s right to behave as one pleases without being punished for it, even if the behavior breaks an informal norm. All else being equal, we should therefore expect peer punishment to be generally less appropriate in more individualistic cultures. Consistent with this hypothesis, cross-national studies found that peer punishment of norm violators was used less frequently in more individualistic countries (Brauer & Chaurand, 2010), and more commonly in (collectivistic) East Asia than in (individualistic) North America (Wang & Leung, 2010).

It should be noted that Wang and Leung (2010) proposed another pathway whereby individualism would affect the use of peer punishment: As people in individualistic cultures are less strongly embedded in groups, they have greater relational mobility. With greater ease
of exiting relations comes a decrease in the need to punish others, because people can choose to avoid others instead. This theorizing has been validated in a computational model (Roos, Gelfand, Nau, & Carr, 2014). Given that the frequency of use of peer punishment affects how appropriate it is judged to be (e.g., Eriksson et al., 2015), individualism could lower the appropriateness of peer punishment through this pathway.

**METHOD**

Above we described two hypotheses: (1) peer punishers are judged more negatively when they act on their own rather than as part of a group, and (2) peer punishers are judged more negatively than non-punishers, especially in individualistic countries. To examine these hypotheses we used geometric animations taken from Eriksson et al. (2016), which were inspired by the classic animation study of Heider and Simmel (1944). Because animations do not rely on verbal descriptions but on direct viewing, they may be more salient than vignettes and less subject to “reading between the lines”. Most importantly, the animation paradigm is suitable for cross-cultural studies because it does not rely strongly on language. Findings are therefore unlikely to be explained by connotative differences across various cultures with different languages. Another advantage is that the task is relatively simple, and does not rely on extensive instructions.

We collected data in eight countries in two waves, using different sets of animations and different samples. All animations showed a group of individuals who took turns to harvest a common resource, and one individual who broke the sharing norm by harvesting too much. Animations differed in the kind of sanction the norm-breaker received: an aggressive confrontation (Study 1) or an informal fine (Study 2).

**Selection of Countries**

The study 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 include also the lower range we further invited researchers in Russia (39), United Arab Emirates (25), China (20), and Pakistan (14).

**Participants**

Participants were recruited from student populations at universities in the selected countries. In the UAE, the majority of participants came from other countries, mainly from countries across the Middle East and North Africa. In the other samples, only a very small minority of participants had a nationality different from the sampling country. Below we report results based on the entire samples. Exclusion of participants whose nationality differ from the sampling country do not qualitatively change any results, neither in the cross-national analyses nor for the UAE in particular.

We excluded 28 and 26 participants in Studies 1 and 2, respectively, who failed the manipulation check (see below). Analyses were performed on the data from the remaining 1,420 participants in Study 1 (37% male; mean age 21.1 years, SD = 4.50, age range = 16 to 63) and 1,183 participants in Study 2 (34% male; mean age 21.8 years, SD = 5.27, age range = 17 to 59). Sample sizes and sample characteristics by country are reported in Table 1.

________________________________________________________________________

**INSERT TABLE 1 ABOUT HERE**

________________________________________________________________________

**Procedure**

As an earlier study of animations found no difference in results between online and lab conditions (Eriksson et al., 2016), the recruited students took the study either online or under more controlled conditions at their universities. They were instructed that they were going to
watch two animations of triangles and reported how they felt about the behavior of different triangles. The animations can be accessed at https://pontusstrimling.com/animations-for-cross-cultural-study/.

INSERT FIGURE 1 ABOUT HERE

Animations were displayed on a computer screen. They showed a white stage on which 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) positioned in the four corners. The triangles took turns at harvesting the circles by moving one circle to their own corner, as illustrated in 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.

In the animation of group managed punishment by aggressive confrontation (CWP), Green then moved to the center. Finding no circles there, Green went to gather Pink and Blue from their corners to convene. Subsequently, they moved in synchrony to face the norm violator in its corner. Blue, who was then in the center, delivered the sanction with Green and Pink as onlookers. The sanction consisted of Blue making a quick move toward Purple who lurched backward. Thus, the animation represented an aggressive confrontation that was intimidating without involving physical contact. The triangles all returned to their respective corners and the animation ended.

In the animation of individually managed punishment by aggressive confrontation (IWP), Green instead 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 sanction as in the previous animation.
The animation of group managed punishment by informal fine (CSEn8co), was similar to CWP except that the meeting ended with all three triangles entering Purple’s corner and together they removed all the circles that were present, bringing them back to the center.

In the animation of individually managed punishment by informal fine (ISEn8c), Green never left its corner before Blue found the circles gone and went alone to levy the same informal fine as in the previous animation.

In each study, the order of the two animations (i.e., group managed punishment and individually managed punishment) was counterbalanced. After each animation, participants responded to a series of questions as described below. In Pakistan, the United Arab Emirates, and the United States, the questions were given in English. In the other countries, questions were translated from English to the dominant language (and back-translated by a separate translator, following standard practice to identify and resolve translation problems).

**Measures**

Following Eriksson et al. (2016), participants in Study 1 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.” 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. As there were two animations, this gives a total of six rating measures. Our main interest lies in the Punisher and Others ratings, all of which
had good internal consistency ($\alpha > .8$). Possibly due to floor effects, the Deviant ratings had weaker internal consistency ($\alpha > .6$).

In order to clarify that we wanted participants to judge the animations as if the situation occurred in their own community, we rephrased the three items in Study 2, starting 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)”. The internal consistency of The Punisher ratings, Others ratings, and Deviant ratings all improved their internal consistency by about 0.05 compared to Study 1.

As a manipulation check, participants were asked to describe in free-text format what they found to be the difference between the two animations. Participants were excluded if they stated that they did not find any difference at all or if they made false claims about the difference between the animations (e.g., claiming that the difference between the animations lay in how many circles each triangle took or in the degree of punishment that was deployed). As a secondary manipulation check, participants were asked whether they found the triangles to look alive and goal-directed/intentional; around 85% agreed that they did. This manipulation check was not used for exclusion of participants.

No other outcome measures were collected. The study ended with demographic questions 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.

**RESULTS**

Descriptive statistics of the rating measures for each study are presented in Table 2 (Study 1) and Table 3 (Study 2). Both studies collected rating measures of three targets
(Punisher, Deviant, and Others) in each of two animations (individually managed punishment and group managed punishment) in eight countries (USA, Netherlands, Sweden, Japan, Russia, UAE, China, and Pakistan). Our research questions relate to two particular contrasts: how Punisher ratings compared between individually managed and group managed punishment, and how the individual Punisher rating compared to the 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.

Ratings in Study 1 (Agitated Confrontation)

A 3 × 2 × 8 (Target [punisher, deviant, others] × Animation [group managed punishment, individually managed punishment] × 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 Others (a point we shall return to in a focused contrast). There was a weak main effect of animation, \(F(1, 1402) = 20.39, p < .001, \eta_p^2 = .01\), and a somewhat larger interaction between target and animation, \(F(2, 2804) = 29.60, p < .001, \eta_p^2 = .02\), reflecting higher Punisher ratings in the animation where punishment was group managed rather than individually managed. We shall return to this point in a focused contrast.

There was a medium main effect of country, \(F(7, 1402) = 11.93, p < .001, \eta_p^2 = .06\), and significant Country × Target interaction, \(F(14, 2804) = 8.28, p < .001, \eta_p^2 = .04\), and Country × Animation interaction, \(F(7, 1402) = 5.64, p < .001, \eta_p^2 = .03\), as well as a three-way
interaction, $F(14, 2804) = 6.46, p < .001, \eta_p^2 = .03$. In particular, in the animation with individually managed punishment, the Punisher ratings were especially low and the Others ratings especially high in Sweden, the Netherlands, and the United States. We return to this finding in the second contrast analysis below.

**Punisher Ratings for Individually vs. Group Managed Punishment (Study 1)**

Our first research question was whether there is a universal preference for group managed punishment over individually managed punishment. We performed a $2 \times 8$ (Animation [group managed punishment, individually managed punishment] $\times$ Country [USA, Netherlands, Sweden, Japan, Russia, UAE, China, Pakistan]) mixed ANOVA on Punisher ratings. To begin with, this analysis revealed a medium-sized main effect of country, $F(7, 1408) = 11.74, p < .001, \eta_p^2 = .06$, reflecting country variation in the absolute levels of Punisher ratings. Importantly, there was a medium main effect of animation, $F(1, 1408) = 77.46, p < .001, \eta_p^2 = .05$, which was only weakly moderated by country, $F(7, 1408) = 2.76, p < .001, \eta_p^2 = .01$. To illustrate this result, the left panel of Figure 2 shows 95% CIs (confidence intervals) of the difference in Punisher ratings between group managed punishment and individually managed punishment in each country. Note that the mean difference was positive in every country, although only marginally significant in some countries.

---

**Punisher Rating vs. Non-Punisher Rating (Study 1)**

Our second research question was how appropriate peer punishment is compared to non-punishment across countries. To answer this question, we analyzed the Punisher rating and the Others rating in the animation involving individually managed punishment. We
performed a 2 × 8 (Target [punisher, others] × Country [USA, Netherlands, Sweden, Japan, Russia, UAE, China, Pakistan]) mixed ANOVA on ratings. This analysis revealed a small-to-medium main effect of country, $F(7, 1408) = 7.49, p < .001, \eta^2_p = .04$, reflecting country variation in the absolute levels of Punisher and Others ratings. Our focus here is the effect of target. The ANOVA revealed a medium effect of target, $F(1, 1408) = 118.98, p < .001, \eta^2_p = .08$, which was substantially moderated by country, $F(7, 1408) = 14.05, p < .001, \eta^2_p = .07$. To illustrate this result, the left panel of Figure 3 shows 95% CIs of the difference between Others rating and Punisher rating in the animation with individually managed 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 mid-level.

**INSERT FIGURE 3 ABOUT HERE**

**Ratings in Study 2 (Informal Fine)**

We now turn to the data of Study 2 on animations involving an informal fine. A 3 × 2 × 8 (Target [deviant, punisher, others] × Animation [group managed punishment, individually managed punishment] × 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, $F(2, 2350) = 2666.69, p < .001, \eta^2_p = .69$, 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, $F(1, 1175) = 92.45, p < .001, \eta^2_p = .07$, as well as a medium Target × Animation interaction, $F(2, 2350) = 90.12, p < .001, \eta^2_p = .07$. As in Study 1, this reflected
higher Punisher ratings in the animation involving group managed, compared to individually managed, punishment.

Moreover, there was a medium main effect of country, $F(7, 1175) = 10.36, p < .001, \eta_p^2 = .06$, and significant Country $\times$ Target interaction, $F(14, 2350) = 4.40, p < .001, \eta_p^2 = .03$, and Country $\times$ Animation interaction, $F(7, 1175) = 4.12, p < .001, \eta_p^2 = .02$, as well as a significant three-way interaction, $F(14, 2350) = 2.88, p < .001, \eta_p^2 = .02$. These findings reflected that for individually managed punishment, countries with relatively low Other ratings (Russia, China, UAE, Pakistan) tended to have relatively high Punisher ratings.

---

**Punisher Ratings for Individually vs. Group Managed Punishment (Study 2)**

Analogous to the analysis of Study 1, we performed a $2 \times 8$ (Animation [group managed punishment, individually managed 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, $F(7, 1175) = 9.40, p < .001, \eta_p^2 = .05$. Our focus here is the effect of animation. The ANOVA revealed a large main effect of animation, $F(1, 1175) = 192.81, p < .001, \eta_p^2 = .14$, which was not significantly moderated by country, $F(7, 1175) = 0.56, p = .83, \eta_p^2 = .00$. This is illustrated in the right panel of Figure 2, which also shows that all samples rated informal fine sanctioning significantly higher when group managed than when individually managed.

Taken together, Studies 1 and 2 substantiate the hypothesis of a culturally universal preference for group involvement over individual administered peer punishment.
Punisher Rating vs. Non-Punisher Rating (Study 2)

Finally, we analyzed how appropriate peer punishment was compared to non-punishment across countries in Study 2. Similar to Study 1, we performed a $2 \times 8$ (Target [punisher, others] × Country [USA, Netherlands, Sweden, Japan, Russia, UAE, China, Pakistan]) mixed ANOVA on ratings in the animation involving an individually managed informal fine. There was a small main effect of country, $F(7, 1175) = 4.40, p < .001, \eta^2_p = .03$, reflecting country variation in the absolute levels of Punisher and Others ratings. We found a medium-to-large effect of target, $F(1, 1175) = 141.67, p < .001, \eta^2_p = .11$, which was moderated by country, $F(7, 1408) = 4.96, p < .001, \eta^2_p = .03$. The right panel of Figure 3 illustrates the favorable ratings of non-punishers over punishers in all countries. While the difference was very clear in the Netherlands, Sweden, and Japan, it was not significant in Pakistan and the United Arab Emirates, with the United States and China in-between.

The two panels of Figure 3 showed both consistent and different patterns. A consistent pattern across studies was that Pakistan, UAE, and Russia exhibited a particularly weak or non-existent preference for non-punishers. However, when punishment constituted an informal fine (Study 2), rather than aggressive confrontation (Study 1), China and (in particular) Japan resembled the Western countries in exhibiting a clear preference for non-punishers.

DISCUSSION

Cultural scholars have proposed a number of dimensions along which cultures exhibit meaningful variation. Some of these dimensions are based on values held by individuals of different cultures. In particular, the individualism-collectivism dimension refers to how individual autonomy is valued in relation to group embeddedness. By contrast, the tightness-looseness dimension is based on external constraints, namely the strength of social norms and
the degree of sanctioning within societies (Gelfand et al., 2006). In the present work, we have bridged these two types of dimensions by measuring how individuals in different countries value norm enforcers. Building on prior work that compared pairs of countries (Bond et al., 1985; Strimling & Eriksson, 2014), our research seems to be the first attempt to systematically investigate cultural universals and cultural variation with respect to judgments of norm enforcement between peers. In particular, it is distinct from another interesting line of research on cultural variation in the effectiveness of peer punishment at promoting cooperation in economic experiments (Herrmann, Thöni, & Gächter, 2008; for a meta-analysis, see Balliet & Van Lange, 2013). It is interesting to note that the individualistic countries, such as the USA, Sweden, and The Netherlands, are countries where the punisher is not evaluated positively, given that economic experiments indicate that peer punishment is especially effective in promoting cooperation in these countries. One possible interpretation is that a punishment system, once in place, is effective because of its mere presence, rather than its strong use. The “mere threat” of the peer punishment already provides a strong incentive to cooperation (Van Lange, Rockenbach, & Yamagishi, 2013, p.10).

Our first finding was a culturally universal tendency to judge peer punishment as more appropriate when managed by a group rather than an individual. This effect was found both when the group designated one member to sanction the norm-breaker while the others watched (Study 1) and when the whole group was actively involved in the sanction (Study 2). Although our data do not warrant definite causal conclusions on the mechanism responsible for the observed effect, the results are in agreement with our hypotheses and consistent with the provided account that the group provides observers with an external standard to guide their judgments. We note that recent research on use of peer punishment has indicated 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). Future research is needed to better understand the mechanisms that may drive our results.

The second finding concerned systematic differences in punisher approval across countries. Economic experiments on peer punishment in social dilemmas have yielded a paradox. Although peer punishment can effectively promote cooperation (for a meta-analysis, see Balliet, Mulder, & Van Lange, 2011), those who punish often receive very little approval, sometimes even disapproval (Cinyabuguma et al., 2006; Kiyonari & Barclay, 2008; Strimling & Eriksson, 2014; Eriksson et al., 2016). In line with these latter studies, we found that use of punishment negatively affected approval ratings. In other words, our participants tended to rate non-punishers more favorably than punishers. A more negative view of peer punishers may lead to a lower behavioral tendency to engage in peer punishment. However, it is not clear that this should have overall negative consequences for the group. Participants who use peer punishment in economic experiments seem to do so with little regard for whether it benefits the group (Eriksson, Cownden, Ehn & Strimling, 2014). Whereas peer punishment may encourage cooperation it also engenders social costs and may lead to escalating conflicts within the group (Nikiforakis & Normann, 2008; Nikiforakis & Engelmann, 2011). Restrictions on peer punishment may harness its positive effects while lowering the social costs associated with it (Eriksson et al., 2013; O’Gorman, Henrich, & Van Vugt, 2009; for a demonstration of the effectiveness of non-punitive alternatives, see Wu, Balliet & Van Lange, 2016). Outside the laboratory it remains challenging to reliably measure the costs and benefits associated with peer punishment (Guala, 2012).

In our studies we found the negative view of peer punishers to be moderated by culture: students in some collectivistic non-Western countries tended to have a more neutral view of punishers. This pattern was particularly clear when sanctions occurred in the form of aggressive confrontation (Study 1). When sanctions occurred in the form of an informal fine,
a more negative view of punishers than of non-punishers prevailed also in some collectivist countries, in particular Japan and China. Thus, our data suggest that cultural variation in meta-norms about punishment may also depend on the specific forms of informal sanctions. Perhaps Japanese and Chinese have a particularly strong sense of individual property rights—a promising avenue for future research.

In the introduction we outlined two mechanisms by which peer punishment may be evaluated more negatively in more individualistic cultures. With the exception of Japan in the case of informal fines, our data were consistent with this theoretical prediction. Thus, we tentatively conclude that individualism-collectivism may be a contributing factor to the cultural variation we observed. However, a sample of eight countries is far from sufficient to tease apart all cultural dimensions of potential relevance. For instance, the Western countries in our sample are not only individualistic but also educated, industrialized, rich, and democratic (Henrich et al., 2010). Of the non-Western countries, only Japan shares all of these characteristics (Japan stood out a bit from the Western countries in the direction toward the less individualistic countries in Study 1, but did not show such tendency in Study 2). Thus, it is premature to draw firm conclusions about what drives cultural variation in meta-norms until a substantially larger number of countries have been examined.

Another limitation of the current study is that we used student samples. This is likely to lead to underestimation of cultural differences (Henrich et al., 2010). Thus, representative samples would likely exhibit greater cultural differences in meta-norms than those demonstrated here.

CONCLUSION

Two studies reveal consistent yet novel evidence for the claim that meta-norms exhibit culturally universal features as well as systematic cultural differences. The latter finding may be of particular importance in a cultural evolutionary perspective. Cultures are not static, and
studies of single countries indicate an ongoing move toward a greater emphasis on individualism in Japan (Matsumoto et al., 1996), as well as in the United States and the United Kingdom (Greenfield, 2013; Twenge et al., 2013). Indeed, Inglehart’s (1997) theory of a large-scale change in values toward autonomy and self-expression can be seen as describing a general pattern of change toward greater individualism. Based on our findings, we might also expect a corresponding trend in meta-norms toward decreasing approval of peer punishment. Speculatively, this may lead to systems where punishment is detached from individual punishment, such as the nurturing of institutionalizing formal punishment, while in addition to punishment other forms of informal regulation—gossip, monitoring—are still embraced as well.

In conclusion, we want to emphasize how our study breaks new ground. Starting with Axelrod (1986), a large body of literature—drawing on game theory, economic experiments, and evolutionary arguments—has been devoted to what meta-norms in social dilemmas should look like (e.g., Binmore, 1998; Fehr & Gächter, 2000; Gintis, 2009; Henrich & Boyd, 2001). The present research complements and extends these approaches by closely examining how punishers are evaluated, and how different forms of punishment are evaluated. These questions should help us understand how much, and when, cultures rely on punishment, and whether they will complement punishment with other forms of regulation. In the final analysis, this helps us understand the workings of democratic institutes, and how general trust in societies may be promoted and sustained. New theoretical and empirical developments are necessary to understand what meta-norms actually look like around the globe, and why. Our study is a first step in this direction.
REFERENCES


Table 1

*Sample Sizes and Sample Characteristics*

<table>
<thead>
<tr>
<th>Country</th>
<th>Study 1</th>
<th></th>
<th>Study 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Male %</td>
<td><em>M</em>&lt;sub&gt;age&lt;/sub&gt;</td>
<td>N</td>
</tr>
<tr>
<td>USA</td>
<td>163</td>
<td>33%</td>
<td>18.8</td>
<td>131</td>
</tr>
<tr>
<td>Netherlands</td>
<td>120</td>
<td>27%</td>
<td>20.7</td>
<td>121</td>
</tr>
<tr>
<td>Sweden</td>
<td>176</td>
<td>22%</td>
<td>27.9</td>
<td>202</td>
</tr>
<tr>
<td>Japan</td>
<td>178</td>
<td>60%</td>
<td>20.6</td>
<td>128</td>
</tr>
<tr>
<td>Russia</td>
<td>165</td>
<td>23%</td>
<td>20.7</td>
<td>113</td>
</tr>
<tr>
<td>UAE</td>
<td>190</td>
<td>41%</td>
<td>19.9</td>
<td>204</td>
</tr>
<tr>
<td>China</td>
<td>150</td>
<td>13%</td>
<td>19.4</td>
<td>179</td>
</tr>
<tr>
<td>Pakistan</td>
<td>278</td>
<td>57%</td>
<td>20.5</td>
<td>105</td>
</tr>
<tr>
<td>Total</td>
<td>1420</td>
<td>37%</td>
<td>21.1</td>
<td>1183</td>
</tr>
</tbody>
</table>

*Note.* N = Sample size. Male % = the percentage of the sample that disclosed their gender as male. *M*<sub>age</sub> = Mean age.
Table 2

Descriptive Statistics of Ratings in Study 1

<table>
<thead>
<tr>
<th>Country</th>
<th>Individually managed punishment</th>
<th></th>
<th>Group managed punishment</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Punisher</td>
<td>Deviant</td>
<td>Others</td>
<td>Punisher</td>
</tr>
<tr>
<td>USA</td>
<td>3.65</td>
<td>2.12</td>
<td>4.88</td>
<td>4.05</td>
</tr>
<tr>
<td></td>
<td>(1.50)</td>
<td>(1.15)</td>
<td>(1.18)</td>
<td>(1.50)</td>
</tr>
<tr>
<td>Netherlands</td>
<td>3.72</td>
<td>2.27</td>
<td>4.88</td>
<td>3.93</td>
</tr>
<tr>
<td></td>
<td>(1.46)</td>
<td>(1.12)</td>
<td>(0.97)</td>
<td>(1.40)</td>
</tr>
<tr>
<td>Sweden</td>
<td>3.26</td>
<td>2.13</td>
<td>4.69</td>
<td>3.41</td>
</tr>
<tr>
<td></td>
<td>(1.75)</td>
<td>(1.20)</td>
<td>(1.21)</td>
<td>(1.68)</td>
</tr>
<tr>
<td>Japan</td>
<td>4.02</td>
<td>2.42</td>
<td>4.66</td>
<td>4.63</td>
</tr>
<tr>
<td></td>
<td>(1.29)</td>
<td>(1.05)</td>
<td>(0.91)</td>
<td>(1.15)</td>
</tr>
<tr>
<td>Russia</td>
<td>4.01</td>
<td>2.44</td>
<td>3.95</td>
<td>4.18</td>
</tr>
<tr>
<td></td>
<td>(1.39)</td>
<td>(1.25)</td>
<td>(0.99)</td>
<td>(1.42)</td>
</tr>
<tr>
<td>UAE</td>
<td>3.97</td>
<td>2.20</td>
<td>4.14</td>
<td>4.18</td>
</tr>
<tr>
<td></td>
<td>(1.65)</td>
<td>(1.21)</td>
<td>(1.09)</td>
<td>(1.58)</td>
</tr>
<tr>
<td>China</td>
<td>3.85</td>
<td>2.40</td>
<td>3.98</td>
<td>4.38</td>
</tr>
<tr>
<td></td>
<td>(1.44)</td>
<td>(1.33)</td>
<td>(0.97)</td>
<td>(1.45)</td>
</tr>
<tr>
<td>Pakistan</td>
<td>4.24</td>
<td>2.49</td>
<td>4.41</td>
<td>4.56</td>
</tr>
<tr>
<td></td>
<td>(1.69)</td>
<td>(1.45)</td>
<td>(1.31)</td>
<td>(1.61)</td>
</tr>
<tr>
<td>Total</td>
<td>3.87</td>
<td>2.32</td>
<td>4.43</td>
<td>4.20</td>
</tr>
<tr>
<td></td>
<td>(1.57)</td>
<td>(1.25)</td>
<td>(1.16)</td>
<td>(1.54)</td>
</tr>
</tbody>
</table>

*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

<table>
<thead>
<tr>
<th>Country</th>
<th>Individually managed punishment</th>
<th>Group managed punishment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Punisher</td>
<td>Deviant</td>
</tr>
<tr>
<td>USA</td>
<td>4.73</td>
<td>2.27</td>
</tr>
<tr>
<td></td>
<td>(1.52)</td>
<td>(1.23)</td>
</tr>
<tr>
<td>Netherlands</td>
<td>4.58</td>
<td>2.10</td>
</tr>
<tr>
<td></td>
<td>(1.58)</td>
<td>(0.96)</td>
</tr>
<tr>
<td>Sweden</td>
<td>4.21</td>
<td>2.14</td>
</tr>
<tr>
<td></td>
<td>(1.59)</td>
<td>(1.29)</td>
</tr>
<tr>
<td>Japan</td>
<td>4.01</td>
<td>2.17</td>
</tr>
<tr>
<td></td>
<td>(1.50)</td>
<td>(1.13)</td>
</tr>
<tr>
<td>Russia</td>
<td>4.55</td>
<td>2.29</td>
</tr>
<tr>
<td></td>
<td>(1.40)</td>
<td>(1.19)</td>
</tr>
<tr>
<td>UAE</td>
<td>4.85</td>
<td>1.96</td>
</tr>
<tr>
<td></td>
<td>(1.55)</td>
<td>(1.06)</td>
</tr>
<tr>
<td>China</td>
<td>4.35</td>
<td>1.90</td>
</tr>
<tr>
<td></td>
<td>(1.56)</td>
<td>(1.02)</td>
</tr>
<tr>
<td>Pakistan</td>
<td>4.52</td>
<td>2.23</td>
</tr>
<tr>
<td></td>
<td>(1.73)</td>
<td>(1.34)</td>
</tr>
<tr>
<td>Total</td>
<td>4.48</td>
<td>2.11</td>
</tr>
<tr>
<td></td>
<td>(1.57)</td>
<td>(1.16)</td>
</tr>
</tbody>
</table>

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 group managed punishment over individually managed punishment in each sample in Study 1 (left) and Study 2 (right). Error bars represent 95% confidence intervals. The dashed line indicates the zero level, corresponding to the punisher receiving the same ratings in the two animations involving group managed and individually managed punishment.
Figure 3. The preference for non-punishers over the punisher in the individually managed punishment animation in each sample in Study 1 (left) and Study 2 (right). Error bars represent 95% confidence intervals.