World Development 181 (2024) 106686

Contents lists available at ScienceDirect

World Development

journal homepage: www.elsevier.com/locate/worlddev

Regular Research Article

The mixed impacts of peer punishments on common-pool resources: Multi-country experimental evidence

Arild Angelsen [a,b], Julia Naime [a,b,*]

[a] School of Economics and Business, Norwegian University of Life Sciences (NMBU), PO Box 5003, 1432 \AAs, Norway
[b] Center for International Forestry Research (CIFOR), Bogor, Indonesia

ARTICLE INFO

Keywords:
Field experiments
Peer punishment
Common-pool resources
Deforestation

ABSTRACT

The conservation of common-pool resources (CPRs), such as tropical forests, is a key challenge of development and environmental policies. Peer sanctioning of excessive resource use increases the cost of free riding and may be an effective way to ensure sustainable management of CPRs, but it entails individual costs to punishers. This paper examines peer punishment patterns and impacts in a cross-country framed field experiment (FFE) with homogeneous and heterogenous agents. The FFE is conducted with 720 forest users in Brazil, Indonesia, and Peru. We first examine the relationship between the appropriation of the common-pool resource (first order cooperation) and peer punishment choices (second order cooperation), distinguishing between prosocial and antisocial punishment. A small share (18.2%) of the participants behaves as self-interested payoff maximisers (*homo economicus*), while the largest group (26.1%) cooperates in both the appropriation and punishment decisions (*homo reciprocans*). Across countries, receiving prosocial punishment, defined as punishment of free riders, increases cooperation, while receiving antisocial punishment reduces cooperation. There are, however, important inter-country variations. In Indonesia, the marginal costs of non-cooperation are higher than in the Brazilian and Peruvian sites, and agent heterogeneity significantly increases peer punishment frequency. We conjecture that the higher punishment frequency in Indonesia is linked to stronger equality norms and a willingness to enforce them. Although peer punishment boosts cooperation across all our study sites, the research highlights how peer punishment patterns and impacts are linked to the institutional and cultural contexts.

1. Introduction

The conservation of common-pool resources (CPRs), such as tropical forests, is a key issue in both environment and development discussions. CPRs create a collective action problem: individuals have an incentive to overexploit the resource, thereby reducing the collective benefits and average individual payoff. Peer sanctioning of excessive resource use increases the cost of free riding and may be an effective way to ensure sustainable management of CPRs. This is argued in both observational studies in the Ostrom tradition (e.g., Ostrom, Walker, & Gardner, 1992) as well as in the experimental literature in the Fehr and Gächter (2000) tradition. While peer punishment can create a collective benefit by increasing cooperation and thus group benefits, it entails individual costs to punishers. Thus, peer sanctioning of free riders constitutes a *second-order* collective action problem (Fehr & Gächter, 2002; Ostrom, 1998; Rustagi, Engel, & Kosfeld, 2010).

Experimental studies report on several shortcomings of peer punishments. One is the collective action problem itself: since punishment is costly to the punisher it is also subject to free riding (Ozono, Kamijo, & Shimizu, 2017). Further, antisocial (i.e., punishment of cooperators) and retaliation punishments exist (Herrmann, Thöni, & Gächter, 2008) and the overuse of punishments may reduce the net benefits compared to an open-access situation (Cason & Gangadharan, 2015; Naime et al., 2022; Ostrom, Walker, & Gardner, 1992). Peer punishment also hinges on good social norms and monitoring networks. Bad social norms or imperfect monitoring can prevent effective sanctioning and reaching the social optimal outcome (Abbink, Gangadharan, Handfield, & Thrasher, 2017; Ambrus & Greiner, 2019; Shreedhar, Tavoni, & Marchiori, 2020).

Most studies on peer punishment are lab experiments (e.g., Cason & Gangadharan, 2015; Chaudhuri, 2011; Chen, Lian, & Zheng, 2023; Gachter & Herrmann, 2009; Herrmann, Thöni, & Gächter, 2008; Lohse & Waichman, 2020). There is a long-standing debate on to what extent results from such lab experiments generalize to other groups, domains

* Corresponding author at: School of Economics and Business, Norwegian University of Life Sciences (NMBU), PO Box 5003, 1432 \AAs, Norway.
E-mail addresses: arild.angelsen@nmbu.no (A. Angelsen), junaime@nmbu.no (J. Naime).

https://doi.org/10.1016/j.worlddev.2024.106686
Accepted 23 May 2024

and contexts (Levitt & List, 2007). This is particularly important for peer punishment mechanisms as the patterns of punishment and the norms surrounding its acceptability vary across cultures (Eriksson et al., 2017; Henrich et al., 2010; Henrich et al., 2006), a result confirmed in this paper.

While lab experiments offer high internal validity (isolation of the variables of interest), framed field experiments (FFE) takes the lab to the field to increase external validity by framing the cooperation problem to a specific domain, and by changing the participant pool from university students to actual resource users (Harrison & List, 2004). Only a few FFEs have been undertaken to study monetary peer punishment of CPR management. These include Lopez, Murphy, Spraggon, and Stranlund (2013) on mollusc harvesting in a coastal community of Colombia, Vollan, Pröpper, Landmann, and Balafoutas (2019) on tree harvesting in a woodland savannah area of Namibia, and Kaczan, Pfaff, Rodriguez, and Shapiro-Garza (2017) on a collective Payment for Environmental Services (PES) system in Mexico. In Uruguay, de Melo and Piaggio (2015) evaluated the impact of social (non-monetary) punishment among small scale fishers.

This paper represents the first multi-country FFE of peer punishment in a CPR game with both homogenous and heterogenous agents. The experiment was conducted with 720 smallholders in Brazilian, Indonesian, and Peruvian sites. During the FFE, a group of six local forest users faced a social dilemma, framed as a decision about how many plots to convert to agriculture from a common access forest. Conserving the forest gave higher aggregate benefits to the group in the form of a collective PES scheme, but deforestation gave more agricultural income to the participant than the individual loss of PES income.

We first examine the relationship between first order (FO) cooperation (i.e., the conservation of the CPR) and second order (SO) cooperation (i.e., the peer punishment). As a novel contribution, we present a typology of six different types of players, depending on their FO and SO cooperation, which illustrates the diversity of players with only one quarter being cooperators in both first and second order cooperation (*homo reciprocans*).

Second, we carefully examine the impacts of punishment by evaluating whether the expected gains from free riding are effectively reduced as compared to the open access situation, as well as whether receiving peer punishment leads to lower deforestation. We find significantly higher punishment frequency in the site in Indonesia than in Brazil and Peru. While punishment frequency varies, the impact is consistent across countries: prosocial punishment increases cooperation, while antisocial punishment reduces cooperation.

Third, we examine these questions when there is an equal and unequal distribution of endowments, operationalized in the experiment as differences in the capacity to extract from the CPR. There is only a handful of lab experiments that have compared the effect of peer punishment on homogenous and heterogenous groups on a Voluntary Contribution Mechanism (VCM) or CPR game (De Geest & Kingsley, 2021; Kingsley, 2016; Nockur, Pfattheicher, & Keller, 2021; Reuben & Riedl, 2013; Zelmer, 2003), and none of them are conducted across multiple countries. We find that the impact of group heterogeneity is only significant in Indonesia. We conjecture that this is due to stronger norms and expectations of equity and fairness in Indonesia, making participants more likely to punish high deforesters, and more so when exposed to unequal distribution of endowments. Additional studies can help to draw more generalizable conclusions about the impact of inequality on peer punishment.

2. Theoretical background and research questions

2.1. The peer punishment literature

Users of CPRs face a conflict between individual and collective benefits. Uncoordinated and self-maximizing behaviour will lead to over-exploitation and eventual depletion of the resource, resulting in the well-known "tragedy of the commons" (Hardin, 1968). However, the tragedy is not unescapable: self-governed communities can successfully manage the commons (Ostrom, 1990). Understanding the capacity of such groups to govern themselves is important because formal, external institutions are not always feasible.

Assuming purely self-maximizing individuals, there should be no peer punishment because punishing entails an individual cost. However, motivated by social preferences such as fairness, individuals often engage in punishing free riders (Falk, Fehr, & Fischbacher, 2005; Fehr & Fischbacher, 2004; Fehr & Gächter, 2000). First order cooperators may perceive uncooperative behaviour as unfair, and a central motivation for peer punishment is to sanction non-cooperative individuals (Falk, Fehr, & Fischbacher, 2005). The probability of receiving punishment is also higher the greater the positive deviation from average conversion (De Geest & Kingsley, 2021; Kingsley, 2016).

Participants also engage in antisocial punishment, which occurs when cooperating participants are punished (Herrmann, Thöni, & Gächter, 2008). Antisocial punishments can lead to less cooperation by those that have received such punishments (de Melo & Piaggio, 2015). Antisocial punishments are more frequent when there is an opportunity to retaliate (Engelmann & Nikiforakis, 2015; Nikiforakis, 2008), suggesting the prevalence of retaliation and revenge emotions as a motivation for antisocial punishments (Fehr & Gächter, 2002).

While lab experiments show that peer punishment increases cooperation, the effects of peer punishment depend on the cost of the punishment received (Chaudhuri, 2011; Sutter, Haigner, & Kocher, 2010); if sufficiently high, it leads to near full cooperation (Nikiforakis & Normann, 2008). Other factors that enhance the effect of peer punishment include previous communication (Koch, Nikiforakis, & Noussair, 2021; Ostrom, Walker, & Gardner, 1992), common identity (Weng & Carlsson, 2015), and previous trust and experience with cooperation dilemmas (Gelcich, Guzman, Rodríguez-Sickert, Castilla, & Cárdenas, 2013; Pfaff, Rodriguez, & Shapiro-Garza, 2019). Peer sanctioning might lose effectiveness if there are opportunities to retaliate (Engelmann & Nikiforakis, 2015; Nikiforakis, 2008). In turn, antisocial punishment may reduce cooperation (Gächter & Herrmann, 2011; Herrmann, Thöni, & Gächter, 2008), although a field experiment comparable to ours find no such effect (Vollan et al., 2019).

The effect of peer punishment also differs across participant pools. For instance, Gächter and Herrmann (2011) show that peer punishment does not increase cooperation in a Russian subject pool. Multiple cross-cultural lab experiments have highlighted the importance of culture and contexts in shaping game outcomes (Bruhin, Janizzi, & Thöni, 2020; Eriksson et al., 2021; Henrich et al., 2006; Herrmann, Thöni, & Gächter, 2008). The nature and magnitude of cross-cultural differences in framed cooperation problems, conducted with non-student subject pool remains relatively unexplored for peer punishment and heterogenous treatments. We contribute to filling this gap.

2.2. Research questions and hypotheses

Our first research question (RQ1) examines the patterns of punishment: *who is punishing who?* This includes an examination of the relationship between FO cooperation and SO cooperation. Early studies showed a positive relationship between the two forms of cooperation (Falk, Fehr, & Fischbacher, 2005; Ones & Putterman, 2007). Albrecht, Kube, and Traxler (2018) examined the relationship between FO and SO cooperation, finding two main behavioural types: the prosocial type, who engages in FO cooperation as well as prosocial punishment, and the free rider type, who gives less contributions in the FO dilemma while also being non-punishers. While the relationship between FO and SO is not always clear-cut (Weber, Weisel, & Gächter, 2018), in a field setting there is evidence that individuals who contribute to the FO public good in an experimental game also contribute more to the SO public good of sanctioning over-extraction of forest resources (Rustagi, Engel, & Kosfeld, 2010). Based on the available evidence, we put forward the following hypotheses:

H1.1: First order (FO) cooperators are more active punishers than FO free riders.

H1.2: Prosocial punishment dominates: the probability of receiving punishment is higher the larger the positive deviation from average conversion.

H1.3: Antisocial punishment is driven by retaliation behaviour.

The second research question (RQ2) examines punishment effectiveness: *how does punishment affect (i) the incentives to cooperate and (ii) the actual behaviour?* For the first part of the question, we explore to what extent peer punishment changes the marginal incentives to deviate from the social norm, defined as the average group deforestation. For the second part, we evaluate the impact of punishment on cooperation, separately analysing the impact of prosocial and antisocial punishments, and paying attention to country differences. Experimental research indicates strong cultural variation in punishment behaviours and acceptability of punishment (Eriksson et al., 2017; Herrmann, Thöni, & Gächter, 2008). Eriksson et al. (2017) find that in countries with a collectivistic culture, punishers and non-punishers are rated equally, while in more individualistic cultures punishers are considered less favourably. Forests are managed communally by local or indigenous communities in Peru and Indonesia, while at the site in Brazil land is owned individually by colonist farmers.

H2.1: Peer punishment is more frequent in Peru and Indonesia (collectivistic cultures) than in Brazil (individualistic culture).

H2.2: Prosocial punishment reduces future deforestation while antisocial punishment increases it.

Our third research question (RQ3) is*: how does endowment inequality affect the answers to RQ1 and RQ2?* There is limited evidence about the effect of inequality on punishment in CPRs, and the evidence in VCM games is mixed. Some studies find negative effects on cooperation (Kingsley, 2016; Nikiforakis, Noussair, & Wilkening, 2012), others find no effect (Nockur, Pfattheicher, & Keller, 2021; Weng & Carlsson, 2015). The differences between the studies can stem from different cost-punishment ratios, the nature of agent heterogeneity, as well as the type of public good. Nockur et al. (2021) use a cost-punishment ratio of 1:2, while Nikiforakis et al. (2012) and Kingsley (2016) use 1:3. Further, Nikiforakis et al. (2012) considers heterogeneity in returns from the public good, while Kingsley (2016) and Nockur et al. (2021) consider endowment heterogeneity. De Geest and Kingsley (2021) is the only study that has evaluated endowment heterogeneity in a CPR context, finding more sanctioning in equal than in unequal settings. While the evidence is mixed, inequality in endowments can decrease cooperation levels (Naime et al., 2022), and thus we put forward the following hypotheses:

H3.1: Punishment frequency is higher in groups with inequality in endowments.

H3.2: The response to punishment is stronger in unequal than in equal groups.

3. Data and methods

3.1. Sample and study sites

The FFE was implemented in 24 villages equally split between three study sites in Pará (Brazil), Central Kalimantan (Indonesia) and Ucayali (Peru) between October 2019 and January 2020. Five experimental sessions were conducted in each village, summing up to 30 participants per village and 240 per site. The average age of the participants was 44 years, and 52 % of them were men.

At country level, the eight villages share similar socioeconomic and institutional characteristics (Sills et al., 2017). However, there are relevant differences across the countries. In Indonesia and Peru, smallholders have communal institutions to manage forests. Each household controls on average an area of ~ 2 ha for subsistence and/or commercial agriculture. In Brazil, households control on average an area of 44.8 ha of forest and 38.7 ha of agricultural land, mostly pastures. (See supplementary information (SI), Table A1 for summary characteristics of the study sites.)

In Brazil and Peru, land tenure is in most cases considered secure, in the sense that collective (Peruvian site) or individual (Brazilian site) boundaries of properties are legally recognized. In contrast, tenure is considered weaker in the Indonesian site as village and households do not have legal recognition of the land they manage. Land is publicly owned, and forest access is based on customary (*adat*) laws, which give individuals land claim when they have invested on that land, e.g., by clearing forest and planting crops (Sills et al., 2014).

3.2. Experimental set up

3.2.1. Baseline stage

The CPR's social dilemma was framed as a linear public good game with extraction. Six participants shared access to a common forest which provided collective benefits in the form of a group-based PES. In each round, participants chose how many forest plots to convert to agriculture, reducing the forest area and thus the group benefits.

The experiment consisted of four stages with six rounds each. To conserve anonymity, each participant was represented by a letter of the alphabet, only known to the participant and the experimenter. The letter was changed in each stage to minimize spillovers across treatments. No communication between participants was allowed to reduce the risk of losing anonymity during the experiment, as well as to better capture individual motivations for conservation and sanctioning. Thus, we recreated a non-cooperative environment with no capacity to engage in verbal agreements.

In the first stage, we introduced the collective action problem. Let x_{it} be the number of plots of forest that the participant decides to deforest, and x_{-it} the deforestation of other participants. Setting the benefits of deforestation to 1, δ represents the individual earnings from the collective PES, i.e., the marginal per capita return (MPCR) of the public good. With a total stock of forest plots equal to S, the monetary pay-off during the baseline stage for participant I in round t is:

$$\pi_{it} = x_{it} + \delta\left(S - x_{it} - \sum x_{-it}\right); x_i \leq \overline{x_i} \tag{1}$$

The two conditions necessary for creating a social dilemma are that: (i) the return of deforestation of forest land (x_{it}) is higher than the individual return of the collective PES ($\delta < 1$), and (ii) the individual return from deforestation is lower than the group benefits from the collective PES ($\delta n > 1$), with n being the number of resource users. Thus, the parameters must satisfy the condition $\delta < 1 < n\delta$. The Nash Equilibrium is when everyone maximizes deforestation ($\overline{x_i}$), while the social optimum is no deforestation. Eq. (1) implies that the collective benefit (PES) is distributed equally among participants.

In accordance with previous studies, the levels of the parameters were set at $S = 60$, and $\delta = 0.4$ (Chaudhuri, 2011; Ngoma, Hailu, Kabwe, & Angelsen, 2020). The stock of forestland S was reset in every round, to avoid effects due to accumulated forest loss and to maintain the focus on our three research questions. We specified that each plot was equivalent to 0.5 ha. Each plot of agricultural land was worth 10 points, while each plot of forest gave 24 points to the group (4 points to each participant). In all sessions, each participant had a payoff table indicating his/her earnings as a function of his/her and others' decisions. Visual support was also provided to explain the collective action dilemma, using a cardboard with 60 green squares. Each square represented a forest plot, and showed the group payoff of 24 points, and the individual payoff of 4 points. Whenever deforestation took place, yellow paper stickers indicating the individual payoff of 10 points replaced the green squares.

3.2.2. The treatments

The inequality treatment was introduced by modifying the maximum number of forest plots that a participant could convert to agricultural land, with a between-session design: half of the experimental sessions had inequality in the capacity to deforest, and the other half had equality. In the Unequal groups, three randomly chosen low-capacity participants could deforest a maximum of four plots, and three high-capacity participants could deforest up to eight plots. In the Equal groups, all participants had a medium capacity to deforest of six plots. The experiment strictly focused on the effects of inequality in deforestation capacity by keeping the marginal benefits of deforestation constant and equal across participants, and the same aggregate deforestation capacity across Equal and Unequal groups.

After the baseline, we sequentially introduced three different treatments: (i) individual monitoring of public deforestation, (ii) external punishment, and (iii) peer punishment. In this paper, we only use data from the baseline and peer punishment stages (see Naime et al. (2022) for a comparison of all three treatments). Individual monitoring was introduced at the second stage, and we randomized between external and peer punishment in the third and fourth stages. Our design choice implies that the peer punishment treatment operated under perfect monitoring conditions, thus it offers a higher bound of the impact that peer punishment can have on cooperation on linear public goods games (Ambrus & Greiner, 2019). All analyses control for the order of the two types of punishment treatments.

The payoff function during the peer punishment stage was as follows:

$$\pi_{it} = x_{it} + \delta\left(S - x_{it} - \sum x_{-it}\right) - kg_{it} - r_{it}3k \;; \quad x_{it} \leq \overline{x_i} \tag{2}$$

k is the cost of assigning punishment points to the peer, g is the number of punishments given by participant i to peers, and r is the number of punishments received by i from peers at round t. A participant can either fully cooperate (zero deforestation), not cooperate (free ride, maximum deforestation), or partially cooperate (deforest less than maximum). The optimal strategy during this stage depends on the expectations of receiving punishments from other participants. Free riding is advantageous as long as $(1-\delta)x_{it} > r_{it}3k$. Given the parameters, if participants are not punished the highest possible gains of non-cooperation for a low, medium and high deforestation capacities are 24 (= 6 * 4), 36 (= 6 * 6), and 48 (= 6 * 8) respectively.

The optimal strategy for risk neutral participants with low deforestation capacity is to fully cooperate if they expect at least one peer punishment if they deforest, while medium and high-capacity participants should fully cooperate if they expect to be punished by at least two peers if they deforest. If a participant expects the probability of being punished to depend on how much he/she deforested (absolute or relative to the group average), partial cooperation can be observed. Risk averse participants would opt more for more cooperative choices.

The payoff was presented in terms of points, and the exchange rate was set such that the expected average payment would be equivalent to the country's rural daily wage. In Indonesia and Peru, payments were in cash, while in Brazil payments were in-kind, with commonly used commodities, due to security concerns and recent robberies in the area. At the end of the game and before making the payment, we administered a post-experiment questionnaire to ask about punishment motivations in an open-ended question.

Multinomial logit regressions indicate that the distribution of trust, social and demographic characteristics are balanced across the four types of experimental sessions, except for the risk and social preferences, which are included as control in subsequent analyses (SI, Table A2). The experimental design followed the Center for International Forestry Research (CIFOR) Research Ethics Review, and verbal consent was requested before running the experiment.

3.3. Variable definition and data analyses

Two different definitions of prosocial and antisocial punishment are found in the literature (Cinyabuguma, Page, & Putterman, 2006). The first one defines prosocial (antisocial) punishment as punishment given to a participant who deforest above (below) the round average. The second definition considers the punisher's own cooperation level: prosocial (antisocial) punishment occurs when the punished participants have higher (lower) deforestation that the punisher. We opted to use the first definition (group average) to focus on group rather than individual norms, as well as it being the more relevant definition to disentangle the impact on group cooperation from each type of punishment.

We defined first-order (FO) cooperators as the participants who converted below the group average during the baseline stage, while FO free riders are those who deforested above the group average. For second-order (SO) cooperation, we identified three types of punishers: (i) the no-punisher, who are those with zero punishments assigned during the stage, (ii) the prosocial punisher, those for whom at least half of the punishments were prosocial, and (iii) the antisocial punishers, those for whom more than half of the punishments were antisocial. This classification is similar to Albrecht et al. (2018). Acknowledging that retaliation can be a significant driver of punishment decisions, we further defined a retaliation punishment as one where participants punish an individual who punished him/her in the previous round.

We conduct Mann–Whitney U test to compare and test for significant differences in punishment frequency between equal and unequal groups, and across country sites. To analyse the relationships between FO and SO cooperation (RQ1), we use multilevel linear and double censored Tobit models, with random effects at the individual and experimental session levels. Multilevel models allow to take into account the nested nature (individual and experimental session levels) of the observations, and Tobit models allow to fit censored data (Skrondal & Rabe-Hesketh, 2004).

To analyse how punishment opportunities change the incentives to cooperate (RQ2), we examined the determinants of punishment using Poisson multilevel models (Moffatt, 2015). The dependent variable was the punishments received, and the key independent variable was the participant's deviation from the group average, including both the linear and the squared terms to allow for any non-linearities. From the Poisson models we calculated the expected number of punishments received as a function of the deviation from group average. Based on the predicted number of punishments, we calculated the marginal gains from increasing deforestation, which allowed us to estimate the optimal strategy across countries and contexts. Finally, to evaluate the impact of the punishments received on future cooperation, we regressed the change in deforestation from one round to the next, as a function of past punishments received.

As controls, we included in all our models social and risk preferences measured following Fehr, Glätzle-Rützler, and Sutter (2013) and Binswanger (1981), respectively (SI, section B). Preferences were measured before the CPR game. In all models, we also controlled for trust preferences at the group and village level, as well as socioeconomic characteristics such as age and gender obtained from the post-experiment questionnaire. We included village fixed effects, a dummy specifying the round of the stage to accommodate for learning effects, a dummy indicating whether the peer punishment or the external sanction was introduced first, and a dummy indicating the order of the experimental session within the village (from 1 to 5).

4. Results and discussion

To answer our first research question (RQ1) regarding the relationship between FO and SO cooperation, we first present the descriptive statistics and punishment patterns (section 4.1), and then investigate the determinants of giving punishment (section 4.2). To answer our second research question (RQ2) regarding the effectiveness of peer punishment, we examine the determinants of receiving punishments (section 4.3), before evaluating the impacts on deforestation levels (section 4.4). In each result sub-section, we highlight differences between the equal and unequal groups (RQ3) and discuss country differences in section 4.5.

4.1. Punishment patterns

4.1.1. Punishment across groups and sites

An average of 14.4 punishments were given during the six rounds of the peer punishment session (2.4 per round). Among these, 64.4 % were classified as prosocial, while 35.6 % were antisocial punishments (punishment to those with forest conversion below the group average). This is consistent with hypothesis H1.2 that prosocial punishments dominate.

Consistent with our hypotheses (H2.1 and H3.1), we find significant differences in overall punishment levels between equal and unequal groups, and across the three country sites (Fig. 1). Considering all countries, significantly more punishments were given in the Unequal groups (16.9 per session) than in the Equal groups (11.7 per session) ($p = 0.06$, Mann–Whitney *U* test). This difference is driven by the Indonesian sample; the difference between equal and unequal groups is significant for Indonesia ($p = 0.03$), but insignificant for Peru and Brazil (see Tables A4 to A7 in SI). Furthermore, the frequency of punishment in the Indonesian site (21.5 per session) was about twice as high as in the two other study sites (Peru: 10.5; Brazil: 10.9). In the Unequal groups in Indonesia, punishment frequency was more than three times than in the Equal groups in Peru (25.7 vs. 8.0, see Fig. 1).

Across the Equal and Unequal groups and the three country sites, there is some variation in the share of prosocial punishments, from 55.7 % for the Unequal groups in Peru to 71.1 % for the Equal groups in Indonesia. Overall, the share of prosocial punishments was only slightly higher in the Equal sessions (66.5 %) than in the Unequal sessions (62.9 %). The share of prosocial to antisocial punishments of approximately 2:1 is within the range of what has been reported previously (e.g., Gächter & Herrmann, 2011).

4.1.2. A typology of first and second order cooperation

Combining our FO cooperation categorization (deforestation above or below the group average in baseline stage) and SO cooperation categorization (i.e., no punisher, mostly prosocial punisher, or mostly antisocial punisher), we introduce in Table 1 a typology of six different types of participants.

We label the participants who are both FO and SO cooperators *Homo reciprocans* (Bowles & Gintis, 2002), as they are punishing individuals who are not reciprocating on their cooperative behaviour. By deforesting less while engaging in prosocial punishments they make a double contribution to the group's benefits. This group consists of slightly more than a quarter of the participants (26.1 %). Almost as large is the *Benigns* (21.7 %), the FO cooperators who did not want to engage in any punishment of their peers. The behaviour of a small group of participants (6.2 %), the FO cooperators who engaged in antisocial punishments, are labelled the *Confused* as they are not consistent between FO and SO cooperation.

Among the FO free riders, a sizeable group are the *Hypocrites* (17.5 %); they convert more forest than the group average and punish those that do the same, displaying double standards. The *Homo economicus* (18.2 %) behave as selfish payoff maximisers, free riding in both the FO and SO cooperation games. Finally, the *Saboteurs*, making up about one tenth of the participants (10.3 %), are FO free riders that also engage in antisocial punishment (i.e., punishing those that contribute the most in the FO cooperation).

Similar to Albrecht, Kube, and Traxler (2018), we find that the most common types are the Non-punishers (39.9 %) and the Prosocial punishers (43.6 %). Likewise, the share of FO free riders who are Non-punishers is slightly lower (18.2 %) than the share of FO cooperators who are Non-punishers (21.7 %). The Antisocial punishers have the lowest share overall (16.5 %). Contrary to Albrecht et al. (2018), we find a higher proportion of FO free riders who engage in antisocial punishment behaviour (10.3 %) as compared to FO cooperators (6.2 %). Further, the high share of Non-punishers is in contrast with the pattern found in cross-country lab experiments of Herrmann, Thöni, and Gächter (2008), where the share of non-punishers is only 17 % (Bruhin, Janizzi, & Thöni, 2020). The non-anonymous setting of lab experiments and abstract decision frame, as compared to our field setting, is a plausible explanation of our higher share of Non-punishers.

Table 1 further highlights one major difference between the equal and unequal groups, as already observed in Figure 1, namely the much higher proportion of individuals who engage in punishments in the

Fig. 1. Average peer punishments by punishment type, Equal or Unequal groups, and country.

Table 1
Typology and proportion of participants, according to first order (FO) and second order (SO) cooperation.

		Enforcement public good (second order (SO) cooperation)			
		Prosocial punishment (SO cooperators)	**Antisocial punishment**	**No punishment (SO free riders)**	**Total**
Common-pool resource (first order (FO) cooperation)	**FO co-operator (Equal group, Unequal group)**	*Homo reciprocans* 26.1 % (25.6 %, 26.7 %)	*Confused* 6.2 %(4.4 %, 8.1 %)	*Benigns* 21.7 %(23.6 %, 19.7 %)	(53.6 %, 54.5 %)
	FO free riders (Equal group, Unequal group)	*Hypocrites* 17.5 %(15.0 %, 20.0 %)	*Saboteurs* 10.3 %(10.6 %, 10.0 %)	*Homo economicus* 18.2 %(20.8 %, 15.6 %)	(46.4 %, 45,6%)
	Total	43.6 %(40.6 %, 46.7 %)	16.5 %(15.0 %, 18.1 %)	39.9 %(44.4 %, 35.3 %)	100 %

Note: Shares represent average for all groups, while share in parentheses are for equal and unequal groups, respectively.

unequal groups compared to the equal groups: 64.8 % vs. 55.6 %. Interestingly, the difference is not explained by a higher share of *Homo reciprocans* (i.e., the FO and SO cooperators), but rather *Hypocrites* (the FO free riders who are SO cooperators, meaning they punish prosocially) and *Confused* (the FO cooperators who punish antisocially). The higher number of *Confused* and *Hypocrites* in unequal settings indicates the ambiguous way in which inequality affects punishment patterns: it increases both the share of prosocial and antisocial punishers.

We also observe substantial differences across the countries for the six typologies (Fig. 2, see also SI, Table A8). Compared to the two other sites, Indonesia has a higher share of *Hypocrites* (high deforesters punishing fellow high deforesters) and *Homo reciprocans* (low deforesters punishing high deforesters), and a lower share of *Benigns* and *Homo economicus*. Across the two types of groups, we note for the Indonesian site the much lower share of the *Benigns* in the unequal groups, while the shares of both *Homo reciprocans* and *Saboteurs* are higher.

Our results differ from De Geest and Kingsley (2021), who found that introducing agent heterogeneity *reduced* punishment frequency in a CPR game. Possible explanations include the framing and experimental pool (lab with an abstract problem and undergraduate students vs. field experiment framed as a real-life problem and actual forest users) as well as the existence in De Geest and Kingsley (2021) of an "outside option" for the participants. The outside option represents the income obtained from not appropriating the resource and is different from the collective benefit. De Geest and Kingsley (2021) argue that inequality allowed participants to better coordinate on a contribution norm, while in our case inequality might have had the contrary effect (i.e., hindering coordination on the norm), which would explain why punishments are higher in unequal environments.

What are the potential motivations behind these patterns of FO and SO cooperation? The motivation for FO cooperation (*Homo reciprocans*, *Confused* and *Benigns*) are often driven by a concern of others' payoff or to avoid the guilt of being a free rider (e.g., Lopez, Murphy, Spraggon, & Stranlund, 2012). Motivations for SO cooperation (*Homo reciprocans* and *Hypocrites*), include fairness and inequality aversion, as by punishing they reduce the higher than average payoff of the free riders (Falk, Fehr, & Fischbacher, 2005; Katuščák & Miklánek, 2022).

The *Confused* and *Saboteurs*, in turn, who punish the FO cooperators, can be driven by negative emotions such as spite and revenge. These participants might gain utility from reducing other's payoff at a cost to

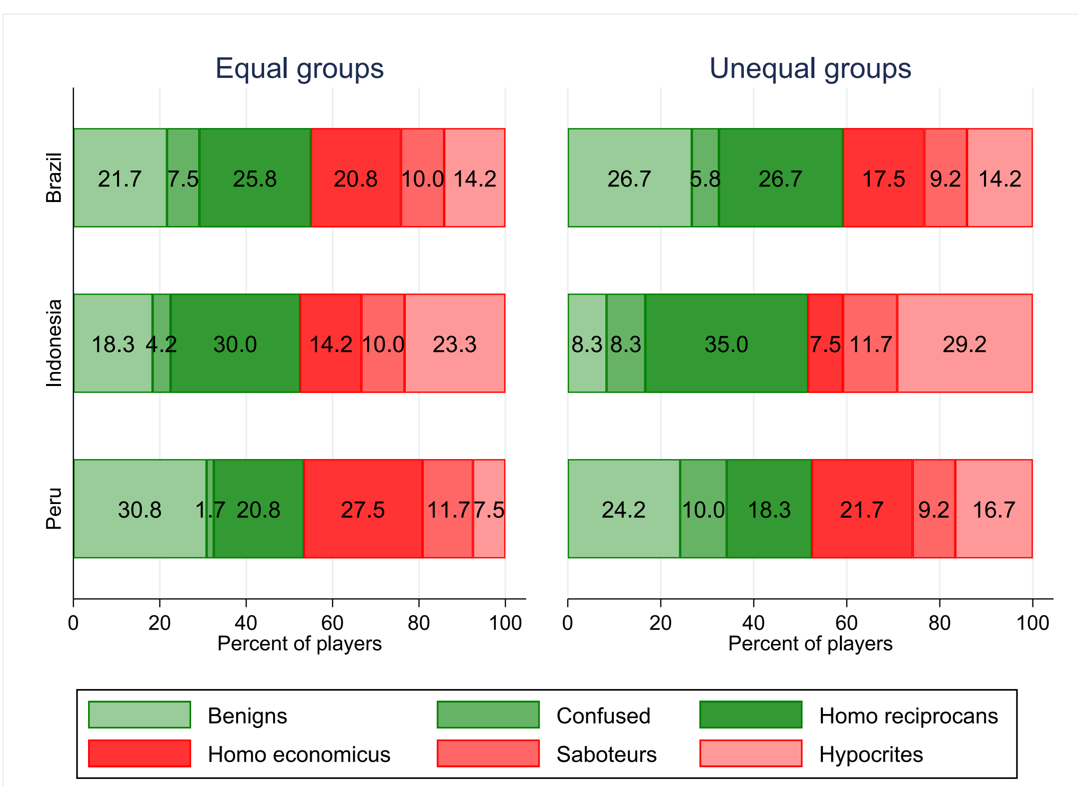

Fig. 2. Types of participants across countries and types of groups.

themselves and the others. *Hypocrites* might have a similar motivation to the *Confused* and *Saboteurs,* and driven by spite – rather than by fairness concerns – gain utility from reducing the payoff of FO free riders. This motivation is consistent with their own FO free-riding behaviour. Another possible motivation for antisocial punishment is to target the non-punishers and avoid earning less than non-punishing subjects (Thöni, 2013). A proportion of the punishments can also be linked to confusion or anchoring (e.g., Katuščák & Miklánek, 2022; Ostrom, Walker, & Gardner, 1992).

4.2. Who are the punishers?

We now move to the analysis of who assigns punishments. Since the motivations and behavioural patterns are likely to differ between prosocial and antisocial punishments, we conduct a separate analysis for the two punishment types. Tables 2 and 3 present the results of Tobit regression models. Our main independent variable of interest is the degree of FO cooperation, defined as a participant's deviation from the group average in the round. We control for lagged punishment received as retaliation can also be a motivation for giving a punishment.

As hypothesized in H1.1, participants that deforested less than the group average (FO cooperators) are more likely to give prosocial and less likely to give antisocial punishments. In contrast, FO free riders are more likely to give antisocial and less likely to give prosocial punishments. Further, high group conversion during the round was associated with more frequent punishments.

Looking into country differences, in the site with the highest punishment frequency, Indonesia, the four coefficients linking FO cooperation to SO cooperation are all significant and in general larger compared to the other countries (Tables 2 and 3, column 3), suggesting that the fairness and retaliatory motivations for punishing are strongest in this site. As robustness check, we ran supplementary regressions with country interactions and find consistent results in the country differences (SI, Table A12). We also regressed punishments based on the cooperation in the baseline stage and find consistent results: FO free riders are more likely to give antisocial punishments and less likely to give prosocial punishments (SI, Tables A13 and A14).

Lagged retaliatory punishments have a small impact on the likelihood of giving prosocial punishments when considering the whole sample, but are insignificant at the country level (Table 2). The picture is different for antisocial punishments, where lagged punishments significantly increase the antisocial punishment in all the country sites (Table 3). This supports hypothesis H1.3, that antisocial punishment is more likely to be driven by retaliation motivations. Our results are consistent with previous analysis establishing the relationship between FO and SO cooperation (Albrecht, Kube, & Traxler, 2018) and retaliatory behaviour and antisocial punishments (Nikiforakis, 2008). The relationship between antisocial punishments and retaliation is stronger than found in other countries (e.g., Vollan et al. (2019) in Namibia). Finally, Wald tests indicate that there is no significant difference between the coefficients of low and high-capacity participants for giving antisocial ($p = 0.49$) or prosocial punishments ($p = 0.47$) for the total sample.

4.3. What are the gains and costs of free riding?

Prosocial punishment make deforestation above the group average more costly. Figure 3 displays how the expected number of punishments received varies by a participant's deviation from the group average, distinguishing between the equal or unequal groups (panel a) and the country sites (panel b). The principal conclusion is that high converters are much more likely to be punished by fellow group members, and that the targeting of free riders varies by site, being more pronounced in Indonesia. There is no significant difference between the expected punishments in the Equal and Unequal groups (Fig. 3).

For Indonesia, four punishments can be expected if the positive deviation from the group norm is four. In contrast, the number of punishments in Brazil and Peru range from 0.27 when there is no deviation from the group norm, to about one when the deviation is four units. Thus, the rate at which the punishment probability increases as a function of deviation from group average in Indonesia is much higher than in Brazil or Peru.

Addressing the first part of RQ2, the expected marginal payoff loss from punishments received is given in Table 4. The numbers indicate the payoff loss from deviating one more unit from the group average. The participants' optimal strategy varies across countries. Considering the full sample, it is optimal to deforest two units above group average. In Brazil and Peru, the optimal deforestation is higher, at 3 units above group average. In Indonesia, the optimal strategy is to deforest just at the average group deforestation.

The result is consistent with lab evidence; individuals' willingness to punish depends on the intensity of the violation (Fehr & Gächter, 2000, 2002; Masclet, Noussair, Tucker, & Villeval, 2003) and is in line with the graduated sanctions criteria of successful collective governance

Table 2
Tobit model of giving prosocial punishment.

Prosocial punishment given	(1)	(2)	(3)	(4)
	Total	Brazilian site	Indonesian site	Peruvian site
FO cooperator	0.27***(0.06)	0.13(0.11)	0.42***(0.07)	0.30***(0.12)
FO free rider	−0.14**(0.06)	−0.13(0.11)	−0.13*(0.08)	−0.19(0.13)
# Lagged punishments received (1 round)	0.11**(0.05)	0.22(0.21)	0.08(0.06)	0.19(0.12)
Deforestation capacity				
Low capacity	0.30(0.19)	−0.37(0.40)	0.50*(0.27)	0.34(0.29)
High capacity	0.36*(0.19)	0.37(0.34)	0.15(0.23)	0.64**(0.32)
Round dummy (1–6)	−0.07***(0.03)	−0.06(0.05)	−0.10**(0.04)	−0.00(0.04)
Random effects	Yes	Yes	Yes	Yes
Village fixed effects	Yes	Yes	Yes	Yes
Individual levelcovariates	Yes	Yes	Yes	Yes
Observations	3600	1200	1200	1200
Log likelihood	−2135.94	−594.97	−929.48	−544.67
χ2	191.62	311.98	351.50	298.62
p-value	0.00	0.00	0.00	0.00

Notes:
1) The variable FO cooperator (FO free rider) indicates how much below (above) the group average the participant's forest conversion was during the round.
2) Random effects are at the individual and experimental session level. Model is censored at 0 and 3. Clustered standard errors by experimental session in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3
Tobit model of giving antisocial punishment.

Antisocial punishment given	(1)	(2)	(3)	(4)
	Total	Brazilian site	Indonesian site	Peruvian site
FO cooperator	−0.33***(0.09)	−0.27**(0.12)	−0.36**(0.14)	−0.47**(0.19)
FO free rider	0.21***(0.06)	0.15(0.11)	0.27***(0.09)	0.12(0.13)
# Lagged punishments received (1 round)	0.39***(0.07)	0.86***(0.13)	0.19**(0.07)	0.64***(0.19)
Deforestation capacity				
Low capacity	0.59***(0.21)	0.15(0.32)	0.50(0.31)	0.71**(0.31)
High capacity	0.45**(0.20)	0.38(0.28)	0.11(0.30)	0.62*(0.32)
Round dummy (1 to 6)	−0.04(0.04)	0.06(0.08)	−0.04(0.06)	−0.11*(0.06)
Random effects	Yes	Yes	Yes	Yes
Village fixed effects	Yes	Yes	Yes	Yes
Individual level covariates	Yes	Yes	Yes	Yes
Observations	3600	1200	1200	1200
Log likelihood	−1405.45	−389.28	−550.29	−427.26
χ2	1784.55	432.99	1801.54	2529.77
p-value	0.00	0.00	0.00	0.00

Notes:
1) The variable FO cooperator (FO free rider) indicates how much below (above) the group average the participant's forest conversion was during the round.
2) Random effects are at the individual and experimental session level. Model is censored at 0 and 3. Clustered standard errors by experimental session in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

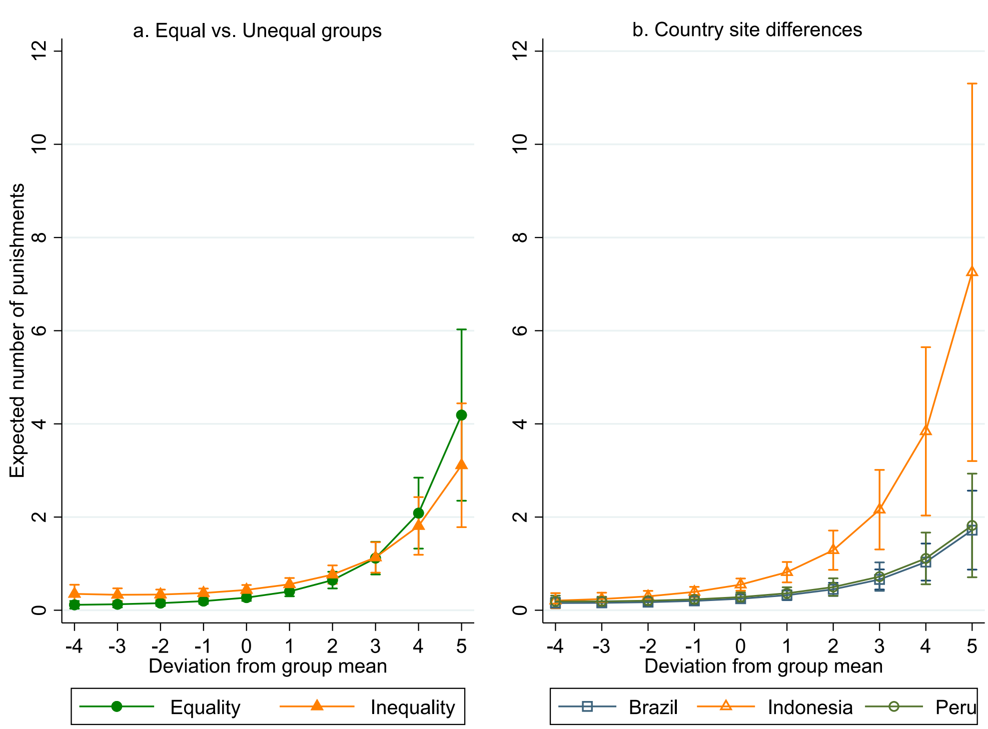

Fig. 3. Expected punishments received depending on the deviation from the group mean, for equal and unequal groups, and by country. *Note 1*: Negative deviations imply antisocial punishments, while punishment of positive deviations implies prosocial punishment. Predictive margins with 95 % confidence intervals. See SI (Table A16) for full model specification and coefficients. *Note 2*: Model predictions indicate that the expected number of punishments in Indonesia exceed the experimental limit of maximum five punishments received per participant.

(Ostrom, 1990; Wilson, Ostrom, & Cox, 2013).

We also evaluated how expected punishment received varies by the number of deforested plots instead of the deviation from group average, finding similar results (SI, Fig. A1): punishments increase sharply with conversion, no effect of inequality, and significant differences between countries. The expected marginal payoff loss considering the number of plots deforested is presented in SI, Table A15.

4.4. *Do punishments make free riders cooperate?*

We have demonstrated how the punishment patterns change the payoffs of different deforestation choices (Table 4). Did the participants act accordingly (second part of RQ2)? Specifically, we ask: how does receiving punishment reduce the forest conversion in the next round? The results in Table 5 are in line with hypothesis H2.2. On average,

Table 4
Marginal loss (in number of points) from punishment, per country and depending on the deviation from the group average.

Deviation from group average	Marginal loss from higher expected punishment (expected increase in punishment * 30 points) from deforesting one more unit			
	Total sample	Brazil	Indonesia	Peru
-4	-3.9	-0.51	1.05	-1.74
-3	0.15	-0.06	1.59	-0.78
-2	0.75	0.33	2.46	-0.033
-1	1.56	0.84	3.87	0.69
0	2.82	1.59	**6.27**	1.62
1	4.98	2.82	10.47	3.09
2	**9.03**	5.07	18.09	5.7
3	17.13	**9.45**	32.4	**10.83**
4	34.35	18.48	60.15	21.69
5	73.26	38.1	115.89	46.35

Note: Grey cells indicates that the marginal effect is significantly different from zero at the 10 % level. Bold numbers indicate the deviation at which it becomes unprofitable to deviate more from the average, considering that one unit of deforestation brings a marginal net benefit of 6 points.

Table 5
Impact of punishment on deforestation levels.

	(1)	(2)	(3)	(4)
Change in deforestation	Total	Brazilian site	Indonesian site	Peruvian site
Deviation from group average	0.66***(0.03)	0.72***(0.04)	0.70***(0.05)	0.46***(0.05)
Lagged punishments received				
Anti-social punishments	0.88***(0.11)	1.13***(0.23)	0.71***(0.14)	0.97***(0.17)
Anti-social punishments^2	−0.31***(0.06)	−0.37***(0.12)	−0.23***(0.08)	−0.36***(0.08)
Pro-social punishments	−1.16***(0.09)	−1.92***(0.23)	−1.03***(0.09)	−0.99***(0.17)
Pro-social punishments^2	0.15***(0.03)	0.41***(0.08)	0.11***(0.03)	0.20***(0.06)
Deforestation Capacity				
Low capacity	0.09*(0.05)	0.02(0.13)	0.16***(0.06)	0.05(0.06)
High capacity	−0.05(0.04)	−0.13(0.09)	−0.01(0.04)	−0.02(0.06)
Constant	−0.11(0.14)	−0.16(0.25)	0.40**(0.17)	−0.37*(0.20)
Random effects	Yes	Yes	Yes	Yes
Village fixed effects	Yes	Yes	Yes	Yes
Individual level covariates	Yes	Yes	Yes	Yes
Observations	3960	1320	1320	1320
Log Likelihood	−6074.10	−2285.49	−1559.72	−1970.52
χ2	1090.67	1223.47	2571.06	719.28
p-value	0.00	0.00	0.00	0.00

Note: Random effects are at the individual and experimental session level. Clustered standard errors by experimental session in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

receiving one prosocial punishment *reduced* deforestation by approximately 1 unit in the next round. In contrast, receiving one antisocial punishment *increased* deforestation. A major finding is thus that – across all country sites – antisocial punishments have a double negative effect on the group: besides being costly to both the punisher and the punished, it also reduces future cooperation and thus the public good (i.e., the collective PES payments).

The detrimental effect of antisocial punishment is characteristic of peer punishment contexts in which the decision to punish is uncoordinated (Boyd, Gintis, & Bowles, 2010; Herrmann, Thöni, & Gächter, 2008) and is also observed in the context of social, non-monetary punishments (de Melo & Piaggio, 2015). The results are consistent with the previous lab experiments showing the negative effects of antisocial punishment on cooperation (e.g., Gächter & Herrmann, 2011), but contrast the more recent evaluation of peer punishment in Namibia, where antisocial punishment does not significantly affect cooperation rates (Vollan et al., 2019).

Figure 4 presents the predicted change in deforestation depending on the number of antisocial and prosocial punishments received in the previous round, by country. The pattern is clear: prosocial punishment reduces deforestation at a decreasing rate, while antisocial punishment increases deforestation (also at a decreasing rate). There are no statistical differences in the effect of prosocial and antisocial punishments across countries. We further find that participants with low deforestation capacity respond more to punishment, but only in Indonesia (which supports hypothesis H3.2.).

One design factor that can increase effectiveness is to introduce punishment coordination among participants (Boyd, Gintis, & Bowles, 2010). Various studies have examined variations of the peer punishment rules, for example, voting on whom to punish (Nockur, Pfattheicher, & Keller, 2021; Pfattheicher, Böhm, & Kesberg, 2018) or introducing communication before punishing (Koch, Nikiforakis, & Noussair, 2021). Similarly, delegating the enforcement and punishment decisions to a small number of individuals (i.e., leaders or monitors) can solve some of the issues of peer punishment identified in the study, but not always (Kosfeld & Rustagi, 2015; Nosenzo & Sefton, 2014): the effect crucially

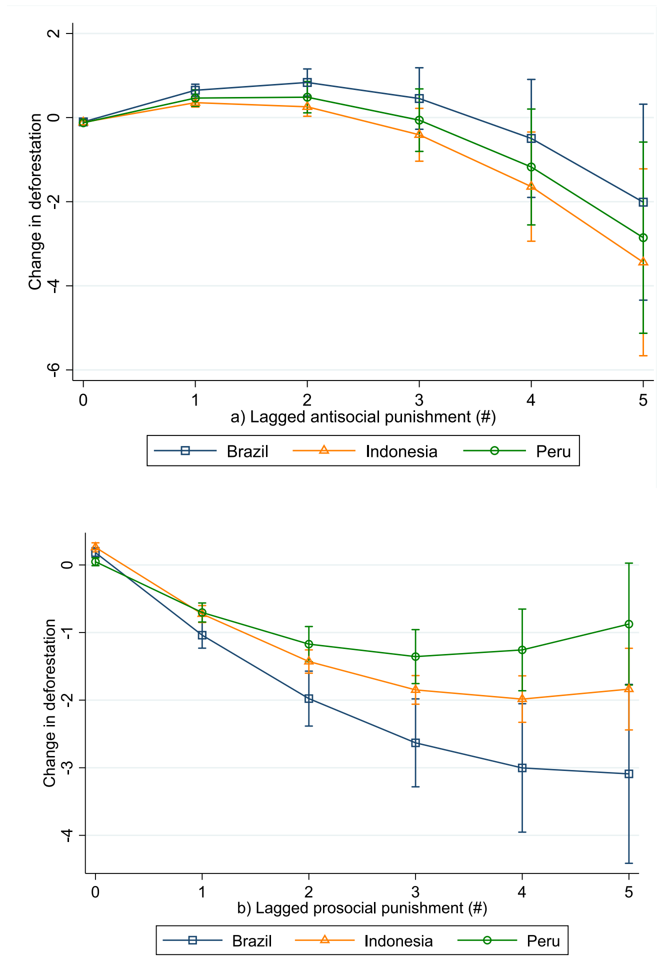

Fig. 4. Predicted change in deforestation depending on the number of antisocial or prosocial punishments received, by country.

depends on the motivations of those leading the punishment.

4.5. Explaining differences across countries

We found strong and significant differences across the three sites. The Indonesian participants punished twice as often as the participants from Brazil and Peru, they punished more in unequal groups (no significant difference for Brazil and Peru), and they punished any forest conversion above group average. How can the much higher punishment frequency in Indonesia be explained?

First, in all our models we control for trust, social and risk preferences, thus any systematic difference across participants in these variables are controlled for in the regression models. Second, other socioeconomic characteristics may affect the outcomes (SI, Table A1). While there are significant differences across socioeconomic variables, these do not provide a consistent explanation of the differences across sites. For example, for key characteristics such as forest and agricultural land sizes, Brazil is the outstanding case while Peru and Indonesia are quite similar with much less land per household. Further, the Brazilian site is the only with individual land rights, yet the punishments levels are similar to the Peruvian site with communal ownership. One notable difference is, however, the much lower deforestation rate in the Indonesian site. This could be one reason why high deforestation in the Indonesian experiment is more frequently sanctioned. Note, however, that the national annual deforestation rate is 2–3 times higher in Indonesia (0.78 %, compared to 0.23 % (Peru) and 0.30 % (Brazil), for the years 2010–2020) (FAO, 2020).

Third, and our main conjecture, the observed differences in punishment patterns may reside in the different cultural norms across the three countries. Although challenged by legal structures, rural communities in Indonesia – including our study site – frequently refer to the customary (*adat*) rules to manage community forests. More generally, some evidence support the existence of stronger norms and preferences for equality and fairness in Indonesia than in the two South American countries. Table 6 presents the responses to two questions in the World Values Surveys: respondents in Indonesia expect the government to prioritize income equality higher as compared to respondents in Brazil and Peru. Furthermore, Indonesians also have higher expectations about being treated fairly, indicating their belief in the existence of a strong fairness norm in the country.

An experiment on distributional rules for PES payments, and with a basic design comparable to ours, also lends support to a claim of stronger egalitarian norms in Indonesia (Cook, Grillos, & Andersson, 2023). The experimental group elected a leader, who then decided on the payment distribution. Elected leaders in Indonesia always chose to distribute the payments in an egalitarian fashion, while in Peru 27 % chose a non-egalitarian distribution (Brazil was not included in the study).

The existence of egalitarian norms may not alone explain the higher punishment frequency. A large cross-country study on perceptions of appropriate responses to norm violations (Eriksson et al., 2017) found a higher acceptance of both physical and verbal confrontation of norm violations in Indonesia than in Brazil and Peru (Table 6). In short, egalitarian norms and a high willingness to deal with norm-violators may explain why the Indonesian participants punished twice as often as the Brazilian and Peruvian participants.

5. Conclusion

We extend the literature on cooperation dilemmas by conducting a framed field experiment (FFE) in Brazil, Indonesia and Peru, and classifying participants depending on how they behave in the first order (FO) and second order (SO) cooperation problems. We found a positive correlation between FO cooperation (conserving the common-pool resource) and SO cooperation (punishment of free riders). Mirroring that, there is also a positive correlation between FO free riding and antisocial punishment. Yet a significant proportion of both FO cooperators and FO free riders (40 %) do not engage in peer punishment. Our typology of six different participant types illustrates the diversity of individual behaviour. Only the *Homo reciprocans* and the *Homo*

Table 6
Differences across countries in norms and responses to norm violations.

Variable	Brazil	Indonesia	Peru	Data source
Equality preferences (1–10)[1]	4.85	6.76	5.73	WVS7[4]
Perceived fairness of others (1–10)[2]	4.99	6.67	5.56	WVS5[4]
Appropriate response to norm violations:				
Physical confrontation (0–5)[3]	1.68	2.46	1.82	Eriksson et al. (2017)
Verbal confrontation (0–5)[3]	2.76	3.04	2.77	

Notes:
[1]Mean response to the item "The state makes people's incomes equal", with the overall question being: "… please tell me for each of the following things how essential you think it is as a characteristic of democracy. Use this scale where 1 means "not at all an essential characteristic of democracy" and 10 means it definitely is "an essential characteristic of democracy")
[2]Mean response to the question: "Do you think most people would try to take advantage of you if they got a chance, or would they try to be fair? … 1 means that "people would try to take advantage of you," and 10 means that "people would try to be fair".
[3]Mean score on a six-point scale, from extremely inappropriate (0) to extremely appropriate (5).
[4]https://www.worldvaluessurvey.org/wvs.jsp.

economicus show consistent behaviour in the two cooperation problems.

Peer punishment can deliver on conservation outcomes and reduce deforestation in the context of collective PES. A large share of participants engages in prosocial punishment, i.e., they punish individuals deforesting above the group average. The punished free riders reduce their deforestation levels in the next round compared to those not being punished. However, self-enforcement in the form of peer punishment entails a risk of engaging in antisocial behaviour which – besides being costly to both the punisher and the punished – has a negative effect on future cooperation. Approximately one third of the punishments were antisocial. Future examination of how patterns of antisocial and prosocial punishment evolve over time can help increase the understanding of peer punishment dynamics.

Important differences in punishment frequencies were observed across the country sites, with Indonesia having the strongest sanctioning of free riders; any positive deviation from group average was penalized such that the optimal individual strategy was to conform with the rest of the group. Relatedly, the effect of inequality in endowments on peer punishment varied: it increased the frequency of punishments in Indonesia, but not in Brazil or Peru. We conjecture that a strong norm of equity and fairness in Indonesia, combined with an acceptance for punishing norm violation, makes punishment of above-average resource exploiters more likely, particularly in an unequal setting. Our findings serve as a warning against sweeping generalizations of experimental results across cultural contexts and population pools.

CRediT authorship contribution statement

Arild Angelsen: Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Julia Naime:** Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This research is part of CIFOR's Global Comparative Study on REDD+ (www.cifor.org/gcs). The funding partners that have supported this research include the Norwegian Agency for Development Cooperation (Norad, Grant No. QZA-16/0110 & No. QZA-21/0124), International Climate Initiative (IKI) of the German Federal Ministry for the Environment, Nature Conservation, and Nuclear Safety (BMU, Grant No. 20_III_108), and the CGIAR Research Program on Forests, Trees and Agroforestry (CRP-FTA) with financial support from the CGIAR Fund Donors. We thank Frode Alfnes and Unai Pascual for valuable comments to an earlier version of our manuscript. All participants in the study sites in Brazil, Indonesia and Peru are gratefully acknowledged.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.worlddev.2024.106686.

References

Abbink, K., Gangadharan, L., Handfield, T., & Thrasher, J. (2017). Peer punishment promotes enforcement of bad social norms. *Nature Communications, 8*(1), 1–8.
Albrecht, F., Kube, S., & Traxler, C. (2018). Cooperation and norm enforcement - The individual-level perspective. *Journal of Public Economics, 165*, 1–16. https://doi.org/10.1016/j.jpubeco.2018.06.010
Ambrus, A., & Greiner, B. (2019). Individual, Dictator, and Democratic punishment in public good games with perfect and imperfect observability. *Journal of Public Economics, 178*. https://doi.org/10.1016/j.jpubeco.2019.104053
Binswanger, H. (1981). Attitudes Toward Risk: Theoretical Implications of an Experiment in Rural India. *The Economics Journal, 91*.
Bowles, S., & Gintis, H. (2002). Homo reciprocans. *Nature, 415*(6868), 125–127.
Boyd, R., Gintis, H., & Bowles, S. (2010). Coordinated Punishment of Defectors Sustains Cooperation and Can Proliferate When Rare. *Science, 328*.
Bruhin, A., Janizzi, K., & Thöni, C. (2020). Uncovering the heterogeneity behind cross-cultural variation in antisocial punishment. *Journal of Economic Behavior & Organization, 180*, 291–308. https://doi.org/10.1016/j.jebo.2020.10.005
Cason, T., & Gangadharan, L. (2015). Promoting cooperation in nonlinear social dilemmas through peer punishment. *Experimental Economics, 18*(1), 66–88. https://doi.org/10.1007/s10683-014-9393-0
Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics, 14*(1), 47–83. https://doi.org/10.1007/s10683-010-9257-1
Chen, J., Lian, Z., & Zheng, J. (2023). Self-serving reward and punishment: Evidence from the laboratory. *Scientific Reports, 13*(1), 13997. https://doi.org/10.1038/s41598-023-41256-5
Cinyabuguma, M., Page, T., & Putterman, L. (2006). Can second-order punishment deter perverse punishment? *Experimental Economics, 9*(3), 265–279. https://doi.org/10.1007/s10683-006-9127-z
Cook, N. J., Grillos, T., & Andersson, K. P. (2023). Conservation payments and perceptions of equity: Experimental evidence from Indonesia, Peru, and Tanzania. *Current Research in Environmental Sustainability, 5*, Article 100212. https://doi.org/10.1016/j.crsust.2023.100212
De Geest, L. R., & Kingsley, D. C. (2021). Inequality and Peer Punishment in a Common-Pool Resource Experiment. *Strategic Behavior and the Environment, 9*(1–2), 1–26. https://doi.org/10.1561/102.00000099
de Melo, G., & Piaggio, M. (2015). The perils of peer punishment: Evidence from a common pool resource framed field experiment. *Ecological Economics, 120*, 376–393. https://doi.org/10.1016/j.ecolecon.2015.05.011
Engelmann, D., & Nikiforakis, N. (2015). In the long-run we are all dead: On the benefits of peer punishment in rich environments. *Social Choice and Welfare, 45*(3), 561–577. https://doi.org/10.1007/s00355-015-0884-5
Eriksson, K., Strimling, P., Andersson, P. A., Aveyard, M., Brauer, M., Gritskov, V., & Yamagishi, T. (2017). Cultural Universals and Cultural Differences in Meta-Norms about Peer Punishment. *Management and Organization Review, 13*(4), 851–870. https://doi.org/10.1017/mor.2017.42
Eriksson, K., Strimling, P., Gelfand, M., Wu, J., Abernathy, J., Akotia, C. S., & Anum, A. J. N. C. (2021). Perceptions of the appropriate response to norm violation in 57 societies. *Nature Communications, 12*(1), 1–11.
Falk, A., Fehr, E., & Fischbacher, U. (2005). Driving forces behind informal sanctions. *Econometrica, 73*(6).
FAO. (2020). Global Forest Resources Assessment 2020: Main report. Rome. doi:https://doi.org/10.4060/ca9825en.
Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior, 25*(2), 63–87. https://doi.org/10.1016/s1090-5138(04)00005-4
Fehr, E., & Gächter, S. (2000). Cooperation and Punishment in Public Goods Experiments. *American Economic Review, 90*(4).
Fehr, E., & Gächter, S. (2002). Altruistic Punishment in Humans. *Nature, 415*, 137–139. https://doi.org/10.1038/415137a
Fehr, E., Glätzle-Rützler, D., & Sutter, M. (2013). The development of egalitarianism, altruism, spite and parochialism in childhood and adolescence. *European Economic Review, 64*, 369–383. https://doi.org/10.1016/j.euroecorev.2013.09.006
Gachter, S., & Herrmann, B. (2009). Reciprocity, culture and human cooperation: Previous insights and a new cross-cultural experiment. *Philosophical Transactions of the Royal SocietyLond B Biol Sci, 364*(1518), 791–806. https://doi.org/10.1098/rstb.2008.0275
Gächter, S., & Herrmann, B. (2011). The limits of self-governance when cooperators get punished: Experimental evidence from urban and rural Russia. *European Economic Review, 55*(2), 193–210. https://doi.org/10.1016/j.euroecorev.2010.04.003
Gelcich, S., Guzman, R., Rodríguez-Sickert, C., Castilla, J. C., & Cárdenas, J. C. (2013). Exploring External Validity of Common Pool Resource Experiments: Insights from Artisanal Benthic Fisheries in Chile. *Ecology and Society, 18*(3). https://doi.org/10.5751/es-05598-180302
Hardin, G. (1968). Tragedy of the commons. *Science, 162*, 1443–1448.
Harrison, G., & List, J. A. (2004). Field Experiments. *Journal of Economic Literature, 42*, 1009–1055.
Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., & Ziker, J. (2010). Markets, Religion, Community Size, and the Evolution of Fairness and Punishment. *Science, 327*, 1480–1484.
Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., & Henrich, N. (2006). Costly punishment across human societies. *Science, 312*(5781), 1767–1770.
Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial Punishment across Societies. *Science, 310*. https://doi.org/10.1029/2003jb002610

Kaczan, D., Pfaff, A., Rodriguez, L., & Shapiro-Garza, E. (2017). Increasing the impact of collective incentives in payments for ecosystem services. *Journal of Environmental Economics and Management, 86*, 48–67. https://doi.org/10.1016/j.jeem.2017.06.007

Katuščák, P., & Miklánek, T. (2022). What drives conditional cooperation in public good games? *Experimental Economics*. https://doi.org/10.1007/s10683-022-09756-9

Kingsley, D. C. (2016). Endowment heterogeneity and peer punishment in a public good experiment: Cooperation and normative conflict. *Journal of Behavioral and Experimental Economics, 60*, 49–61. https://doi.org/10.1016/j.socec.2015.12.002

Koch, C., Nikiforakis, N., & Noussair, C. N. (2021). Covenants before the swords: The limits to efficient cooperation in heterogeneous groups. *Journal of Economic Behavior & Organization, 188*, 307–321. https://doi.org/10.1016/j.jebo.2021.05.003

Kosfeld, M., & Rustagi, D. (2015). Leader Punishment and Cooperation in Groups: Experimental Field Evidence from Commons Management in Ethiopia. *American Economic Review, 105*(2), 747–783. https://doi.org/10.1257/aer.20120700

Levitt, S. D., & List, J. A. (2007). What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *Journal of Economic Perspectives, 21*, 153–174.

Lohse, J., & Waichman, I. (2020). The effects of contemporaneous peer punishment on cooperation with the future. *Nature Communications, 11*(1), 1815. https://doi.org/10.1038/s41467-020-15661-7

Lopez, M. C., Murphy, J. J., Spraggon, J. M., & Stranlund, J. K. (2012). Comparing the Effectiveness of Regulation and Pro-Social Emotions to Enhance Cooperation: Experimental Evidence from Fishing Communities in Colombia. *Economic Inquiry, 50* (1), 131–142. https://doi.org/10.1111/j.1465-7295.2010.00344.x

Lopez, M. C., Murphy, J. J., Spraggon, J. M., & Stranlund, J. K. (2013). Does government regulation complement existing community efforts to support cooperation? Evidence from field experiments in Colombia. In *Handbook on Experimental Economics and the Environment*: Edward Elgar Publishing.

Masclet, D., Noussair, C., Tucker, S., & Villeval, M. C. (2003). Monetary and Non monetary Punishment in the Voluntary Contributions Mechanism. *The American Economic Review*.

Moffatt, P. G. (2015). *Experimetrics: Econometrics for experimental economics*. Macmillan International Higher Education.

Naime, J., Angelsen, A., Molina-Garzón, A., Carrilho, C. D., Selviana, V., Demarchi, G., & Martius, C. (2022). Enforcement and inequality in collective PES to reduce tropical deforestation: Effectiveness, efficiency and equity implications. *Global Environmental Change, 74*. https://doi.org/10.1016/j.gloenvcha.2022.102520

Ngoma, H., Hailu, A. T., Kabwe, S., & Angelsen, A. (2020). Pay, talk or 'whip' to conserve forests: Framed field experiments in Zambia. *World Development, 128*. https://doi.org/10.1016/j.worlddev.2019.104846

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics, 92*(1–2), 91–112. https://doi.org/10.1016/j.jpubeco.2007.04.008

Nikiforakis, N., & Normann, H.-T. (2008). A comparative statics analysis of punishment in public-good experiments. *Experimental Economics, 11*(4), 358–369. https://doi.org/10.1007/s10683-007-9171-3

Nikiforakis, N., Noussair, C. N., & Wilkening, T. (2012). Normative conflict and feuds: The limits of self-enforcement. *Journal of Public Economics, 96*(9–10), 797–807. https://doi.org/10.1016/j.jpubeco.2012.05.014

Nockur, L., Pfattheicher, S., & Keller, J. (2021). Different punishment systems in a public goods game with asymmetric endowments. *Journal of Experimental Social Psychology, 93*. https://doi.org/10.1016/j.jesp.2020.104096

Nosenzo, D., & Sefton, M. (2014). Promoting cooperation: The distribution of reward and punishment power. In P. A. M. van Lange, B. Rockenbach, & T. Yamagishi (Eds.), *Social dilemmas: New perspectives on reward and punishment Oxford*. Oxford University Press.

Ones, U., & Putterman, L. (2007). The ecology of collective action: A public goods and sanctions experiment with controlled group formation. *Journal of Economic Behavior & Organization, 62*(4), 495–521. https://doi.org/10.1016/j.jebo.2005.04.018

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.

Ostrom, E. (1998). A Behavioral Approach to the Rational Choice Theory of Collective Action: Presidential Address, American Political Science Association, 1997. *American Political Science Review, 92*(1), 1–22. https://doi.org/10.2307/2585925

Ostrom, E., Walker, J. M., & Gardner, R. (1992). Covenants with and without a sword: Self governance is possible. *American Political Science Review, 86*(2), 404–417.

Ozono, H., Kamijo, Y., & Shimizu, K. (2017). Punishing second-order free riders before first-order free riders: The effect of pool punishment priority on cooperation. *Scientific reports, 7*(1), 1–9.

Pfaff, A., Rodriguez, L. A., & Shapiro-Garza, E. (2019). Collective Local Payments for ecosystem services: New local PES between groups, sanctions, and prior watershed trust in Mexico. *Water Resources and Economics, 28*. https://doi.org/10.1016/j.wre.2019.01.002

Pfattheicher, S., Böhm, R., & Kesberg, R. (2018). The Advantage of Democratic Peer Punishment in Sustaining Cooperation within Groups. *Journal of Behavioral Decision Making, 31*(4), 562–571. https://doi.org/10.1002/bdm.2050

Reuben, E., & Riedl, A. (2013). Enforcement of contribution norms in public good games with heterogeneous populations. *Games and Economic Behavior, 77*(1), 122–137. https://doi.org/10.1016/j.geb.2012.10.001

Rustagi, D., Engel, S., & Kosfeld, M. (2010). Conditional Cooperation and Costly Monitoring Explains Success in Forest Commons Management. *Science, 330*, 961–963.

Shreedhar, G., Tavoni, A., & Marchiori, C. (2020). Monitoring and punishment networks in an experimental common pool resource dilemma. *Environment and Development Economics, 25*(1), 66–94. https://doi.org/10.1017/s1355770x19000457

Sills, E. O., de Sassi, C., Jagger, P., Lawlor, K., Miteva, D. A., Pattanayak, S. K., & Sunderlin, W. D. (2017). Building the evidence base for REDD+: Study design and methods for evaluating the impacts of conservation interventions on local well-being. *Global Environmental Change, 43*, 148–160. https://doi.org/10.1016/j.gloenvcha.2017.02.002

Sills, E. O., Atmadja, S. S., de Sassi, C., Duchelle, A. E., Kweka, D. L., Resosudarmo, I. A. P., & Sunderlin, W. D. (2014). *REDD+ on the ground: A case book of subnational initiatives across the globe*: Cifor.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Crc Press.

Sutter, M., Haigner, S., & Kocher, M. G. (2010). Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations. *Review of Economic Studies, 77*, 1540–1566. https://doi.org/10.1111/j.l467-937X.2010.00608.x

Thöni, C. (2013). Inequality aversion and antisocial punishment. *Theory and Decision, 76* (4), 529–545. https://doi.org/10.1007/s11238-013-9382-3

Vollan, B., Pröpper, M., Landmann, A., & Balafoutas, L. (2019). Self-Governance and Punishment: An Experimental Study among Namibian Forest Users. *Economic Development and Cultural Change, 67*(4), 935–967. https://doi.org/10.1086/700098

Weber, T. O., Weisel, O., & Gächter, S. (2018). Dispositional free riders do not free ride on punishment. *Nature Communications, 9*(1), 2390.

Weng, Q., & Carlsson, F. J. J. o. P. E. (2015). Cooperation in teams: The role of identity, punishment, and endowment distribution. *126*, 25-38.

Wilson, D. S., Ostrom, E., & Cox, M. E. (2013). Generalizing the core design principles for the efficacy of groups. *Journal of Economic Behavior & Organization, 90*, S21–S32. https://doi.org/10.1016/j.jebo.2012.12.010

Zelmer, J. J. E. E. (2003). Linear public goods experiments: A meta-analysis. *6*, 299-310.