Luck or merit? Inequality and punishment in a public goods game

Ananish Chaudhuri[1]

Scott Claessens[2]

Michalis Drouvelis[3]

Abhijit Ramalingam[4,*]

Brock Stoddard[5]

Quentin Atkinson[6]

July 2025

We provide causal evidence that the less well-off are more accepting of inequality arising from merit rather than luck using a public goods experiment with heterogeneous endowments. In one treatment, participants can vote to implement a costly centralised mechanism that punishes free riding. Behaviour in this treatment is compared to a control treatment where no such sanctioning in available. The less well-off benefit monetarily and can reduce inequalities in earnings by implementing deterrent punishments for free riding but do so only when endowment differences are based on luck rather than merit, suggesting greater tolerance for merit-based inequality.

JEL Codes: C72; C91; C92; D91; H41

Keywords: Inequality; Public goods; Cooperation; Punishment; Voting; Online experiment

Acknowledgements: We thank Yudou Yang for excellent research assistance. We gratefully acknowledge funding from the University of Auckland via a Faculty Research Development Fund. Ethics approval was obtained from the University of Auckland Human Participants Ethics Committee (Ref. 024639). We thank participants at numerous seminars and conferences for helpful feedback. We are responsible for all errors in the paper.

[1] Department of Economics, University of Auckland and CESifo, Munich, a.chaudhuri@auckland.ac.nz.

[2] School of Psychology, University of Kent, scla896@aucklanduni.ac.nz.

[3] Department of Economics, University of Birmingham and CESifo, Munich, m.drouvelis@bham.ac.uk.

[4] Department of Economics, Appalachian State University, 416 Howard Street, Peacock Hall 3101A, Boone, NC 28608, USA. abhi.ramalingam@gmail.com. [*] Corresponding Author.

[5] Department of Economics, Appalachian State University, stoddardbv@appstate.edu.

[6] School of Psychology, University of Auckland, q.atkinson@auckland.ac.nz.

1. Introduction

Inequality has featured prominently both in academic and popular discourse in recent times. See, for instance, the voluminous research undertaken at the International Inequalities Institute of the London School of Economics.[1] Piketty (2014) points out that income and wealth inequality in developed nations are approaching levels not seen since before the Second World War. However, a key issue here is that there are individual differences in attitudes towards and tolerance of inequality depending on its source. Simply put, inequality can be the outcome of merit or luck (random chance). Almås et al. (2019) show that across a large cross-section of countries, there is higher (lower) tolerance of inequality, and even extreme inequality – measured via imputed values of the Gini coefficient – when such inequality results from differences in merit (luck). Starmans et al. (2017) and Almås et al. (2010) provide evidence that these feelings are present even among children and adolescents. Macchia et al. (2020) report that many, particularly in developing nations, are tolerant of inequality since the relatively well-off actually feel a greater sense of well-being from their higher income rank. Indeed, as Almås et al. (2019) highlight, cross-country tolerance of inequality decreases with an increase in per capita GDP, with citizens of richer countries expressing greater aversion toward inequality.

Alesina et al. (2004) document a clear division in attitudes on two sides of the Atlantic, with Europeans being less tolerant of inequality than Americans. Alesina and Angeletos (2005) highlight how different beliefs about the fairness of social competition and the causes of inequality may influence redistributive policies. They show that based on beliefs regarding the causes of inequality and the degree of social mobility, there may exist multiple social equilibria. A society that believes individual effort (or the "signal" according to Alesina and Angeletos,

[1] https://www.lse.ac.uk/international-inequalities.

2005) determines income, will choose low taxes and low redistribution (and end up with potentially higher inequality) while a society that puts more weight on luck ("noise") in the form of accidents of birth, connections and/or corruption, may well choose higher taxes and higher redistribution resulting in lower inequality. In equilibrium, perceptions of the signal to noise ratio will have implications for the choice of taxes, other redistributive policies and the size of the welfare state.

To the extent that the merit versus luck issue raises questions about earned versus unearned income and/or wealth, those who are less well-off (henceforth, for expositional ease, the "poor") may feel more aggrieved when inequality is a result of luck while the better off (the "rich") may feel more entitled to their wealth, especially when this comes from merit. Moreover, it is not clear if the poor and rich will support the same policies depending on the underlying causes of inequality.

In this paper, we explore the effects of these two different sources of inequality on cooperative behaviour as well as preferences for redistribution through punishment of free riding in a paradigmatic linear public goods game (Ledyard, 1995). Specifically, we ask: (1) How does the source of inequality (luck or merit) impact cooperation with or without a centralised sanctioning mechanism that punishes free riding? (2) Do the poor and the rich vote for differing levels of sanctioning? (3) Does this depend on the source of inequality?

It is difficult, if not impossible, to ascertain the true cause of variation in wealth using natural data. Typically, both luck and merit are involved. A practical and cost-effective way to establish a causal relationship is to rely on controlled experiments (e.g., Falk and Heckman, 2009; Chaudhuri, 2021) where the true reason behind wealth dispersion can be exogenously manipulated. Inequality in this study is implemented by varying subjects' initial endowment of resources. Some readers of earlier drafts have questioned the external validity of trying to assess a multi-dimensional concept such as inequality via lab experiments. But studies that rely on surveys suffer from drawbacks too, at least because such surveys often capture attitudes rather than behaviour. Natural experiments can be useful but are often hard to find, and their generalisability is also unclear. Indeed, Smith (1976), in developing his induced value theory, points out that experiments can serve both as an empirical pretest prior to field testing as well as a means of interpreting field data. In any event, as the many studies cited in our work highlights, there is now a voluminous literature that explores responses to and tolerance of inequality via economic experiments. Clark and D'Ambrosio (2015, Section 13.2.2) provide an overview of this line of work.

In an online experiment, we create three different types of unequal societies and allow participants to vote on policy that could reduce inequality. In our "merit" treatment, individuals' endowments (and therefore rich or poor status) are determined by relative performance in a prior real-effort task. Our "luck" treatment creates a situation where participants perform the same real-effort task to start with, but endowments are determined purely by random chance. In the "uncertain" treatment, endowments may be determined, with equal chance, by merit (based on performance in the prior real-effort task) or by luck.

In our "uncertain" treatment, participants know that there is a half-chance that rich or poor status depends only on merit and a half-chance that it depends only on random chance. The "uncertain" treatment is novel and important since, in reality, the distinction of whether rich/poor status is the result of merit or luck is often unclear. This treatment then allows us to understand whether poor and rich participants perceive this third situation to be similar to luck-based or merit-based inequality.

Comparing across these different groups allows us to understand whether and how differences in the source of wealth heterogeneity leads to differences in contribution patterns. Then, we examine individual policy preferences in such settings. We look at sanctioning of free riding by allowing participants to establish a costly centralised scheme for punishing free riding in the different societies. Participants in a group can vote for and pay the same small price (of course, a relatively larger proportion of endowment for the poor compared to the rich) to implement a central sanctioning scheme. The punishment scheme is infallible and represents a certain punishment on free riding, with the level of the punishment being determined by majority vote. In any round, the part of one's endowment that is not contributed toward the public good is subject to this punishment if and when participants successfully implement a sanctioning mechanism. This stylised setting minimises potential confounds, and the approach is similar to those implemented in other studies exploring centralised punishment for free riding. See, for instance, Kingsley (2016), Markusen et al. (2014), Putterman at al. (2011) and Tyran and Feld. (2006).

Our findings may be summarised as follows. In general, compared to the rich, the poor contribute a much larger proportion of their endowment to the public good. But in the institution that allows punishment for free riding, the poor free ride significantly more when inequality is luck-based than when it is merit-based. Further, within this punishment institution, the poor vote for punishment more often when inequality results from luck than from other sources. Given that punishment for free riding results in greater redistribution, increased earnings and reduced inequality, it is striking that this tendency on the part of the poor to punish free riding is preeminent in the luck treatment and not the others. We interpret this as a signal that even the poor find merit-based inequality to be more tolerable and are therefore less willing under merit-based inequality to free-ride or to punish the rich for contributing a relatively smaller proportion of their endowment towards a public good. By and large, there is little difference in the behaviour of the rich in terms of voting or contribution across different sources of inequality.

A final interesting observation is that patterns of choices made by the rich and poor, with regards to cooperation and punishment, are similar in both the merit-based and uncertain treatments, suggesting both perceive the uncertain treatment as being similar to the merit-based treatment. It appears that people treat a system that has some semblance of merit as being similar to a system based on merit.

We proceed as follows. In Section 2, we explain our experimental design and procedures. In Section 3, we present hypotheses based on existing research findings. In Section 4, we present our results, and in Section 5, we make some concluding remarks.

2. Experimental design and procedures

There were two parts in all sessions of the experiment. In Part 1, participants completed a real-effort task, performance on which, depending on the treatment, determined a participant's endowment in Part 2. In Part 2 of the experiment, participants played a repeated public goods game that could additionally include opportunities for punishing free riding. Instructions to all parts and treatments of the study are available in Appendix A.

2.1 Part 1: Real effort task

Participants first individually and independently completed a word coding task as in Erkal et al. (2011) for three minutes. In the task, participants are presented with a sequence of randomly formed three-letter 'words', which they translate to a numeric code using a table that assigns a number to each letter of the alphabet. Participants are presented with one 'word' at a time and are presented with the next one upon submitting a code. Each correct code submitted earns a participant a piece rate of 30 tokens. Participants are not informed of whether they submitted a correct code and are not shown a running total of the number of attempted or accurately completed codes. They are also not informed of the total number of correct codes and earnings from Part 1 until the conclusion of the experiment. This avoids any potential wealth effects, however small. Participants are not aware whether and how Part 1 earnings may impact Part 2 of the study since the instructions to the second part only after the completion of Part 1.

2.2 Part 2: Public goods game

In all treatments, participants play a two-stage linear public goods game in fixed groups (partner matching) of four that is repeated for 10 rounds.

2.2.1 The base decision setting

In the base game, each player receives an endowment of $e_i > 0$ tokens and independently and simultaneously decides how many tokens, $c_i \in [0, e_i]$, to contribute to a group account (the public good), with the remainder, $(e - c_i)$, being automatically allocated to his/her private account. A player earned one token for each token retained in his/her private account, and a fraction m $(0 < m < 1 < 4m)$ of the total contribution to the public good by all group members, denoted as $C = \sum_{j=1}^{4} c_j$. Our chosen parameter value of the marginal per capita return (MPCR) of m = 0.5 satisfies the conditions above. Given the restrictions on m, the base game is a social dilemma that captures the conflict between privately and socially optimal actions. Assuming self-interested players who only care about monetary payoffs, the unique Nash equilibrium of the game is zero contribution by each group member. The social optimum is 100% contribution by each member. Under finite repetition, the subgame perfect Nash equilibrium is zero contribution by all, and the social optimum remains unchanged.

2.2.2 Voting for punishment

In one treatment that allows for sanctioning of free riding, the public goods game is preceded by an initial stage, where group members simultaneously vote to implement one of four punishment rates – 0, 30%, 60%, or 80% – in their group in that round. An implemented punishment rate reduces each group member's private-account earnings by that percentage, whether they voted for it or not. A punishment rate is chosen by a simple majority, i.e., a 'first-past-the-post' rule. Any punishment rate that receives at least two votes in a group is implemented for that round. If two rates each receive two votes, one of them is chosen randomly for implementation. Regardless of what punishment rate a participant chooses, if a positive punishment rate is implemented in a round, each group member pays a fee of 4 tokens from his/her earnings in the round. There is no fee if a zero-punishment rate is implemented.

A punishment rate of zero does not change the fully free-riding equilibrium. Neither does a punishment rate of 30%, since retaining 70% of a token in one's private account is more profitable than contributing it to the public good and receiving a return of 50% (since $m = 0.5$). Hence, a punishment of 30% is *non-deterrent*. However, punishment rates of 60% and 80% ensure that the loss from retaining a token in the private account is greater than 50%, which would be the loss from contributing it to the group account. Thus, with a higher punishment rate, full contribution becomes the dominant strategy. Hence, punishment rates of 60% and 80% are *deterrent*.

We chose two different values for the deterrent punishment because we wanted players to decide, first, whether to implement a punishment or not (0%) and, second, whether to implement a non-deterrent (30%) or deterrent punishment (60% or 80%). Providing three choices of 0, 30, and either 60 or 80 may have made 30% focal.[2] The non-deterrent 30% fine is analogous to the "endogenous mild law" treatment implemented by Tyran and Feld (2006). As they argue, a non-deterrent fine or "mild law" makes the contribution to a public good more of an obligation and "induces compliance not by deterrence but by norm-activation" (p. 136). Further, as Markusen et al. (2014, p. 302) argue, one reason for implementing a non-deterrent fine is that

> "…the penalty required to achieve deterrence may be considered to violate social standards of reasonableness, in part because of the possibility that violation occurred

[2] The data show that when participants chose to implement a punishment rate, they either voted for the non-deterrent (30%) punishment or the 80% punishment rate, i.e. the larger of the two deterrent punishment rates. There were very few votes for the 60% punishment rate.

due to error or ignorance or that a rule-complying individual is wrongly penalized. A non-deterrent sanction may nonetheless deter most rule violation when it expresses a norm that citizens internalize or when being caught violating the rule brings informal as well as formal penalties, e.g. social disapproval."

Note that such costly punishment is standard in the literature dealing with centralised and decentralised punishment. In particular, the centralised sanction in our setting is not a tax that is levied on non-contributors and is not redistributed among any of the group members. The sanction amount is simply lost to the group and is thus costly and potentially inefficient.[3] See discussions of centralised punishment mechanisms in, for instance, Baldassari and Grossman (2011), Kosfeld et al. (2009), Markusen et al. (2014), Putterman et al. (2011) and decentralised punishment in Fehr and Gächter (2000) and Chaudhuri (2011). Given that implementation depends on voting by participants, our sanctioning mechanism incorporates elements of democratic punishment as in Ambrus and Greiner (2019) and Pfattheicher et al. (2018).

That all players paid the same fee of 4 tokens implies that the poor paid a share of their endowment that was four times that of the rich (20% vs. 5%). We chose this payment scheme rather than one that exacted the same share of endowment from the poor and the rich for its simplicity and ease of explanation to participants. Moreover, the fee as a share of endowment 20% for poor and 5% for rich) mirrors the 4:1 imbalance in the ratio of endowments between the rich and the poor.

In the second stage, group members are informed of the punishment rate implemented in the first stage before they make contribution decisions in the round. The punishment is

[3] Prior research has noted that the efficiency implications of punishments are ambiguous; whether punishments raise efficiency or not depends crucially on the cost-benefit ratio of said punishments. See Chaudhuri (2011); Egas and Riedl (2008) and Nikiforakis and Normann (2008).

automatically and universally applied to all group members, regardless of how they voted. The payoff to a group member in each round is given by the following:[4]

$$\pi_i = \begin{cases} (e_i - c_i) + mC & \text{if } p = 0 \\ (1-p)(e_i - c_i) + mC - 4 & \text{if } p \in \{0.3, 0.6, 0.8\} \end{cases},$$

where p is the implemented punishment rate.

2.2.3 Treatments

In all treatments, two 'poor' group members receive a per-round endowment of 20 tokens each, and two 'rich' group members receive a per-round endowment of 80 tokens each.[5] In all cases, endowments are decided once at the beginning of Part 2 and then remain fixed for all 10 rounds in Part 2. In the luck treatment, two of the four group members are randomly assigned to be poor (have 20 tokens) while the other two are rich (have 80 tokens). This assignment is independent of performance in the word coding task in Part 1. In merit, endowments are decided based on performance in the effort task in Part 1. All four group members are ranked according to their performance in the word coding task. The two group members with the higher (lower) performance in Part 1 receive the higher (lower) endowment.[6]

In the uncertain treatment, there is a 50% chance that endowments within a group will be allocated according to merit (i.e., the top two performers in the group are rich and the bottom two are poor). Else, endowments are allocated according to luck (i.e. two randomly chosen participants in the group are poor and the other two are rich). Group members in this treatment

---

[4] There is no possibility of negative earnings and the minimum possible is zero. To see this, consider the case of a poor participant who votes for a deterrent 80% punishment at a cost of 4 tokens, which is implemented. Then this participant allocates all of his/her 20 tokens to the private account. With the 80% punishment rate this participant is fined 16 tokens leaving him/her with 4 tokens. Once the 4-token fee is deducted this participant ends up with zero.

[5] We chose an equal number of poor and rich members within unequal groups to avoid complications that might arise due to the presence of a majority and a minority within groups (see Oxoby and Spraggon, 2013), such as tacit coalition formation when voting on institutional choice.

[6] Any ties are broken randomly.

know the exact rule used to decide endowments, know their endowments, but do not know exactly how that endowment came about. We have a 2 × 3 design, leading to a total of 6 treatments. Table 1 summarises our treatments and lists the number of participants (and independent groups) in each.

Table 1: Experimental design: summary of treatments and number of participants

	Source of inequality			
Vote for punishment?	Luck	Merit	Uncertain	Total participants (Groups)
No	*Luck*	*Merit*	*Uncertain*	
	76 (19)	68 (17)	72 (18)	216 (54)
Yes	*Luck – Pun*	*Merit – Pun*	*Uncertain – Pun*	
	68 (17)	48 (12)	48 (12)	164 (41)
Total	144 (36)	116 (29)	120 (30)	380 (95)

Figures in parentheses are the numbers of independent groups of four members.

2.3 Procedures

The experiments for this study were carried out via the on-line platform Prolific (https://www.prolific.co). To minimise location, timing and cultural differences among participants, we restricted participation to adults (18+ years old) based in the United Kingdom. The average age of participants was 34.8 (min = 18, max = 82) with 69.5% of participants being female. Participants log into the Prolific site where they first read a description of the study design, which includes the projected length and the expected payment from taking part. Those who consented to participate were directed by Prolific to Heroku, a cloud application platform that ran our oTree program (Chen et al., 2016).

Participants were told there would be two parts to the experiment but were only presented with instructions for Part 1 to start with. Upon completion of Part 1, subjects read the instructions for their assigned treatment, worked through a few illustrative examples and control questions[7], and played a hypothetical practice round with automated others. They then arrived at a waiting room. As soon as a group of four was assembled in the waiting room, they were assigned their endowments according to the treatment (luck/merit/uncertain) and then began the public goods game (Part 2). Participants were told the source of inequality in their groups in the instructions and were reminded of this on the decision screen in each round. Participants remained anonymous at all times. Group composition remained unchanged throughout the session.

Participants were informed of their performance and earnings in Part 1 only at the end of the experiment. In each round in Part 2, group members were informed of the outcome of the vote (i.e., the implemented punishment rate) before they made contribution decisions. They were not informed of the votes of other group members at any point. At the end of each round, participants were informed of total contributions to the public good in their group in the round, and their own earnings in the round. They were not informed of individual contributions or earnings of other group members.

At the end of the experiment, participants answered a few demographic questions (see Appendix A).[8] Finally, participants rated the fairness (on a 5-point Likert scale: 1 = Very unfair, to 5 = Very fair) of the mechanism used to determine the distribution of endowments in their group in Part 2. We could have asked about perceived fairness at the outset immediately after the participants learn how endowments had been decided. However, this has the potential to prime participants and lead to experimenter demand effects affecting behaviour. On balance, we felt that asking this question later rather than earlier made more sense. Because these

---

[7] In case participants answered a question incorrectly, they were presented with an explanation of the required calculation and were asked to answer the question again. They could not proceed until they had answered each question correctly.

[8] In addition, they answered a series of questions designed to measure their 'Social Dominance Orientation' (Sidanius and Pratto, 1999) and 'Right Wing Authoritarianism' (Altemeyer, 1981).

fairness ratings are not immediately relevant, we present an analysis of the same in Appendix B, rather than in the main text.

Participants were paid their earnings from Part 1 and accumulated earnings from all 10 rounds in Part 2. Token earnings were converted to cash using an exchange rate of 250 tokens to GBP 1. Participants took around 40 minutes on average to complete the experiment, and earned an average of GBP 11, including a GBP 5 fixed payment for completing the experiment. The experiments reported in the study were approved by the University of Auckland Human Participants Ethics Committee (Ref. 024639).

We collected data for a total of 492 participants. Out of these 380 took part in the three inequality treatments without or with punishment. In addition, we have data for an additional 112 participants who took part in "equal" treatments where every participant had the same endowment of 50 tokens, with 64 participants in a treatment without punishment and 48 in a treatment with punishment. We do not include the data from these equal treatments in the current paper since the focus here is on understanding the impact of the source of inequality rather than comparing behaviour in equal and unequal societies.

2.4 Dropouts and sample size

Participants could exit the experiment at any time by closing their internet browser. Once they quit, they could not re-join the experiment. In Part 2, this presents a problem for group members who do not quit the study. To deal with this issue, we implemented a time limit for decisions in a group. If a group member did not enter a decision in a round (either because they quit or due to inattention) within 45 seconds, a vote for zero punishment was entered for that member (in the Punishment treatments) and his/her entire endowment was contributed to the group account in that round. If that group member did not enter a decision on time for two consecutive rounds, that member and his/her entire group were taken to the end of the study. That member (and those who quit at any stage) was not paid at all, while the other group members were paid for the rounds they completed. All participants were informed of this drop-out procedure at the beginning of the experiment and were shown the time remaining to submit decisions on every decision screen.

We experienced a relatively large dropout rate but even then our sample size is large by experimental standards. The experiments described in this study are more elaborate and longer lasting compared to typical online experiments. However, the dropouts are not unusual by the standards of online experiments. Arechar et al. (2018) provides an excellent overview of the methodological and practical challenges facing those undertaking such complex interactive experiments in an online environment. They conduct an online public goods experiment with decentralised punishment. In their experiment, 18% of all participants dropped out leading to only 53% of all groups that began the experiment completing it. Our experiments are longer lasting than theirs since we have the additional coding task at the outset. A detailed ex-ante and ex-post power analysis and an analysis of dropouts during the online sessions are discussed in Appendices B and C, respectively.

3. Hypotheses

3.1 Contribution behaviour in the absence of punishment

Prior results suggest that in the presence of endowment heterogeneity, and the absence of punishment, the rich contribute a smaller proportion of their endowment to the public good compared to the poor (e.g., Anderson et al., 2008; Cherry et al., 2005; Buckley and Croson, 2006; Hargreaves Heap et al., 2016; Zelmer, 2003). This is in line with Sugden's (1984) theory of reciprocity which suggests that the rich will only match the poor's contribution in absolute rather than relative terms. Therefore, the contribution of the rich proportional to their endowment will always be lower than that by the poor.

A large literature in self-serving biases (see Babcock and Loewenstein, 1997 for an overview) suggests that the rich may feel entitled to their wealth and believe that this is the result of merit rather than luck, even where it is mostly due to luck. Similarly, the poor may feel hard done by, even where inequality is the result of merit, and may attribute this to luck alone.

This leads to our first hypothesis.

Hypothesis 1*: In the absence of punishment:*

(a) the rich will contribute a smaller proportion of their endowment to the public good.

(b) the rich will free ride more when inequality is based on merit or uncertainty than luck, and

(c) the poor will free ride more when inequality is based on luck or uncertainty rather than based on merit.

3.2 Contribution and voting behaviour in the presence of punishment

We need to consider the incentives of the rich or the poor to vote a for a particular fine separately. Poor group members have every incentive to vote for a deterrent fine. In the absence of such, the best response is to fully free ride. In this scenario, a poor group member earns 20 tokens, i.e., just his/her endowment for the round. On the other hand, with full contributions, a poor player earns 96 tokens (= (200 × 0.5) – 4 token fee). The rich have little incentive to implement a deterrent punishment. The rich earn 80 tokens by free riding. If a deterrent fine is implemented and everyone contributes their entire token endowment, the rich stand to make 96 tokens (100 tokens – the 4 token cost). However, if there is no fine and a rich player free-rides, but at least 32 tokens are contributed by others whether due to other-regarding preferences or errors, the rich player stands to earn the same amount.

The above voting tendencies are likely to be moderated by the level of acceptance of inequality, which in turn depends on the source of inequality. As mentioned earlier, people are generally more accepting of inequality when it is based on merit than when it is based on luck

(Starmans et al. 2017; Almås et al. 2010). The poor are less likely to be accepting of luck-based inequality. Hence, we hypothesise that the poor are more likely to vote for higher punishment rates in luck than in the other two treatments. Prior research suggests that when participants earn their endowments, they feel more attached to it (e.g., Hoffman et al., 1994; Cherry et al., 2002) and are, therefore, less inclined to engage in pro-social behaviour. Thus, we hypothesise that the rich are more likely to vote for punishment that does not penalise free riding in merit than in the other conditions. This leads to our second and third hypotheses.

Hypothesis 2: *(a) The poor will vote for a deterrent punishment, particularly when inequality is based on luck or uncertainty rather than merit, and*

(b) The rich will vote for zero punishment, particularly when inequality is based on merit or uncertainty rather than luck.

Hypothesis 3: *(a) In the presence of a non-deterrent punishment, contribution patterns of both the rich and the poor will be along the same lines as in Hypothesis 1.*[9]

(b) In the presence of a deterrent punishment, contributions by both rich and poor will be close the social optimum. This will result in higher earnings for the poor and lower inequality among the two classes.

While the literature looking at punishments in public goods games is voluminous, the literature looking at how punishments affect cooperation in unequal groups is less so. Kingsley (2016) suggests that the effectiveness of peer punishment is weakened by endowment heterogeneity. Waichman (2020) finds peer punishment successfully increases efficiency when endowment heterogeneity exists, but only when the rich also have stronger capabilities. Nockur

---

[9] (i) The rich will contribute a smaller proportion of their endowment to the public good matching contributions by the poor in absolute but not in relative terms; (ii) the rich will free ride more when inequality is based on merit or uncertainty than luck and (iii) the poor will free ride more when inequality is based on luck or uncertainly rather than based on merit.

et al. (2021) look at the impact of democratic peer punishment (majority vote in four member groups) and centralised punishment in the presence of endowment heterogeneity, where such heterogeneity is always generated randomly thereby corresponding to our luck treatment. However, the centralised punishment in their study was neither deterrent nor non-deterrent, leaving participants indifferent between contributing to the public account and free riding.[10] Nockur et al. find that the presence of endowment heterogeneity does not make much of a difference. Punishments increase contributions over a no punishment benchmark, and although peer punishment leads to higher contribution than centralised punishment, efficiency is higher with centralised punishment. (This may be an artefact of the relatively low 1:2 cost benefit ratio of peer/democratic punishments in this study.)

4. Results

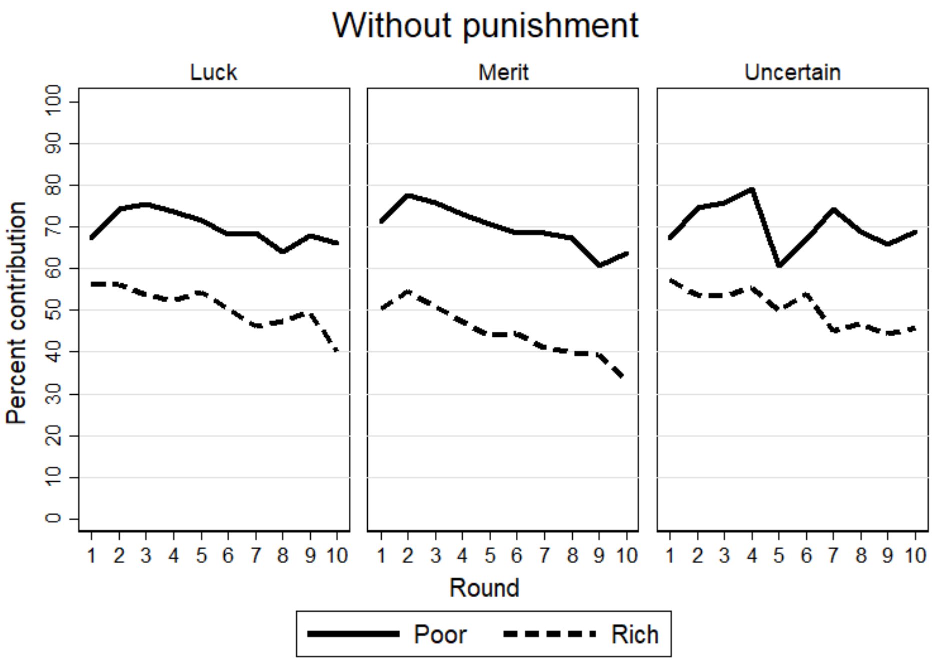
4.1 Cooperation in the absence of sanctioning.

In this part we look at contribution patterns in the absence of punishments using data from 216 participants (54 groups) in the relevant treatments. As noted earlier, we expect the rich to contribute less than the poor in proportional terms. Further, we expect that the contributions of the poor will be lower when inequality results from luck or uncertainty than from merit. We expect greater free riding on the part of the rich when inequality results from merit or uncertainly and less so in luck. The issues have been studied before in the citations we provided earlier. These results are designed to serve as a benchmark for our subsequent results on the choice of sanctioning mechanism by the poor and the rich across the different treatments. Figure 1 shows the pattern of contributions across the three treatments, luck, merit and

[10] The MPCR from the public account is 0.4. But one token placed in one's private account is penalised by a 0.6 token fine. Assuming pure self-interest, this should render participants indifferent between contributing a token to the public account and placing it in the private account.

uncertain. It is clear that in proportional terms, the poor contribute more than the rich in all three treatments. Our conjecture that the poor will free ride relatively more in luck and uncertain is not validated. In all three treatments the poor contribute approximately 70% of their token endowment, while the rich contribute around 50% in the luck and uncertain treatments, and about 45% in merit.

Figure 1: Contribution pattern over ten rounds in the absence of punishments



We corroborate this via regression analysis. Table 2 presents an individual level panel random effects regression of individual percent (of endowment) contributions on a dummy for the rich (80 token endowment), a time trend and the one-period lagged average percent contribution of the other group members. This "rich" dummy is negative and significant at 1% for all three treatments. While the proportional contribution by the rich is lower in the merit treatment compared to luck or uncertain this difference is not statistically significant either on the basis of non-parametric tests or random effects panel regression.

Table 2: Regression of proportional contributions in the absence of sanctioning

	Luck	Merit	Uncertain
Rich	-20.600*** (5.489)	-27.020*** (6.510)	-21.340*** (5.323)
Round	-1.275*** (0.345)	-1.637*** (0.420)	-0.960** (0.460)
Lagged percent cont. of others in group	0.162* (0.0846)	0.250*** (0.0883)	0.163*** (0.0518)
Constant	68.920*** (7.957)	67.230*** (7.416)	67.570*** (6.158)
Observations	684	612	648

*Note: Panel RE regressions. Standard errors clustered on groups in parentheses. * p < 0.1; ** p < 0.05; *** p < 0.01. We have 76 participants in luck, 68 in merit and 72 in uncertain playing for ten rounds. We lose 76 observations in luck, 68 in merit and 72 in uncertain when we include lagged contributions.*

We conclude this section by noting that only one of the three conjectures stated in our first hypothesis is supported. In proportional terms, the poor contribute more in all three treatments. But the contribution patterns of the rich and the poor are similar across all three treatments with no significant cross-treatment differences for either the rich or the poor.

4.2 Cooperation in the presence of punishment

Next, we look at data for 164 participants (41 groups) who first voted for a fine level in the first stage prior to deciding on the contributions in the second stage. Table 3 presents average punishment rates voted for by the rich and the poor in the different Punishment treatments. Here we treat the average of each group over ten rounds as an independent observation. The table also presents two-sided p-values from signrank tests for differences between punishment rates voted for by the rich and poor within each treatment. The poor vote for higher punishment on average than do the rich in all treatments. However, and in keeping with Hypothesis 2, this difference is statistically significant only in luck. It is noteworthy that another part of our conjecture that the poor will vote for higher punishment in the uncertain treatment, owing to self-serving bias, is not borne out.

Table 3: Average individual vote for punishment rates

	Obs.	Poor	Rich	Signrank p-value
Luck	*17*	*41.88*	*24.76*	*0.0056*
		(19.49)	*(13.86)*	
Merit	12	39.17	28.17	0.2393
		(25.19)	(19.94)	
Uncertain	12	33.63	29.33	0.4556
		(21.99)	(17.59)	

Figures in parentheses are standard deviations. The unit of observation is the average punishment rate voted for by the two rich (or poor) group members over all 10 rounds within each group.

Figure 2 presents the distribution of votes for punishment rates (0%, 30%, 60% or 80%) by poor and rich group members in each treatment. Figure 3 aggregates the four different fine rates into two categories: non-deterrent (0% or 30%) and deterrent (60% or 80%). From Figure 2, it is clear the modal choice for punishments on the part of the rich is 0%. While the differences in the votes of the poor are less dramatic, if one looks at the luck treatment in Figure 2, then it is clear that the proportion of 0% votes by the poor in luck is much lower than the proportion of 0% votes in merit or uncertain. Figure 3 tells a similar story for the poor. The proportion of deterrent and non-deterrent votes is roughly equal in luck while in merit or uncertain, the poor tended to vote much more for non-deterrent fines and less for deterrent fines.

Figure 2: Distribution of votes for different punishment rates (0%, 30%, 60% or 80%)

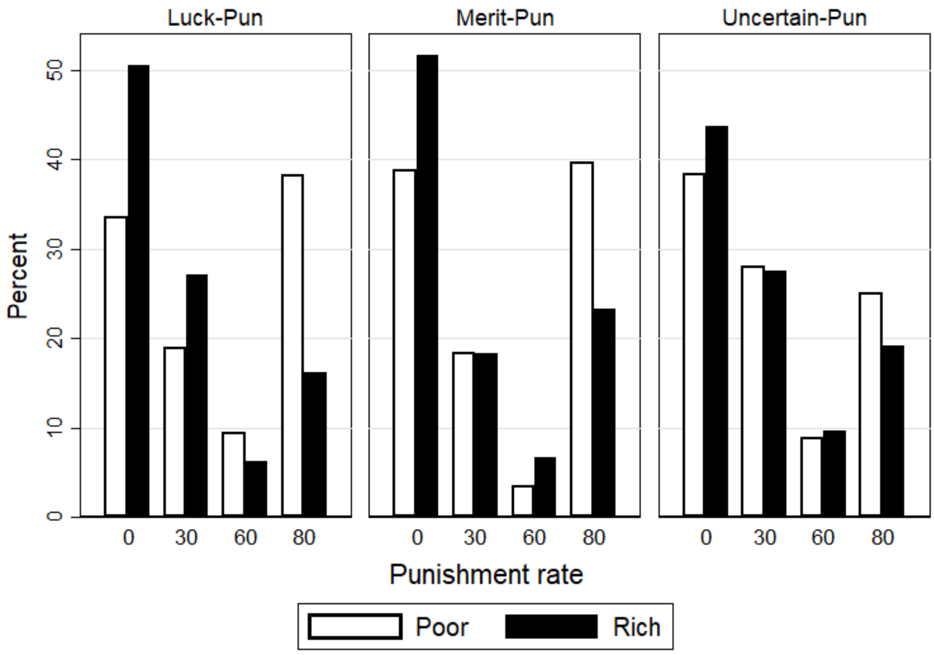
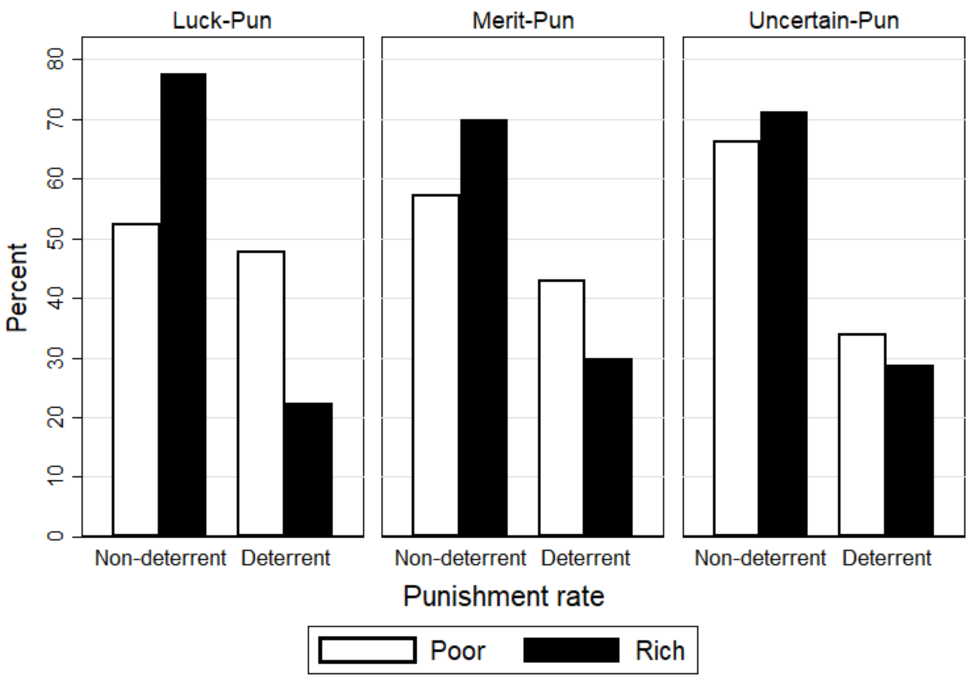


Figure 3: Distribution of votes for non-deterrent (0% or 30%) or deterrent (60% or 80%) punishments



The results presented in Table 2 and Figures 2 and 3 suggest that the poor tend to vote for higher relatively higher punishment rates only in the luck treatment compared to merit or luck. We provide corroboration for this insight by running individual-level probit regressions for votes by each fine rate. In each regression, the dependent variable is 1 if the vote of the player in a round is equal to the vote specified at the top of the column (i.e., a vote for a punishment rate of 0%, 30%, 60% or 80%) and = 0 otherwise. The regressors include two treatment dummies with luck being the reference category, a dummy for poor individuals (with endowment of 20), and interactions between the treatment dummies with the poor dummy to look for differential voting patterns between the poor and the rich. Table 4 presents marginal effects estimated after the regressions.

Table 4: Probit regressions of individual votes: Marginal effects

Individual Probit regressions	(1) Vote = 0	(2) Vote = 30	(3) Vote = 60	(4) Vote = 80
Merit	0.027	-0.292	0.039	0.259
	(0.2571)	(0.255)	(0.263)	(0.278)
Uncertain	-0.172	0.0133	0.234	0.115
	(0.213)	(0.248)	(0.212)	(0.299)
Poor	-0.440**	-0.273	0.224	0.688***
	(0.210)	(0.173)	(0.165)	(0.229)
Merit × Poor	0.112	0.273	-0.557	-0.224
	(0.340)	(0.390)	(0.374)	(0.368)
Uncertain × Poor	0.301	0.286	-0.275	-0.491
	(0.326)	(0.295)	(0.294)	(0.299)
Constant	0.0147	-0.611***	-1.540***	-0.987***
	(0.124)	(0.136)	(0.141)	(0.192)
Observations	1640	1640	1640	1640

Excluded treatment: Luck. Marginal effects from Probit regressions. Std. errors clustered on group in parentheses. * *p < 0.10,* ** *p < 0.05,* *** *p < 0.01*

Table 4 shows that in the luck treatment, the poor voted proportionately more for the highest 80% punishment and against having no punishment at all.[11] This is despite the fact, as we will show below, that implementing an 80% fine results in much higher earnings for the poor and effectively equalize the earnings between the rich and the poor equal. We will discuss this after we show what happens to contributions in the presence of sanctioning. But we take this as validation of Hypothesis 2; the poor do not vote for higher sanctions in merit or uncertain most likely because they find the inequality in those two treatments more acceptable.

4.2 Differences in contributions in the presence of fines

Table 5 shows mean percentage (of endowment) contributions by group members in all treatments and two-sided p-values for signrank tests comparing the rich and poor within treatments. The table aggregates over contributions at all punishment rates, including zero punishment. As was true earlier in the absence of punishments, in proportional terms the rich free ride more than the poor. What is different now is that there is no statistically significant difference in the contribution rates of the rich and the poor in the luck treatment. This is driven entirely by the fact that, compared to the merit and uncertain treatments, the poor free ride relatively more in luck. The average percentage contribution of the poor in luck is the lowest of the three treatments whereas there no substantial differences across treatments for the rich. This suggests that the poor free ride relatively more in the luck treatment compared to merit or uncertain. This provides partial support to our conjecture that the poor will free ride more in luck and uncertain except it is true only for luck and not for uncertain and it is only true in the presence of sanctions for free riding and not otherwise.

---

[11] While the poor were more likely to vote for deterrent punishment than the rich were in *Luck*, it did not necessarily lead to more deterrent punishment being implemented. The percentage of rounds deterrent punishment was implemented by treatment: *Luck* - 29.4%; *Merit* - 30.8%; *Uncertain* - 29.2%.

Table 5: Average individual percentage contributions

Treatment	Obs.	Poor	Rich	SR p-value
Luck	17	70.82	61.75	0.0929
		(16.26)	(18.53)	
Merit	12	81.63	61.95	0.0186
		(12.38)	(19.68)	
Uncertain	12	83.50	65.32	0.0076
		(12.6)	*(23.33)*	

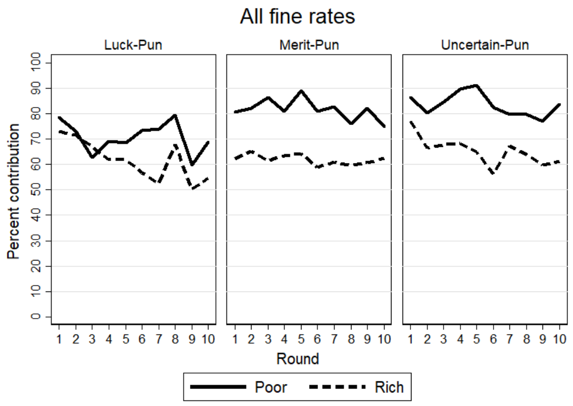
Figures in parentheses are standard deviations. The unit of observation is the average percent contribution by the two rich (or poor) group members over all 10 rounds within each group.

Figure 4 depicts the contribution patterns of the rich and the poor in the presence of punishment. Panel A aggregates contributions over all possible fine rates while Panels B through D show contribution patterns separately for each of the three fine rates 0%, 30% and 80%. We have excluded the figure of the 60% fine given that we have very few observations where this fine rate prevailed.

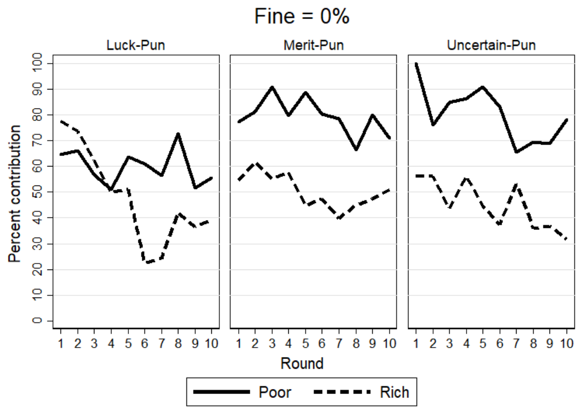
Table 6 presents average percent contributions of the poor and rich in the presence of non-deterrent (0% or 30%) versus deterrent (60% or 80%) punishment rates. The tables also present p-values from non-parametric signrank tests comparing contributions of the poor and the rich within each treatment and punishment level. Table 6 (a) shows that, in the presence of non-deterrent punishment, the contributions of the poor are lowest in luck. Further, contributions are significantly lower in luck than in merit and uncertain (Ranksum test; p = 0.04 and 0.01, respectively; n = number of groups in each pair of treatments). Table 6(b) shows that in the presence of deterrent contributions by both the rich and the poor are higher and average around 80%.

Figure 4: Contribution pattern over ten rounds in the presence of punishments

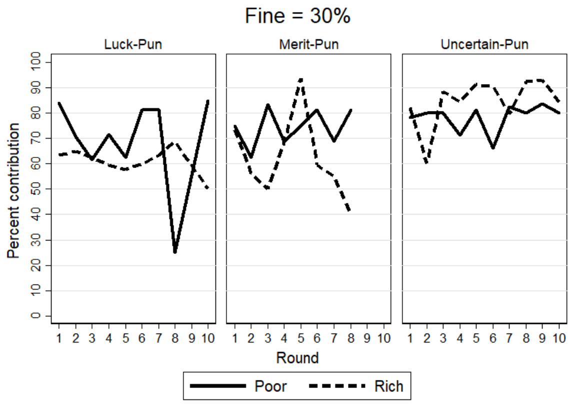
Panel A



Panel B



Panel C



Panel D

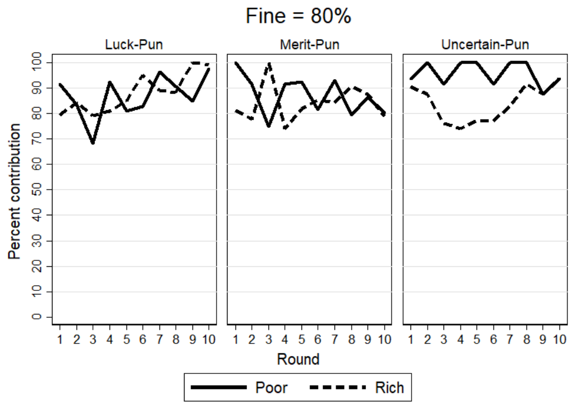


Table 6. Average percentage contributions in the presence of punishment

	(a) Non-deterrent punishment				(b) Deterrent punishment			
	Obs.	Poor	Rich	SR *p*	Obs.	Poor	Rich	SR *p*
Luck	17	62.20	52.08	0.1298	13	87.54	86.31	0.7267
		(20.28)	(17.51)			(12.83)	(14.15)	
Merit	12	77.51	52.13	0.0047	8	87.07	84.56	0.8886
		(17.32)	(19.43)			(11.52)	(16.11)	
Uncertain	11	82.12	55.91	0.0099	7	91.65	85.79	0.4990
		(14.20)	(23.47)			(11.96)	(19.23)	

Figures in parentheses are standard deviations. One group in Uncertain never voted for zero or 30% punishment. Four groups in each of Luck and Merit, as well as five groups in Uncertain, never voted for deterrent punishment (60% and 80%).

Looking at Figure 4 and Table 6, we can see that the percent contributions of the rich do not vary across treatments. As conjectured, the rich still free ride at higher levels than the poor, as a proportion of their endowment, when there is non-deterrent (0% or 30%) punishment. This difference is significant in merit and uncertain, but not in luck. The lack of significant differences in luck is driven entirely by the lower percent contributions of the poor in luck than in the other treatments. The extent of free riding by the rich, i.e., percent contribution of the rich, is not different across treatments. Not surprisingly, Table 6(b) shows that when punishment is deterrent (and these are mostly 80% fines), there are no statistically significant differences between the percentage contributions of the poor and the rich in any treatment, and across the three treatments. Moreover, percentage contributions are 'high' and close to optimal levels across the board.

In Table 7, we present individual-level panel random effects regressions of individual percent contributions in a round on a time trend, one-round lagged percentage contributions of other group members, a dummy for a deterrent (60% or 80%) punishment rate, a dummy for rich group members, and an interaction between the two dummies. For expositional ease, we present regressions for each punishment treatment separately, and report standard errors clustered on independent groups.

Table 7. Panel regressions of individual contributions: the effects of punishment

	Luck	*Merit*	*Uncertain*
Round	-1.183	-1.235*	-0.465
	(0.721)	(0.720)	(0.552)
Lagged percentage cont. of others in group	0.025	0.115	0.146***
	(0.068)	(0.101)	(0.051)
Deterrent punishment dummy	31.67***	13.65***	8.745
	(5.454)	(4.868)	(8.587)
Rich dummy	-10.96*	-26.98***	-28.13***
	(6.345)	(6.955)	(8.186)
Det. Punishment × Rich	4.870	20.50***	27.40**
	(7.942)	(6.203)	(12.39)
Constant	65.93***	77.37***	73.46***
	(7.374)	(8.061)	(8.159)
Observations	612	432	432

Panel RE regressions. Std. errors clustered on group in parentheses. * *$p < 0.10$,* ** *$p < 0.05$,* *** *$p < 0.01$*

This table provides additional support for the differences in luck compared to merit and uncertain. The rich contribute a smaller percentage of their endowments to the public good in all three inequality treatments, except in luck where the Rich dummy is only marginally significant. This suggests that the poor and rich contribute similar proportions of their endowments in luck.[12]

We note that lagged contributions of peers are significant in the uncertain treatment but not in the other two. Own contributions are typically positively correlated with lagged contribution of others. In the absence of punishment, this dependence can be explained via conditional cooperation. With punishment, evidence shows that those who contribute less than others (or the average) tend to attract greater punishment, in which case it makes sense to pay

[12] In running these regressions, we look at the treatments separately and therefore use only a part of the total data. To rectify this, we re-run these regressions using all the data and appropriate treatment dummies and interaction effects. These results provide similar insights. In the interests of parsimony, we have placed these results in **Appendix E.**

attention to what others are doing. But in our study, the punishment is *absolute* and independent of what others are doing. So, there is less of a need to condition one's own contribution on that of others. But in the uncertain treatment, since no one is certain of the circumstances surrounding endowments, it makes sense to look to others for cues as to what might be 'appropriate' behaviour.

Collectively, the evidence here suggests that in the absence of punishment or with a non-deterrent punishment, the rich do free-ride more than the poor except when inequality obtains from luck. But the reason that there are no differences in the case of luck is not that the rich increase their percentage contribution to match those of the poor, but that the poor reduce their proportional contributions to match those of the rich. In the other treatments, the poor contribute proportionally higher amounts. Furthermore, the poor vote to implement the costly 80% fine primarily when inequality is the result of luck but not when it arises from merit and uncertainty.

4.4 Impact of fines on earnings

Earlier, we noted that implementing an 80% fine increases the earnings of the poor and reduces inequality. We provide evidence in favour of this before discussing the implications of this finding. Table 8 shows average token earnings of the rich and the poor across the three treatments in the absence and presence of punishments for free riding. The key insight from Table 8 is that the earnings of the poor and the rich are virtually identical, averaging around a shade more than 80 tokens, with an 80% fine. This also appears true with a 60% fine, which makes sense since that is also a deterrent fine, but we have very few observations in this category and therefore these results may not as robust as those for the 80% fine. The results in Table 8 make clear that the poor are monetarily far better with an 80% fine and that this fine eliminates inequality between rich and poor earnings.

Table 8: Average earning across treatments in the absence and presence of punishments

Punishment	*Fine rate*	*Treatment*	*Poor*	*Rich*
No		Luck	60.72	94.14
		Merit	55.84	94.14
		Uncertain	60.69	94.24
Yes	0%	Luck	57.36	90.80
		Merit	58.69	94.64
		Uncertain	55.95	96.46
	30%	Luck	67.74	83.15
		Merit	62.18	82.79
		Uncertain	80.52	88.31
	60%	Luck	83.25	87.75
		Merit	54.00	71.00
		Uncertain	93.25	92.00
	80%	Luck	82.04	83.89
		Merit	81.88	83.81
		Uncertain	82.01	84.36

Note: The 60% fine was chosen infrequently and there are few observations for this category.

We reinforce this insight in Table 9, where we present panel random effects regressions of individual earnings broken up by the respective fine rate. We include two dummies for the merit and uncertain treatment with the luck treatment as the reference category. We also include a dummy for rich. What stands out is that on average the rich earned significantly more with a non-deterrent fine (approximately 26 tokens more with a 0% fine and 14 more with a 30% fine). The earnings for the rich are still higher with a 60% fine (with the caveat that there are few observations in this category). But with an 80% fine, there are no differences in the earnings of the rich and the poor. This suggests that the poor are always better off voting for an 80% fine. Given that there are two poor and two rich in each group, this does not guarantee that an 80% fine will be implemented. But when it is, the poor are much better off monetarily with a corresponding reduction in inequality. It is instructive that the poor choose to implement the 80% fine only in the luck treatment and not the other two. Much prior research suggests that luck-based inequality is less acceptable than merit-based inequality (e.g., Starmans et al. 2017; Almås et al., 2020). The observed behaviour of the poor in our experiment is in line with this aversion to luck-based inequality. An obvious interpretation of our results is that the poor perceive luck-based inequality as more 'unfair' and, more importantly, are willing to pay a cost to reduce inequality arising from luck.

Table 9: Random Effects regression on earnings by each fine rate

	(1) Fine=0%	(2) Fine=30%	(3) Fine=60%	(4) Fine=80%
Merit	1.163 (1.077)	0.475 (2.280)	-14.65*** (3.123)	-1.335 (4.494)
Uncertain	2.931* (1.187)	5.039* (2.003)	5.176 (3.299)	-0.580 (5.176)
Rich	26.26*** (3.859)	13.81*** (2.441)	4.844*** (1.615)	-0.866 (1.159)
Round	-0.778*** (0.246)	0.158 (0.461)	-0.039 (0.360)	1.288*** (0.440)
Lagged Earnings	0.426*** (0.058)	0.178*** (0.053)	0.044 (0.058)	0.109*** (0.036)
Merit × Poor	3.987 (3.994)	-4.345 (6.865)	-11.83*** (1.435)	-1.220 (1.406)
Uncertain × Poor	-3.726 (5.602)	8.762*** (3.309)	6.713*** (1.670)	-1.210 (1593)
Constant	32.38*** (4.405)	52.16*** (4.435)	78.21*** (6.096)	67.51*** (6.050)
Observations	756	264	32	424

Exclude treatment: Luck. The endowment of 20 (poor) is captured by the rich dummy. Panel RE regressions. Standard errors clustered on groups in parentheses.
** p < 0.1; ** p < 0.05; *** p < 0.01*

5. Conclusion

We report data from an online repeated public goods experiment that varies the source of inequality and punishment opportunities. We find that in proportional terms the rich contribute less than the poor regardless of the source of inequality. Within the punishment that allows sanctions for free riding we find that the poor (a) free ride proportionately more when inequality is the result of luck as opposed to merit or uncertain and (b) it is only when inequality results from luck that the poor respond by voting for a costly deterrent punishment. This second finding is instructive because as we document above, the poor have every incentive to vote for a deterrent punishment in all cases since this results in higher earnings for them and lower inequality. We interpret these two findings taken together to imply that the poor are more tolerant of inequality that results from reasons other than luck and they respond more forcefully when inequality is the result of luck than the result of merit. These results provide causal support for previous correlational evidence showing that people are more tolerant of merit based, rather than luck based, inequality (Almås et al. 2019, Starmans et al. 2017).

Another interesting insight is that participants appear to view the uncertain treatment as being on par with the merit treatment. In retrospect, it is possible that self-serving bias and/or misattribution of cause are certainly factors in decision making, but not in the way we initially conceived. Misattribution of cause seems to work via convincing participants that the uncertain treatment, which has an element of merit to it, is effectively similar to merit. Wilson (2003, p. 300) writes that it is possible to conflate real and apparent merit by arguing as follows:

> *As long as A_1 is believed to work harder than A_2 and/or is perceived as having more talent …whether or not she actually does, A_1 deserves to enjoy a higher level of well-being than A_2. It would be unjust if A_2 were to obtain as much as or more than A_1.*

While Wilson argues against this as being a defensible ethical standard, it is conceivable that a similar standard may have operated in the minds of our study participants. Wilson (p. 302) goes on to argue that "*Beliefs regarding performance are adjusted in order to satisfy the preference that good things happen to people judged to be good overall.*"

Most people have intrinsic notions of justice and at least by some standards, merit is considered more "just" than luck. It appears that the uncertain treatment, which appears to encompass an element of merit, is also perceived to be more just than luck. See Konow (1996, 2000) for further arguments along similar lines. Finally, Cappelen et al. (2007) argue for three distinct fairness ideals: *egalitarianism*, which argues for equalising resources; *libertarianism*, which argues for allocating resources on the basis of merit and *liberal egalitarianism*, which falls in between the two. Cappelen et al. find that 35% of their participants adhere to the libertarian ideal. For such participants, some of whom may be poor, it is clear that a merit-based system is preferred and given the arguments above regarding conflation of merit and uncertain, it may not be surprising that behaviour is similar in the two treatments. Recent years have seen a sustained debate on the question of inequality; particularly since the publication of Piketty (2014). As Piketty points out, over the last three decades the industrialised West experienced levels of inequality not seen since the early years of the 20th century. Our findings suggest that behaviour may differ substantially based on the source of inequality and their perceived fairness of lack thereof.

References

Alesina, A. and Angeletos, G-M. (2005). Fairness and redistribution. American Economic Review, 95(4), 960-980.

Alesina, A., Di Tella, R. and MacCulloch, R. (2004). Inequality and happiness: Are Europeans and Americans different? Journal of Public Economics, 88(9-10), 2009-2042.

Almås, I., Cappelen, A.W., and Tungodden, B. (2020). Cutthroat capitalism versus cuddly socialism: Are Americans more meritocratic and efficiency-seeking than Scandinavians? Journal of Political Economy, 128 (5), 1753-1788.

Almås, I., Cappelen, A.W., Sørensen, E., and Tungodden, B. (2010). Fairness and the development of Inequality Acceptance. Science, 328 (5982), 1176-1178.

Almås, I., Cappelen, A., Sørensen, E., and Tunggoden, B. (2019, November 25-26). Fairness across the world: Preferences and beliefs [Keynote presentation]. Australia New Zealand Workshop in Experimental Economics, Monash University, Clayton, Victoria, Australia.

Altemeyer, B. (1981). Right-wing Authoritarianism. Winnipeg, Manitoba. University of Manitoba Press.

Ambrus, A., and Greiner, B. (2019). Individual, dictator, and democratic punishment in public good games with perfect and imperfect observability. Journal of Public Economics, 178, 1–23.

Anderson, L.R., Mellor, J.M., and Milyo, J. (2008). Inequality and public good provision: An experimental analysis. Journal of Socio-Economics, 37(3), 1010-1028.

Arechar, A.A., Gächter, S., and Molleman, L. (2018). Conducting interactive experiments online. Experimental Economics, 21(1), 99-131.

Babcock, L., and Loewenstein, G. (1997). Explaining Bargaining Impasse: The Role of Self-Serving Biases. Journal of Economic Perspectives, 11(1), 109-126.

Baldassarri, D., and Grossman, G. (2011). Centralized sanctioning and legitimate authority promote cooperation in humans. Proceedings of the National Academy of Sciences, 108 (27), 11023–11027.

Buckley, E., and Croson, R. (2006). Income and wealth heterogeneity in the voluntary provision of linear public goods. Journal of Public Economics, 90, 935-955.

Cappelen, A.W., Hole, A.D., Sorensen, E. O., and Tungodden, B. (1997). The Pluralism of Fairness Ideals: An Experimental Approach. American Economic Review, 97(3), 818-827.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. Experimental Economics, 14(1), 47-83.

Chaudhuri, A. (2021). Behavioural Economics and Experiments. London and New York: Routledge.

Chen, D.L., Schonger, M., and Wickens, C. (2016). oTree – An open-source platform for laboratory, online and field experiments. Journal of Behavioral and Experimental Finance, 9, 88-97.

Cherry, T.L., Frykblom, P., and Shogren, J.F. (2002). Hardnose the dictator. American Economic Review, 92(4), 1218-1221.

Cherry, T.L., Kroll, S., and Shogren, J.F. (2005). The effect of endowment heterogeneity and origin on public good contributions: evidence from the lab. Journal of Economic Behavior and Organization, 57 (3), 356-365.

Clark, A. and D'Ambrosio, C. (2015). Attitudes to income inequality: Experimental and Survey Evidence. Chapter 13, pages 1147-1208 in Atkinson, A. and Bourguignon, F. (Eds.) Handbook of Income Distribution, Volume 2A. Amsterdam: Elsevier.

Croson, R. (2007). Theories of Commitment, Altruism and Reciprocity: Evidence From Linear Public Goods Games. Economic Inquiry, 45(2), 199-216.

Egas, M., and A. Riedl, (2008). The economics of altruistic punishment and the maintenance of cooperation. Proceedings of the Royal Society B, 275(1637), 871-878.

Erkal, N., Gangadharan, L., and Nikiforakis, N. (2011). Relative earnings and giving in a real-effort experiment. American Economic Review, 101(7), 3330-3348.

Falk, A., and Heckman, J.J. (2009). Lab Experiments are a Major Source of Knowledge in the Social Sciences. Science, 326(5952), 535-538.

Fehr, E., and Gächter, S. (2000). Cooperation and Punishment in Public Goods Experiments. American Economic Review, 90(4), 980-994.

Hargreaves Heap, S., Ramalingam, A., and Stoddard, B.V. (2016). Endowment inequality in public goods games: A re-examination. Economics Letters, September 2016, 4-7.

Hoffman, E., McCabe, K., Shachat, K., and Smith, V. (1994). Preferences, property rights, and anonymity in bargaining games. Games and Economic Behavior, 7(3), 346-380.

Kingsley, D. (2016). Endowment heterogeneity and peer punishment in a public good experiment: Cooperation and normative conflict. Journal of Behavioral and Experimental Economics, 60, 49-61.

Konow, J. (1996). A positive theory of economic fairness. Journal of Economic Behavior & Organization, 31 (1), 13-35.

Konow, J. (2000). Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions. American Economic Review, 90(4), 1072-1091.

Kosfeld, M., Okada, A., and Riedl, A. (2009). Institution Formation in Public Goods Games. American Economic Review, 99(4), 1335-1355.

Ledyard, J. (1995). Public goods: A survey of experimental results. Chapter 2, pages 111-194 in Kagel, J. & Roth, A. (Eds.), Handbook of Experimental Economics. Princeton, NJ: Princeton University Press.

Macchia, L., Plagnol, A.C., and Powdthavee, N. (2020). Buying happiness in an unequal world: Rank of income more strongly predicts well-being in more unequal countries. Personality and Social Psychology Bulletin, 46(5), 769-780.

Markusen, T., Putterman, L. and Tyran, J-R. (2014). Self-Organization for Collective Action: An Experimental Study of Voting on Sanction Regimes. Review of Economic Studies, 81, 301-324.

Nikiforakis, N., and Normann, H-T. (2008). A comparative statics analysis of punishment in public-good experiments. Experimental Economics, 11, 358-269.

Nockur, L., Pfatteicher, S., and Keller, J. (2021). Different punishment systems in a public goods game with asymmetric endowments. Journal of Experimental Psychology, 93, 104096.

Oxoby, R.J., and Spraggon, J. (2013). A Clear and Present Minority: Heterogeneity in the Source of Endowments and the Provision of Public Goods. Economic Inquiry, 51(4), 2071-2082.

Pfattheicher, S., Bohm, R., and Kesberg, R. (2018). The advantage of democratic peer punishment in sustaining cooperation within groups. Journal of Behavioral Decision Making, 31(4), 562–571.

Piketty, T. (2014). Capital in the 21st Century. Cambridge. MA: Harvard University Press.

Putterman, L., Tyran, J-R., and Kamei, K. (2011). Public goods and voting on formal sanction schemes. Journal of Public Economics, 95(9-10), 1231-1222.

Ramalingam, A., and Stoddard, B. (2023). Does reducing inequality increase cooperation? Working Paper. Boone, NC: Appalachian State University

Sidanius, J., and Pratto, F. (1999). Social Dominance: An Intergroup Theory of Social Hierarchy and Oppression. Cambridge: Cambridge University Press.

Smith, V.L. (1976). Experimental Economics: Induced Value Theory. American Economic Review (American Economic Association Papers and Proceedings), 66(2), 274-279.

Starmans, C., Sheskin, M., and Bloom, P. (2017). Why people prefer unequal societies. Nature Human Behaviour, 1(4), Article 0082.

Sugden, R. (1984). Reciprocity: The supply of public goods through voluntary contributions. Economic Journal, 94(376), 772-787.

Tyran, J-R. and Feld. L. (2006). Achieving Compliance When Legal Sanctions Are Non-Deterrent. Scandinavian Journal of Economics 108(1):135-156. DOI:10.1111/j.1467-9442.2006.00444.x.

Waichman, I. (2020). Linking wealth and punishment effectiveness: Punishment and cooperation under congruent heterogeneities. Economic Inquiry, 58(1), 86-103.

Weber, R., Camerer, C. Rottenstreich, Y., and Knez, M. (2001). The Illusion of Leadership: Misattribution of Cause in Coordination Games. Organization Science, 12 (5), 582-598.

Wilson, C. (2003). The Role of a Merit Principle in Distributive Justice. The Journal of Ethics, 7(3), 277-314.

Zelmer, J. (2003). Linear Public Goods Experiments: A Meta-Analysis. Experimental Economics, 6, 299–310.

Appendices: Supplementary Material available online

Appendix A: Experimental instructions

These are the instructions for the Fine treatments. Instructions for the No Fine treatments simply left out the bits related to voting and fine rate.

Welcome to the study!

Welcome to this study, funded by the University of Auckland, New Zealand.

The entire study should take no longer than **45 minutes**. Please make sure that you have that time at your disposal.

The study is designed to understand how people make allocation decisions in a group setting. You must be at least 18 years old to participate in this study. All you have to do during the study is enter numbers using your computer keyboard.

- Your responses will be kept anonymous and confidential.
- You will be paid a **guaranteed £5.00** for your participation upon completing the study. Depending on your decisions and the decisions made by others in the study, as well as the time you spend waiting to be matched, you will also earn an additional **bonus amount**. On average, this bonus amount will add up to about **another £5.00** for a total of about **£10.00**.
- Your participation is completely voluntary and you can withdraw at any time during the session. However, if you withdraw, you will not receive any payments from the study and will reduce the payments of other study participants.
- If you have any questions about this study, then you can email Principal Investigator Ananish Chaudhuri (a.chaudhuri@auckland.ac.nz).

Please participate in the study only if you have 45 minutes available, and you intend to complete the entire study immediately.

Study overview

The study consists of **multiple parts**. The instructions for **Part 1** are given here. You will receive instructions for other parts as you proceed.

In each Part, you will earn bonus payments **in tokens**. At the end of the study, tokens you earn will convert to cash at the rate of **250 tokens = £1.00**. On average, you can expect to make another £5.00 in bonus payment on top of the guaranteed £5.00 payment.

Part 1 – Instructions

In Part 1, all participants will perform the same encoding task.

You will be presented with a set of three letters that form "words" with no specific meaning.

You will also be shown a table with numbers corresponding to each letter of the alphabet.

Your task is to encode the set of three letters by substituting them with numbers from the table.

Part 1

Please use the code provided to translate the random strings into the corresponding numerical code. You have 3 minutes to complete as many encodings as you choose. For every correct encoding you submit, you will earn 30 tokens. There is no penalty for incorrect submissions.

Part 2 – Instructions

This is the main part of the study. In this Part, you will play an economic decision-making game over a number of rounds.

Before we get to the actual game, we will explain how the game works and give you a practice round.

The game:

- In this game, you will be anonymously placed in a **group of four** with **three other real participants**.
- In each round of the game, each group member will be given some tokens. This is the **token endowment**.
- Each group member will simultaneously decide how many of these tokens to allocate to the **group account**. Any tokens not allocated to the group account will be kept in the group member's **private account**.
- Tokens allocated to the group account will be **doubled** and then **redistributed equally** among **all** four group members.
- Every group member gets back the same amount from the group account.

Part 2 – Instructions

Example:

Suppose you have been anonymously placed in a group of four. Every group member is given a token endowment of 50 tokens.

You keep 10 tokens in your private account and allocate 40 tokens to the group account. The other three group members allocate an additional 60 tokens to the group account, taking the total in the group account to 100 tokens. This is doubled to 200 tokens and redistributed equally among the four group members. Each member gets back 50 tokens from the group account.

At the end of the round, you have 60 tokens; the 50 tokens you got back from the group account, plus the 10 tokens you kept in your private account.

Part 2 – Instructions

In each round, before making allocation decisions, participants will also choose a **fine rate**.

- The fine rate applies to **any tokens kept in the private account**.
- All four members will vote for **one** of the following fine rates: **0%, 30%, 60%, or 80%**.
- The fine rate that receives at least two votes will be implemented for that round. If two fine rates each receive two votes, one of those fine rates will be chosen randomly. If all fine rates each receive one vote, one of those fine rates will be chosen randomly.
- If the implemented fine rate is 0%, there is no fine.
- If the implemented fine rate is 30%, 60%, or 80%, then each participant in your group will be fined for tokens that they kept in their private account, plus an additional fixed fee of 4 tokens. This additional fixed fee will apply to each participant in the group, regardless of their allocation to the group account.

Example:

The fine rate chosen by majority is 30%.

You keep 10 tokens in your private account. You will be fined 30% of 10 tokens (3 tokens), and will pay the additional fixed fee of 4 tokens.

Part 2 – Instructions

Summary of earnings from a decision round:

If the fine rate is 0%:

Earnings = Tokens you get back from the group account + Tokens in your private account

If the fine rate is 30%, 60%, or 80%:

Earnings = Tokens you get back from the group account + Tokens in your private account - Fine - Fixed fee of 4 tokens

Part 2 – Instructions

Example:

Every group member is given a token endowment of 50 tokens. The fine rate chosen by the majority is 30%.

You keep 10 tokens in your private account and allocate 40 tokens to the group account. The other three group members allocate an additional 60 tokens to the group account, taking the total in the group account to 100 tokens. This is doubled to 200 tokens and redistributed equally among the four group members. Each group member gets 50 tokens back from the group account.

Your fine = 30% of the 10 tokens kept in your private account = 3 tokens

Fixed fee = 4 tokens

Your earning in the round = 50 from group account + 10 kept in private account - 3 fined - 4 fixed fee = ***53 tokens****.*

After all four group members have made their allocation decisions in the round, you will be shown the total number of tokens allocated to the group account, and your bonus payment for the round.

Part 2 – Comprehension

We will now ask you some questions to make sure you understand the game. Feel free to use a calculator if you would like.

Suppose that each of the four group members has 50 tokens. The fine rate chosen by the majority is 80%.

You decide to keep 30 tokens in your private account and allocate 20 tokens to the group account. The other three group members allocate an *additional* 120 tokens to the group account.

1. How many tokens are there in the group account? [*140 tokens*]
2. How many tokens are there in the group account after being doubled? [*280 tokens*]
3. How many tokens will you get back from the group account? [*70 tokens*]
4. How many tokens will you earn in this round? Remember that tokens you kept in your private account will be fined at 80% with an additional 4 token fixed fee. [*72 tokens*]

Part 2 – Practice Round

Practice round (does not count for bonus payment):

In this practice round, you are playing with three **computerised** group members.

You have a token endowment of 50 tokens.

Which fine rate do you vote for in this round? 0% / 30% / 60% / 80%

The tokens you keep in the private account will be subject to the fine rate chosen by the majority.

Part 2 – Practice Round

Practice round (does not count for bonus payment):

In this practice round, you are playing with three **computerised** group members.

You have a token endowment of 50 tokens. The majority of the group voted for a fine rate of **30%**.

How many tokens would you like to allocate to the group account in this round? 0–50 tokens

The tokens you keep in the private account will be subject to the fine rate chosen by the majority.

Part 2 – Practice Round Results

- Fine rate voted by majority: **30%**
- Tokens in private account: **{{kept}}**
- Total tokens in group account: **{{total}}**
- Tokens you got back from the group account: **{{back}}**
- Fine = 30% of {{kept}} = **{{fined}}**
- Fixed fee = **4 tokens**

Earnings for this round = {{back}} from group account + {{kept}} kept in private account - {{fined}} fined - 4 tokens additional fixed fee = ***{{earn}}***

Part 2 – Start the Game

We are now ready to start the actual game, in which you can earn real money.

There will be 10 rounds of this game. You will earn bonus payments in each of the 10 rounds.

You will be anonymously put in a **group of four** with three other **real participants**. You all received the same instructions and information. You will remain in the same group for the entire game. You will be paid £0.05 for every minute you spend waiting to be matched (maximum £0.50).

You will find out your **token endowment** after you are matched with other participants. The token endowment for each participant will remain unchanged for the entire game.

In each round of the game, you will vote for a fine rate and allocate to the group account.

IMPORTANT NOTE:

- In any round, if you do not enter your vote within the stipulated time, then your vote will be set to 0%. If you do not enter your allocation within the stipulated time, then your entire endowment will be allocated to the group account.
- If you fail to vote and/or allocate for two rounds in a row, then the game will end for you and for the other three participants in your group.
- This will mean that you will earn a smaller bonus payment. Your group members will also earn a smaller bonus payment.
- To avoid this, please pay attention and engage with the game.

Are you ready to proceed?

Yes, take me to the game

Matching you with other real participants… please wait!

Please wait for matching. We will pay you £0.05 for every minute you spend waiting to be matched (max £0.50).

You may see this screen for a long time - this is normal! You shouldn't have to wait too long. If we cannot find participants to match you with after ten minutes, we will skip Part 2 and allow you to finish the study alone.

Please be ready to begin the game **immediately** once you are matched.

Part 2 – Endowments

You have successfully been matched with three other real participants.

IF EQUALITY TREATMENT:

The token endowment for all participants is the same.

In each round, each group member will have an endowment of 50 tokens.

ELSE IF MERIT TREATMENT:

The token endowment for all participants is not the same.

The token endowment has been determined by the number of words correctly encoded by group members in Part 1.

In each round, the top two performers from Part 1 will each have 80 tokens.

In each round, the bottom two performers from Part 1 will each have 20 tokens.

ELSE IF LUCK TREATMENT:

The token endowment for all participants is not the same.

The token endowment has been determined on the basis of random chance.

In each round, two randomly chosen group members will each have 80 tokens.

In each round, the other two group members will each have 20 tokens.

ELSE IF UNCERTAIN TREATMENT:

The token endowment for all participants is not the same.

The token endowment has been determined in the following way.

There is a 50% chance that the token endowment has been determined by the number of words correctly encoded by group members in Part 1. In each round, the top two performers from Part 1 will each have 80 tokens and the bottom two performers from Part 1 will each have 20 tokens.

There is a 50% chance that the token endowment has been determined on the basis of random chance. In each round, two randomly chosen group members will each have 80 tokens and the other two group members will each have 20 tokens.

You will not know which of the above rules has been used to determine token endowments in your group.

END IF

Your token endowment for each round of the game is: **{{endowment}}**

<u>Part 2 - Round X - Vote</u>

You have a token endowment of **{{endowment}}**.

IF EQUALITY TREATMENT:

Remember, in each round, each group member has an endowment of 50 tokens.

ELSE IF MERIT TREATMENT:

Remember, in each round, the top two performers from Part 1 each have an endowment of 80 tokens, and the bottom two performers from Part 1 each have 20 tokens.

ELSE IF LUCK TREATMENT:

Remember, in each round, two randomly chosen group members each have an endowment of 80 tokens, and the other two group members each have 20 tokens.

ELSE IF UNCERTAIN TREATMENT:

Remember, there is a 50% chance that in each round, the top two performers from Part 1 each have an endowment of 80 tokens, and the bottom two performers from Part 1 each have 20 tokens. Alternatively, there is a 50% chance that in each round, two randomly chosen group members each have an endowment of 80 tokens and the other

two group members each have 20 tokens. You do not know which of the above rules has been used to determine token endowments in your group.

END IF

Which fine rate do you vote for in this round? 0% / 30% / 60% / 80%

The tokens you keep in the private account will be subject to the fine rate chosen by the majority.

NOTE: If you do not make your decision within the time limit, then your vote will be set to 0% fine rate.

<u>Part 2 - Round X - Allocate</u>

You have a token endowment of **{{endowment}}**.

IF EQUALITY TREATMENT:

Remember, in each round, each group member has an endowment of 50 tokens.

ELSE IF MERIT TREATMENT:

Remember, in each round, the top two performers from Part 1 each have an endowment of 80 tokens, and the bottom two performers from Part 1 each have 20 tokens.

ELSE IF LUCK TREATMENT:

Remember, in each round, two randomly chosen group members each have an endowment of 80 tokens, and the other two group members each have 20 tokens.

ELSE IF UNCERTAIN TREATMENT:

Remember, there is a 50% chance that in each round, the top two performers from Part 1 each have an endowment of 80 tokens, and the bottom two performers from Part 1 each have 20 tokens. Alternatively, there is a 50% chance that in each round, two randomly chosen group members each have an endowment of 80 tokens and the other two group members each have 20 tokens. You do not know which of the above rules has been used to determine token endowments in your group.

END IF

The majority of the group voted for a fine rate of **{{fine_rate}}**.

How many tokens would you like to allocate to the group account in this round? 0–{{endowment}} tokens

The tokens you keep in the private account will be subject to the fine rate chosen by the majority.

NOTE: If you do not make your decision within the time limit, then your entire token endowment will be allocated to the group account.

Part 2 - Round X - Results

- Fine rate voted by majority: **{{fine_rate}}**
- Tokens in private account: **{{kept}}**
- Total tokens in group account: **{{group_account}}**
- Tokens you got back from the group account: **{{back}}**
- Fine = {{fine_rate}} of {{kept}} = **{{fined}}**
- Fixed fee = **{{fixed}}**

Earnings for this round = {{back}} (from group account) + {{kept}} (kept in private account) - {{fined}} (fined) - {{fixed}} (fixed fee) = **{{earn}}**

Part 2 - End of Game

IF TIMEOUT:

Unfortunately, you were kicked out of the game because one or more of the participants in your group were responding too slowly. We saved your earnings up to the point where you were kicked out of the game.

ELSE:

This is the end of the game.

END IF

You encoded {{encoded}} words correctly in Part 1 and earned {{part1earn}} in Part 1.

You earned {{part2earn}} in Part 2.

You also earned {{ waitEarn }} for waiting to be matched.

This means you will receive **{{bonus}}** as the bonus payment in this study, resulting in a total overall payment of **{{total}}**.

Part 3 – Exit survey

Before completing the study, please answer a few demographic questions and questions about the game. You do not have to answer the demographic questions if you would prefer not to.

What is your gender? Male / Female / Other

What is your age? 18-100

What is your ethnic group? Asian or Asian British / Black or Black British / Mixed / White / Other ethnic group

Are you currently employed? Yes / No

Are you currently married? Yes / No

Are you religious? Yes / No

What is the highest level of education you have completed? Primary school / GCSEs or equivalent / A-Levels or equivalent / University undergraduate programme / University postgraduate programme / Doctoral degree

Please estimate your annual household income. Below £10,000 / £10,001 - £20,000 / £20,001 - £30,000 / £30,001 - £40,000 / £40,001 - £50,000 / Above £50,000

How fair do you think the distribution of endowments was in your group? Extremely fair / Somewhat fair / Neither fair nor unfair / Somewhat unfair / Extremely unfair

How were endowments distributed in your group? Equally / By performance in Part 1 / By random chance / By performance in Part 1 with 50% chance and by chance with 50% chance

<u>Part 3 – Exit survey</u>

Please also state the extent to which you agree or disagree with the following statements.

- If you are reading this question, please answer with 'Strongly agree'
- It is okay if some groups have more of a chance in life than others.
- Inferior groups should stay in their place.
- To get ahead in life, it is sometimes okay to step on other groups.
- We should have increased social equality.
- It would be good if groups could be equal.
- We should do what we can to equalise conditions for different groups.
- It is always better to trust the judgment of the proper authorities in government and religion than to listen to the noisy rabble-rousers in our society who are trying to create doubt in people's minds.
- It would be best for everyone if the proper authorities censored magazines so that people could not get their hands on trashy and disgusting material.
- Our country will be destroyed some day if we do not smash the perversions eating away at our moral fibre and traditional beliefs.
- People should pay less attention to The Bible and other traditional forms of religious guidance, and instead develop their own personal standards of what is moral and immoral.
- Atheists and others who have rebelled against established religions are no doubt every bit as good and virtuous as those who attend church regularly.
- Some of the best people in our country are those who are challenging our government, criticising religion, and ignoring the 'normal way' things are supposed to be done.

<u>Part 3 – Exit survey</u>

Finally, do you have any feedback about the study?

Thank you!

This is the end of the study.

Thank you for your participation.

You will receive £5 soon for your participation. Any additional bonus payment will be paid to you within 72 hours once we have finished calculating the payment to all the participants.

Please click here to finish the study and return to Prolific.

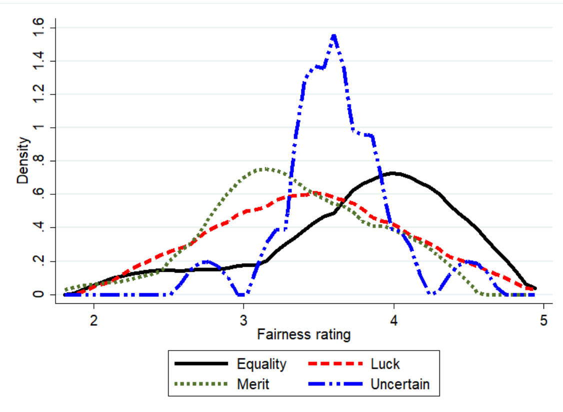
Appendix B: Exploring ex-post fairness perceptions

As mentioned in the experimental design section, all participants were asked to report their perceived level of fairness (on a 5-point Likert type scale: Scale: 1 = Very unfair, to 5 = Very fair) of the source of their endowments. While an admittedly crude proxy for a true measure of fairness perceptions[13], we examine if our self-reported fairness measure potentially correlates with observed behaviour in the prior public goods game in the different treatments. In particular, we compare the fairness ratings across treatments to see if participants perceived luck as being more unfair than merit as a source of inequality.

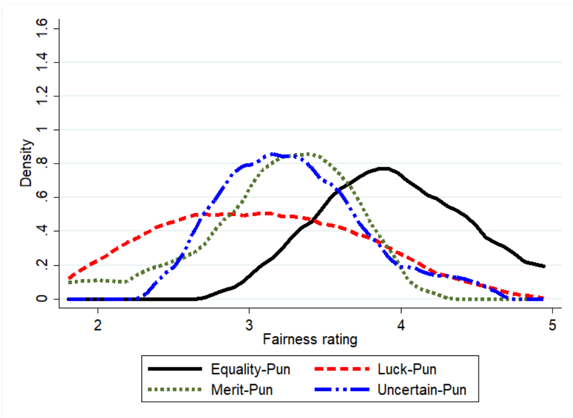
Figure B1 shows the kernel density estimates of the fairness ratings of the four different treatments: equality, luck, merit and uncertain. Table B1 shows a more disaggregated picture and presents average fairness ratings by the rich and poor.

Figure B1. Kernel density estimates of fairness ratings

A. *No-Punishment* (NP) treatments



B. *Punishment* (P) treatments



Scale: 1 = Very unfair, to 5 = Very fair.

[13] Our participants reported their perceived fairness at the end of the experiment. Hence, their perceptions were undoubtedly coloured by their experience in the public goods game. Respondents in earlier studies, who find that luck is a less acceptable than merit as a source of inequality, were made aware of the different sources of inequality and then they ranked the sources. This was not the case in our experiment. Participants were only aware of how endowments were determined in their own treatment and were not aware of the other endowment scenarios. Since they were unable to rank scenarios in terms of fairness, they had to enter a rating with little context. These features make our measure a proxy at best.

Table B1. Average fairness ratings

	No-Punishment treatments				*Punishment* treatments			
	Obs.	Overall	Poor	Rich	Obs.	Overall	Poor	Rich
Equality	64	3.77	-	-	48	3.98	-	-
		(1.05)				(0.81)		
Luck	76	3.43	3.45	3.42	68	3.07	3.18	2.97
		(1.23)	(1.22)	(1.24)		(1.15)	(1.22)	(1.09)
Merit	68	3.33	3.24	3.41	48	3.17	3.08	3.25
		(1.06)	(1.05)	(1.08)		(1.04)	(0.88)	(1.19)
Uncertain	72	3.61	3.56	3.67	48	3.29	3.29	3.29
		(0.89)	(0.94)	(0.86)		(1.17)	(1.19)	(1.16)

Figures in parentheses are standard deviations. The number of poor and rich is half the number of participants in the treatment. Fairness rating scale: 1 = Very unfair, to 5 = Very fair.

First, all three unequal treatments are considered less fair than equality, whether or not there is an opportunity to vote for punishment. For Wilcoxon ranksum (RS) tests, the unit of observation is an independent group, ranging from 12 to 19 groups per treatment. We average fairness ratings across the members of a group to yield one independent observation per group.[14] In the absence of punishments, relative to equality, overall fairness ratings are significantly lower in Merit (RS $p = 0.0148$) and marginally so in Luck (RS $p = 0.0762$), but not in Uncertain (RS $p = 0.1306$). In the presence of punishments, compared to Equality-Pun, overall fairness ratings are significantly lower in Luck-Pun, Merit-Pun and in Uncertain-Pun (RS $p < 0.001$ in all cases).

However, we do not find significant differences between the three unequal treatments, whether punishments are available or not (RS $p > 0.10$ for all pairwise comparisons). We note that while the differences may not be significant, the Luck treatment is nevertheless considered

[14] Fairness ratings were elicited at the end of the experiment after all four group members interacted with each other for 10 rounds in the preceding public goods game. It is more than likely that fairness views were influenced by the experience of the game and, hence, by the behaviour of the other group members. Therefore, we treat a group as an independent observation rather than an individual.

more unfair than Merit or Uncertain, with greater probability mass in the left-tail of the distribution (Figure B1), particularly in the treatment with punishment.[15]

[15] Nockur et al. (2021), who also elicit fairness ratings at the end of the interaction like us, find no significant differences in fairness ratings between the rich and the poor in their luck-based treatment.

Appendix C: Power analysis

Power analysis in studies like this is not straight-forward owing to the fact that there are multiple different hypotheses being considered. Furthermore, given repeated play and partners matching, the only independent observation is the group rather than the individual.

Ex ante power analysis: Using a Related Study to Determine Sample Size Required

We conducted two separate sets of power analyses: (i) for differences in choice of punishment rate between the rich and the poor and (ii) for differences in contributions.

In order to undertake power calculations, we looked at the data from a pilot study that also implemented endowment heterogeneity and centralised punishment along similar lines to the current study except in that study the information regarding the source of inequality was presented differently from the current study. In the sense that we made the source of inequality more salient in the current study while this was not emphasized in the pilot study. We use the averages from that study to determine the sample size required for statistical power (80% chance of not committing a Type-II error by mistakenly failing to reject a false null hypothesis).

We examine two key predictions comparing the rich and poor in this ex-ante power analysis: (i) the rich contribute a lower percentage of their endowment to the public good than the poor, and (ii) the rich vote for lower fine rates than the poor.

Pooling across all the inequality treatments in the related study, Table C1 reports the means, standard deviations, and correlations between rich and poor percentage contributions and fine-rate votes.

Table C1 Pilot summary statistics of rich and poor (pooling across inequality treatments)

	Mean (St dev)	
Inequality	% Contributions	Fine-Rate Votes
Poor	75.34% (16.71%), n=43	36.51% (20.93%), n=43
Rich	63.11% (19.14%), n=43	28.48% (17.51%), n=43
	Correlation	
Poor & Rich	-0.3256	-0.2061

Based on the earlier data, the sample needed in our main study to find similar differences in percentage contributions and fine rate votes significant at the 5% level for 80% statistical power are, respectively, 47 independent pairs and 111 independent pairs. If we use a one-sided test, which is justifiable since we have clear predictions that the rich will contribute a smaller proportional amount compared to the poor and that the poor will choose higher fines, the required sample size for statistical power to detect treatment differences in percentage contributions and fine-rate votes would be 37 independent pairs and 88 independent pairs.

Most power calculations are conducted based on an individual being the unit of observation. This is not true in studies like ours where an independent observation is a group of four rather than an individual participant. This nested structure makes power calculations less than straightforward. We therefore do not account for group-level interdependence in our power calculations. Above we note what sample size is suggested by usual power calculations based on differences in means using individuals as the unit of observation. To account for the nested structure in our analysis, we included random effects for groups as well as individuals as necessary in our formal econometric analysis.

In the main study, pooling across inequality treatments, we have 380 subjects (95 groups) – 216 subjects (54 groups) without punishment and 164 subjects (41 groups) with punishment. All 380 subjects (95 groups) are part of the relevant sample size for comparing differences in percentage contribution between rich and poor. The relevant sample size for comparing differences in fine rate votes between rich and poor is 164 subjects (41 groups).

Note that the sample size for fine-rate votes is lower than what is required in the power analysis. Given our budget and the power analysis, we aimed to recruit 400 subjects, including subjects in equality treatment in each of the two conditions, one with and one without punishments. In the event, we recruited 807 subjects but after dropouts we ended up with 492 participants in all: 280 in the no punishment condition and 212 in the punishment condition. We provide a detailed analysis of dropouts in Appendix D. However, as noted below for Result 2, in the data set used in the main paper, we find a larger difference between rich and poor in their voting behaviour, leading to better implied power.

Ex-post power analysis: Implied power of hypothesis tests based on actual results

For each result, the following analysis reports the implied power of test of means. Power analysis for tests across treatments use a two-sample t-test. Tests within a treatment (i.e., comparing the behaviour of rich and poor) use a paired-means t-test.

Result 1: Punishment Voting (a group of four constitutes one independent observation)

Luck Poor vs. Rich: power > 0.9999, n = 17
Merit Poor vs. Rich: power > 0.9999, n = 12
Uncertainty Poor vs. Rich: power = 0.8683, n = 12

Result 2: Contributions by Punishment Rates (a group of four constitutes one independent observation)

Non-Deterrent Punishment

Luck Poor vs. Rich: power > 0.9999, n = 17
Merit Poor vs. Rich: power > 0.9999, n = 12
Uncertainty Poor vs. Rich: power > 0.9999, n = 11

Deterrent Punishment

Luck Poor vs. Rich: power = 0.8689, n = 13
Merit Poor vs. Rich: power = 0.2679, n = 8

Uncertainty Poor vs. Rich: power = 0.4337, n = 7

No Punishment Treatments

Luck-NoPunish vs. Equality-NoPunish: power = 0.4178, n = 140
Merit-NoPunish vs. Equality-NoPunish: power = 0.6618, n = 132
Uncertainty-NoPunish vs. Equality-NoPunish: power = 0.1569, n = 136

Punishment Treatments
Luck-Punish vs. Equality-Punish: power = 0.9985, n = 116
Merit-Punish vs. Equality-Punish: power = 0.9963, n = 96
Uncertainty-Punish vs. Equality-Punish: power = 0.9552, n = 96

Note that the sample sizes are different in the deterrent and non-deterrent analysis of contributions by punishment rates. Some groups never voted for deterrent or non-deterrent punishment rates.

Appendix D: Attrition analysis

The experiments described in this study are elaborate by the standards of online experiments, requiring more time than average and involving complex interactions in groups of four, particularly in the treatment where participants first voted for punishment rates prior to making contribution decisions. Ex ante, this made it difficult to know what drop-out rates to expect. In the end, 280 out of 407 (212 out of 400) who started the study completed the study as whole four-person groups in the No-Punishment (Punishment) treatments.

The dropouts happened for a number of reasons. Some had internet connection problems. Some people arrived at the waiting room after completing the word coding task but had to wait for a while prior to being assigned to a group. Some of them chose to quit rather than wait. Some left in the middle of the session, possibly because getting all four group members' decisions entered took time, especially in the Punishment treatments. The longer duration in the Punishment treatments resulted in more dropouts than in the No-Punishment treatments. There is no systematic pattern that we can discern in these dropouts. In particular, it is not the case that those who dropped out were earning more or less than others.

A total of 807 participants started the study. Recall that participants were not matched into groups until after Part 1 (coding task) and starting Part 2. In Part 2, participants read the instructions and answered general review questions about the public goods game before being matched into groups.

One participant dropped out during Part 1.

Twenty-three participants dropped out while waiting to be matched in Part 2.

If, after being matched in Part 2, participants' groups timed out due to another member of their group not entering a decision in two consecutive rounds, active group members could still complete the survey questions in Part 3 and be paid for the parts of the experiment they completed. Eleven participants never completed Part 3 after their groups timed out.

For a group to be dropped in Part 2, a member must time out two periods in a row. It is not possible, then, for groups to drop out after round 1. Two hundred and eighty participants dropped out after being matched and starting Part 2. Table D1 reports the number of individuals, by round, who dropped out. Table D2 reports the number of individuals, by treatment, who dropped out. Dropouts were higher in Punishment treatments. This occurs primarily in the first two rounds when dropping out is possible (2 & 3).

Table D1. Dropouts by round in Part 2

Period	No Punishment	Punishment	Total
2	3	80	83
3	12	68	80
4	16	14	30
5	6	19	25
6	16	8	24
7	11	4	15
8	8	8	16
9	0	7	7
10	0	0	0
Total	72	208	280

Table D2. Dropouts by treatment in Part 2

Treatment	No Punishment	Punishment	Total
Equality	22	50	72
Luck	15	42	57
Merit	22	61	83
Uncertainty	13	55	68
Total	72	208	280

Appendix E: Additional analyses in the punishment treatments

Table E1. Individual contributions in the Punishment treatments: all combined

In Table 7 in the main text, we presented regression results separately for different treatments, each of which utilizes a sub-sample of our data. Here, we present estimates of individual-level panel random effects regressions of proportional contributions that use all the data. All punishment treatments are included as dummy variables with the equality treatment being the reference category. This table reinforces the insights as those obtained from Table 7.

Individual Panel RE regressions	(1)	(2)
Merit-Pun	13.87**	12.06*
	(6.484)	(7.139)
Uncertain-Pun	18.58***	17.37***
	(6.323)	(6.425)
Round	-0.764*	-0.0.758*
	(0.443)	(0.441)
Lagged percent cont.	0.0877*	0.0869*
of others in group	(0.051)	(0.050)
Rich (End = 80) dummy	-5.551	-7.865
	(5.361)	(6.422)
Merit-Pun × Rich	-12.82	-9.842
	(8.485)	(9.339)
Uncertain-Pun × Rich	-13.63*	-11.68
	(7.863)	(9.085)
Deterrent fine dummy	16.38***	13.48**
	(5.345)	(7.225)
Merit-Pun × Det. fine	-3.504	-0.331
	(7.420)	(11.18)
Uncertain-Pun × Det. fine	-10.15	-7.791
	(7.056)	(8.168)
Luck-Pun × Det. fine × End=80	-	6.876
		(8.411)
Merit-Pun × Det. fine × End=80	-	-1.735
		(9.423)
Uncertain-Pun × Det. fine ×	-	1.138*
End=80		(4.786)
Constant	61.10***	62.51***
	(5.518)	(5.776)
Observations	1908	1908

Panel RE regressions. Std. errors clustered on group in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$