Invisible inequality leads to
punishing the poor and rewarding the rich

Oliver P. Hauser1*, Gordon T. Kraft-Todd2, David G. Rand2,3,4, Martin A. Nowak1, Michael I. Norton5

Abstract. How does lack of awareness of income inequality affect behaviour towards the rich and poor? To address this question, we assigned participants either at random or based on merit to one of five income levels (reflective of the U.S. income distribution), who then played a repeated public goods game with punishment, reward or both. When participants did not know the income distribution, they punished the poor and rewarded the rich. When incomes were known, conversely, participants punished the rich and rewarded the poor. What explains this reversal? When incomes were hidden, participants punished those who gave the least in absolute amounts (the poor) and rewarded those who gave the most in absolute amounts (the rich). When incomes were revealed, however, participants were sensitive to people’s ability to contribute: they punished those who contributed the smallest percentage of their income (the rich) and rewarded those who contributed the highest percentage (the poor). Moreover, revealing incomes led to greater overall contributions to the public good, suggesting that awareness of inequality may offer societal benefits.

Keywords: cooperation; inequality; punishment and reward; hidden income; revealed income

1 Program for Evolutionary Dynamics, Harvard University, Cambridge, MA 02138
2 Department of Psychology, Yale University, New Haven, CT 06511
3 Department of Economics, Yale University, New Haven, CT 06511
4 School of Management, Yale University, New Haven, CT 06511
5 Marketing Unit, Harvard Business School, Boston, MA 02163
* Corresponding author: ohauser@fas.harvard.edu
Introduction

Preferences for spending on public goods such as social programmes or the health care system are based, at least in part, on beliefs about income and wealth inequality (Alesina & Angeletos, 2005; Charité, Fisman, & Kuziemko, 2015; Durante, Putterman, & van der Weele, 2014; Kuziemko, Norton, Saez, & Stantcheva, 2015). Yet recent cross-cultural evidence suggests that across the globe, many people are unaware of the true extent of inequality in their country (Cruces, Perez-Truglia, & Tetaz, 2013; Davidai & Gilovich, 2015; Kiatponsan & Norton, 2014; Norton & Ariely, 2011).

Here, we explore the implications of this lack of awareness of inequality, examining how hidden inequality affects societal outcomes and people’s behaviours towards the rich and the poor (relative to when inequality is revealed). We hypothesised that revealing information on inequality might exert a substantial effect on behaviour: if people do not realise how little the poor have, and how much the rich have, they may be less sympathetic to low contributions from those who cannot afford to give more, and less likely to punish the rich for not contributing their “fair share.”

To explore these predictions experimentally, we used a standard paradigm in experimental economics; an incentive-compatible, repeated public goods game (PGG) in groups of 5 players. In each of 10 rounds, every player was assigned an “income” and chose how much of that income to contribute to a common pool; all contributions were doubled and divided equally among the five players (see Supplemental Online Material (SOM) Section 2 for more details about experiment design). We then showed each player the contributions of all other players (player IDs were shuffled every round), and gave each
player the opportunity for costly sanctioning of all other players. In the *punishment* condition, participants could pay 1 unit to decrease any other participant’s payoff by 3 units; in the *reward* condition, participants could pay 1 unit to increase any other participant’s payoff by 3 units. Participants could spend up to 2 units on each other participant.

Such sanctioning schemes have been widely used in previous work on cooperation (Fehr & Gächter, 2002; Hauser, Hendriks, Rand, & Nowak, 2016; Rand, Dreber, Ellingsen, Fudenberg, & Nowak, 2009; Sefton, Shupp, & Walker, 2007; Sutter, Haigner, & Kocher, 2010). In many Western societies, results typically reveal that low contributors are punished, while high contributors are rewarded; anti-social punishment aimed at high contributors, on the other hand, is frequently observed in countries with a weak rule of law (Herrmann, Thöni, & Gächter, 2008). In these studies, all players typically receive identical endowments in each round, and this equality is common knowledge to all players; thus the majority of these experiments, while highly informative regarding the maintenance of cooperation, shed little light on perceptions of, and reactions to, inequality.

Recently, however, scholars have been investigating inequality in the provisioning of public goods. In these experiments, richer participants typically contribute less of their income than poorer ones, leading to reduced social efficiency (Buckley & Croson, 2006; Isaac & Walker, 1988; Keser, Markstädt, & Schmidt, 2014a; Keser, Markstädt, Schmidt, & Schnitzler, 2014b). When punishment is introduced, participants punish others to express their preference for efficiency norms (everyone contributing their entire income) or relative contribution norms (everyone contributing the same relative share of their income), even when such behaviour led to a reduction in average net earnings (Antinyan,
Moreover, uncertainty around others’ actual contributions or incomes also elicits greater punishment (Ambrus & Ben Greiner, 2012; Bornstein & Weisel, 2010). (See SOM Section 1.2 for details on previous literature.)

Building on this previous research, we introduce three novel features to explore the impact of people’s recently demonstrated lack of knowledge regarding the distribution of income (Cruces et al., 2013; Kiatponsan & Norton, 2014): (i) we experimentally vary whether the income distribution, as well as each participant’s specific income, is hidden or revealed to explore the causal effect of knowledge of inequality on behaviour toward the rich and poor; (ii) we use an income distribution that is extremely unequal (the actual United States distribution) to explore behaviour toward the rich and poor under conditions reflective of real-world inequality; and (iii) we allow participants to either punish, reward, or both punish and reward the poor and the rich to explore how these sanctions are utilised to address perceived inequity. Indeed, we expected sanctions to be crucial in addressing inequality and net payoffs. Absent sanctions, visibility of wealth inequality can actually increase inequality in social networks (Nishi, Shirado, Rand, & Christakis, 2015). But in the presence of sanctions, we predicted that revealing incomes would be a solution, not an obstacle: when incomes are revealed, the ability to reward and punish allows participants to shape others’ future contributions and restore equity.

In experiment 1, we randomly assign participants to different incomes – reflective of the income distribution of the United States – and then further randomly assign them to either be aware of the full distribution of incomes (the revealed condition) or to be unaware
(the hidden condition); we examine sanctioning behaviour – both punishment and reward – towards other players in the game. In experiment 2, we examine whether merit might play a moderating role: whereas in experiment 1 incomes are assigned randomly, in experiment 2 incomes are assigned based on task performance (see Cappelen, Konow, Sørensen, & Tungodden, 2013). We again assign participants to either the revealed or hidden condition, and examine their decisions to punish and reward other players.

**Experiment 1**

**Methods**

Participants (N = 855) were recruited on Amazon Mechanical Turk, read the instructions and answered comprehension questions, and were assigned to groups of five to play a two-stage economic game over 10 rounds. Before the start of the game, participants were randomly assigned an income level. Across both experiments, we used the United States pre-tax incomes by quintile to create incomes for the five players. In our first experiment, the top quintile participant received 55 units out of 100 units in the group (or 55% of all income), the next 19 units, the next 13 units, the next 9 units, and the bottom quintile participant 4 units (Fig. 1a) (Congressional Budget Office, 2007). Once assigned to an income level, participants received the same income each round for 10 rounds.

Across all conditions, participants played two stages in each round. In Stage 1, participants could contribute any amount of their income to a common project. Any units contributed were doubled and split equally among all five group members. In Stage 2, participants could see everyone’s contributions and could either punish or reward their group members (depending on the condition). Participants could not spend more in Stage
2 than they had earned in Stage 1. At the end of each round, participants saw their group members’ decision to reward or punish them in Stage 2, and a summary of their payoff in this round. To ensure that participants could not identify one another across multiple rounds (and to avoid retaliation (Nikiforakis et al., 2012)), each player was only known by a series of random letters that changed at the beginning of every round.

![Figure 1. The income distribution in our game and the main experimental manipulation between the hidden and revealed conditions. A Each player in a group of five was randomly assigned to a position in an income distribution. In the first experiment, we used the 2007 U.S. pre-tax income distribution (Congressional Budget Office, 2007): in each of ten rounds, the top quintile participant received 55 units, while the bottom quintile player received 4 units. B When making decisions to punish and reward, participants in the hidden condition saw their own income and the sum of all incomes. C In the revealed condition, participants viewed all players’ incomes.](image)

The experimental design was a 2 (punishment versus reward) X 2 (hidden versus revealed) between-participants design ($N = 600$). In the hidden condition, players had no information about the incomes of the others in their group (Fig. 1b): they made
contributions, viewed others’ contributions, and decided to punish or reward based only on the total amounts contributed by other players. In the revealed condition, in contrast, participants were shown the income of each player as they made their decisions to punish or reward – allowing them to base their decisions not only on the total amount contributed, but also the percentage of available income that each player chose to contribute (Fig. 1c). For example, a player who contributed just three units in the hidden condition may look stingy; learning that this player had only four total units in the revealed condition may dramatically alter perceptions of their contribution.

We expected that in the hidden condition, participants would generally view the (low total) contributions of bottom quintile players unfavourably, inducing punishment, and the (high total) contribution of the top quintile players favourably, inducing reward. In contrast, we expected that in the revealed condition, participants would generally view the (high percentage) contributions of bottom quintile players favourably, inducing reward, and the (low percentage) contribution of the top quintile players unfavourably, inducing punishment.

For details about the experimental design of experiment 1, see SOM Section 2.2.1.

**Results**

We find that, indeed, participants in the hidden condition rewarded richer participants more (coeff = 0.636, \( p < 0.001 \)), whereas those in the revealed condition rewarded poorer participants more (coeff = -0.720, \( p < 0.001 \); interaction between income and revealed dummy, coeff = -1.356, \( p < 0.001 \); Figure 2 and Table S1). We observe a mirror image of these results for decisions to punish: participants in the hidden condition punished poorer
participants more (coeff = -0.282, p = 0.042), whereas those in the *revealed* condition punished richer subjects more (coeff = 0.692, p < 0.001; interaction between income and *revealed* dummy, coeff = 0.974, p < 0.001; Figure 2 and Table S3). Thus, knowledge about economic inequality had a profound effect on sanctioning. (Unless otherwise noted, all statistical analyses use linear regression with standard errors clustered on group, taking income quintile as a continuous predictor variable; note that using log-transformed income instead of income quintile as a predictor generates qualitatively equivalent results, see SOM Section 2.)

![Figure 2](image.png)

*Figure 2. Amount of received reward (top panels) and punishment (bottom panels) depends on income quintile and whether income was hidden (left panels) or revealed.*
Participants rewarded higher income participants more in the hidden condition, but less in the revealed condition. Punishment behaviour is a mirror image of reward: participants punished poorer participants more in the hidden condition, while punishing richer participants more in the revealed condition.

Why did players sanction so differently in the hidden and revealed conditions? Across both conditions, richer players contributed larger total amounts (hidden: coeff = 3.172, \(p < 0.001\); revealed: coeff = 4.734, \(p < 0.001\); Table S5), but lower percentages of their income (hidden: coeff = -0.098, \(p < 0.001\); revealed: coeff = -0.058, \(p < 0.001\); Table S6) (Figure 3). Collapsing across conditions, top quintile participants contributed 20.49 out of 55 units (or 37% of their income) whereas bottom quintile participants contributed 2.83 out of 4 units (or 71% of their income). The pattern of sanctioning we observe therefore follows naturally if sanctions were assigned based on percentage of income contributed in the revealed condition but total amount contributed in the hidden condition.

**Figure 3. Who contributes more?** A In the hidden condition, only absolute contributions could be assessed, such that richer participants appeared to contribute more. B In the
revealed condition, where participants could view contributions relative to income, it was clear that lower income participants contributed a larger fraction of their income.

Supporting this logic, in the revealed condition, participants conditioned their sanctioning decisions on the percentage of the target’s income that was contributed (using percentage contributed as the independent variable; predicting punishment, coeff = -4.664, $p < 0.001$, Figure 4a; predicting reward, coeff = 6.320, $p < 0.001$, Figure 4b; Table S10), more so than on the absolute amount contributed (if anything, absolute contribution predicts the opposite direction from the overall observed pattern: using log-transformed total contribution as the independent variable; predicting punishment, higher absolute contributors are punished more, coeff = 1.365, $p = 0.003$; predicting reward, higher absolute contributors are rewarded marginally less, coeff = -0.980, $p = 0.080$; Table S10).

In fact, in the revealed condition, participants not only rewarded and punished based on percentage contributed; they also targeted the richest players in particular, even when they contributed the same relative amount of their income (using percentage contributed and income as the independent variables; predicting punishment, coeff on quintile = 0.563, $p < 0.001$; predicting reward, coeff on quintile = -0.309, $p = 0.031$; Tables S11 and S12). Thus, sanctions were used partly to encourage future contributions from the rich and partly to reduce the wealth gap in the group.

In the hidden condition, conversely, where only total contribution amounts were known, sanctioning was based on total amount contributed (using log-transformed total contribution as the independent variable; predicting punishment, coeff = -1.863, $p = 0.019$, Figure 4c; predicting reward, coeff = 4.700, $p < 0.001$, Figure 4d; Table S9), but not on
percentage of income contributed (using percentage contributed as the independent variable; predicting punishment, coeff = 0.030, $p = 0.954$; predicting reward, coeff = -0.216, $p = 0.677$; Table S9).

Figure 4. Received punishment and reward depends on percentage of income contributed in the revealed condition (top panels) and on absolute income contributed in the hidden condition (bottom panels). A,B When incomes were revealed, participants who contributed a higher percentage of their income were A punished less and B rewarded more. C,D Conversely, when incomes were hidden, participants who contributed a higher absolute amount were C punished less and D rewarded more. Bubble size is proportional to the fraction of corresponding participants.
We next consider the consequences of income transparency on total contributions and final payoff inequality. Overall, significantly more units were contributed in the revealed condition compared to the hidden condition (coeff = 1.745, \( p = 0.002 \); Table S15). But hiding income not only led to lower total payoffs; it also resulted in greater inequality. The Gini coefficient—a common summary measure of inequality—after the final round of the game was significantly higher in the hidden condition (average 0.238) compared to revealed (average 0.169; Rank-sum, \( p < 0.001 \)). Participants in the bottom (poorest) through fourth (second richest) quintiles maintained (or even increased) their contribution levels over the ten rounds in both the hidden and revealed conditions (no significant change in contribution over round, \( p = 1.00 \) bonferroni-corrected for all quintile-condition pairs, with the exception of the poorest quintile in the hidden condition, who actually increased over time: coeff = 0.047, \( p = 0.040 \) corrected; Tables S17 and S18). However, although participants in the top (richest) quintile in the revealed condition also continued to contribute over time (coeff = -0.382, \( p = 1.000 \) corrected), top quintile players in the hidden condition decreased their contributions over the ten rounds (coeff = -1.077, \( p < 0.001 \) corrected) (Tables S17 and S18). Thus, in the hidden condition, sanctions were less effective at maintaining contributions among those with the greatest ability to contribute to the public good.

**Experiment 2**

Participants in our initial experiment were assigned their income randomly. However, incomes in the real world are not just the product of chance, but also of effort. Earned
incomes could justify inequalities and thus reduce a desire for redistribution through punishment or reward (Cappelen et al., 2013). In a second experiment, we thus assigned income based on participants’ performance in an individual effort task before playing the public goods game. We made several additional changes to the design of the original experiment to extend the external validity of our results, which are described below.

**Methods**

Participants \( (N = 440) \) who were recruited on Amazon Mechanical Turk played a two-phase experiment. In Phase 1, participants completed an individual effort task that affected their income level in the second phase. In Phase 2, participants played a repeated two-stage economic game with sanctions. In the economic game, participants decided how much to contribute to a common pool in Stage 1 and then whether to punish or reward a group member in Stage 2. At the beginning of each round, a series of random letters was generated and assigned to each participant as their player ID for the round.

As with our first experiment, participants were assigned to one of two conditions: participants in the *revealed* conditions saw their own income, the income of the other participants and the sum of all incomes in both Stages 1 and 2, while participants in the *hidden* condition only saw their own income and the sum of all incomes in the group.

The second experiment differed in several ways from the first experiment. First, participants were not randomly assigned their income at the start of the economic game, but earned their position in the income distribution beforehand through an effort task (Abeler, Falk, Goette, & Huffman, 2011). All participants were told that their effort in the first phase of the experiment would determine their earnings potential in the second phase.
of the experiment, without disclosing additional information about Phase 2. Once they had completed the effort task, they were assigned their income based on the rank of performance in their group. The best-performing participant in a group earned the highest income, the second-best performing participant earned the second-highest income, and so on.

Second, participants in the hidden condition in our first experiment did not know anything about the income distribution except that incomes between players were different. To increase external validity, we therefore told participants in our second experiment (who had been recruited exclusively from the U.S.) in both the hidden and revealed conditions in the second experiment that the income distribution used in the game was derived from the United States income distribution by quintile (U.S. Census Bureau, 2013). Thus, if participants had an accurate estimate of U.S. inequality, they might be less likely to ‘punish the poor’ in the hidden condition.

Third, we gave participants a wider range of simultaneous actions towards other participants in the second experiment: they could choose to reward or punish different participants in each round, paying 2 units to increase or decrease another participant’s payoff by 6 units, respectively. Thus, by making both sanctioning options available simultaneously, a participant could reward someone they thought contributed sufficiently and punish someone else who they thought did not contribute enough.

For more details on experiment 2, see SOM Section 3.2.

**Results**
In our second experiment, we found qualitatively similar results as in experiment 1: participants continued to reward the rich and punish the poor in the *hidden* condition (predicting number of units received, where positive values imply receiving on average reward and negative values punishment: coeff = 0.053, \(p = 0.042\)), while this trend reversed completely in the *revealed* condition (coeff = -0.171, \(p < 0.001\); interaction between income and *revealed* dummy: coeff = -0.225, \(p < 0.001\); Table S20). Across conditions, richer participants contributed more in absolute terms (coeff = 3.979, \(p < 0.001\); Table S22), but less as a percentage of their income (coeff = -0.078, \(p < 0.001\); Table S23), than poorer participants.

This sanctioning pattern was linked to reward and punishment decisions: higher absolute contributions received more reward in the *hidden* condition (coeff = 0.571, \(p < 0.001\); Table S26), but higher percentage of income contributed was more rewarded in the *revealed* condition (coeff = 1.775, \(p < 0.001\); Table S26). Holding income contributed constant, we found that participants in the *revealed* condition give fewer units to richer participants (predicting number of units received, controlling for percentage of income contributed: coeff = 1.291, \(p < 0.001\); Table S27), suggesting that they desire more equality in the group even when the rich give the same fraction of their income.

Groups in the *revealed* condition (average Gini index = 0.124) indeed ended up with more equal payoffs than those in the *hidden* condition (Gini 0.255). Furthermore, overall contributions to the public good were higher in the *revealed* condition than in the *hidden* condition (coeff = 3.134, \(p < 0.001\); Table S28). All but the richest participants did not decrease their contributions over time in either condition (all \(p > 0.05\) bonferroni-corrected for participants in quintiles 1 through 4; Tables S30 and S31). The richest
participant, on the other hand, marginally decreased their contributions in the hidden condition (coeff = -1.105, p = 0.080 corrected; Table S31), but did not reduce their contributions in the revealed condition (coeff = -0.056, p = 1.000 corrected; Table S30).

In our second experiment, even when incomes were earned and when participants were informed that the income distribution was reflective of their own country’s distribution, participants continued to punish the poor and reward the rich when the income distribution was hidden, but reward the poor and punish the rich when incomes were revealed.

Discussion

In sum, revealing inequality had substantial effects on people’s decisions to reward or punish others, and on total contributions to the public good. Participants were more likely to punish poorer participants and reward richer participants when inequality was hidden; when income was revealed, participants became more sensitive to people’s ability to contribute – leading them to punish the rich and reward the poor.

Participants’ motivation to punish the rich and reward the poor in the revealed condition could arise from seeking equity in contributions or simply aiming to reduce inequality in the group, or both (Bolton & Ockenfels, 2000; Charness & Rabin, 2002; Fehr & Schmidt, 1999; Rawls, 1971). Though equity concerns were present in our sample, participants in the revealed condition punished the top quintile more than any other player, even when the rich had contributed the same percentage of their income. In other words, when inequality was revealed, participants in our experiments desired not just equity but that the wealthy were slightly less well off.
The observation that our participants were unaccepting of inequality adds to a growing literature on social preferences, egalitarianism and libertarianism (Cappelen, Hole, Sørensen, & Tungodden, 2007; Konow, 2000). Our participants’ unwillingness to tolerate inequality of income persisted even when incomes were assigned by performance. Plausible justifications typically make people more accepting of inequalities (Cappelen et al., 2013), though most previous research has focused on moderate, not extremely high, levels of inequality. Future research is needed to investigate to what extent sanctioning behaviour in our sample can be explained by, for example, political ideology, egalitarian values (Piff, Kraus, Côté, Cheng, & Keltner, 2010), compassion (van Kleef et al., 2008), sense of control (Kraus, Piff, & Keltner, 2009), or social dominance (Sidanius & Pratto, 2001).

While our income distribution was drawn from the real world, our paradigm necessarily offers a stylised examination of the impact of inequality on the public good. For example, we restricted the amount that all participants could pay to punish or reward other players to 2 units per player. The real world may not always provide such an upper bound: given their greater resources, the rich have much greater ability to inflict harm or bestow benefits on others. Still, there are some real-world situations in which all decisions count equally: for instance, casting a vote in democratic elections carries equal weight despite differences in income.

Revealing incomes decreased inequality and increased total contributions in our experimental groups. This finding has implications for policymakers concerned with the public good. While revealing all citizens’ incomes may seem challenging to implement—or even hard to imagine—in some countries, it is common practice in others. For example,
the Norwegian government operates an online database which contains detailed information about all citizens’ income, wealth, and tax contributions (Norwegian Tax Administration, 2015). Notably, Norwegians also have high tax morale (I. Lago-Peñas & Lago-Peñas, 2010); while anecdotal, this example suggests that revealing incomes can be associated with increased support for the public good – mirroring our results. In a world of income transparency, the “haves” may become more generous and the “have-nots” less punished, with positive implications for the common good.
References


