MAX PLANCK INSTITUTE
FOR RESEARCH ON
COLLECTIVE GOODS

Stefania Bortolotti
Felix Kölle
Ivan Soraperra
Matthias Sutter

Discussion Paper
2025/15

SOCIAL RISK, FAIRNESS TYPES, AND REDISTRIBUTION

Social Risk, Fairness Types, and Redistribution*

Stefania Bortolotti
University of Bologna & IZA

Felix Kölle
University of Cologne

Ivan Soraperra
Max Planck Institute for Human Development, Center for Human and Machine, Berlin

Matthias Sutter†
Max Planck Institute for Research on Collective Goods, Bonn & University of Cologne & University of Innsbruck

Abstract

Inequality often arises from strategic interactions among individuals. This is so because risky investments can not only be resolved by chance (natural risk), but also by others' actions (social risk). We study how these different sources of inequality shape fairness judgments and the level of redistribution in a controlled experiment with a total of 2,152 participants. We find significantly less inequality acceptance, and thus much more redistribution, under social risk. In addition to the well-known types of Libertarians, Egalitarians and Choice Egalitarians, we identify a novel, hitherto unnoticed, fairness type — Insurers — who always compensate unlucky risk-takers and are especially prevalent when one is let down by others rather than simply unlucky by chance. This suggests that impartial spectators view betrayal as more deserving of support than bad luck. Our findings show that fairness ideals depend jointly on risk-taking and the way in which risk is resolved, either by nature or another human actor, thus highlighting the important role of strategic interaction for fairness types and redistribution.

Keywords: Inequality, fairness views, social risk, redistribution, experiment.

JEL Classification Numbers: C91; D63; D90

*We would like to thank Ingvild Almas, James Konow and Bertil Tungodden, participants at seminars in Munich and Berlin, as well as at the MPI Workshop, and those attending various conferences for their valuable comments. Financial support from the German Research Foundation through Germany's Excellence Strategy – EXC 2126/1–390838866 is gratefully acknowledged.

†Corresponding author: Sutter ✉: Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str, 10, 53113 Bonn, Germany; matthias.sutter@coll.mpg.de, +49 228 91416-150.

1 Introduction

The rise in economic inequality (Chetty et al., 2014; Piketty, 2014) has spurred a growing literature that seeks to explain public support for redistribution and social welfare policies by examining when and why inequality is perceived as (un)fair (e.g., Stantcheva, 2021; Cohn et al., 2023; Fehr et al., 2024; Harrs and Sterba, 2025; Yusof and Sartor, 2025). Rooted in normative theories of justice, this literature focuses on the concept of meritocratic fairness (Konow, 2000; Fong, 2001; Konow, 2003; Alesina and Angeletos, 2005; Cappelen et al., 2007; Almås et al., 2020; Cappelen et al., 2022a; Amasino et al., 2024; Cappelen et al., 2024). A common finding is that individuals are willing to tolerate substantial inequality when it reflects differences in effort, choices or merit, but support more redistribution when outcomes are shaped by circumstances beyond one's control (Cappelen et al., 2022a,b; Bhattacharya and Mollerstrom, 2022; Andre, 2025).[1]

Yet, this literature largely neglects the *strategic context* in which many economically consequential outcomes arise. Inequality often does not stem from isolated choices, but from interactions among individuals in markets and other social settings. Online transactions, business partnerships, delegated investments, and informal lending, to name just a few, frequently involve risks generated by others' actions. Some individuals succeed because they are paired with reliable counterparts; others suffer because they are let down. These dynamics are particularly salient in environments where legal institutions are weak and contracts are incomplete or prohibitively costly to enforce. We refer to such interpersonal sources of inequality as *social risks*, in contrast to *natural risks* — those arising from external factors such as climate events, geopolitical instability, or macroeconomic shocks. In this paper, we examine whether individuals judge inequality differently when it results from social risks rather than from natural risks and whether that leads to different levels of redistribution.

Our study design builds on Cappelen et al. (2013) and related studies, and consists of two experiments: A *stakeholder* experiment and a *spectator* experiment. In the first *stakeholder* experiment, participants make incentivized choices between a safe option and a risky, but potentially more profitable, option. In the *spectator* experiment, participants are matched with a pair of stakeholders and asked to redistribute earnings, if any, between them. Spectators make a series of such redistribution decisions, varying the stakeholders' choices (*safe* vs. *risky*), the outcome of the risky option (*high* vs. *low*), and the probability of receiving the high outcome.

Our design features three between-subjects treatments that exogenously manipulate

[1]Related work has explored the role of efficiency concerns (Almås et al., 2020), beliefs about the origins of inequality (Cappelen et al., 2022a; Harrs and Sterba, 2025; Cappelen et al., 2024), levels of impoverishment (Faravelli, 2007; Erkal et al., 2011; Dohmen et al., 2017), and past experiences with fairness and social status (Bauer et al., 2016; Barr et al., 2016; Roth and Wohlfart, 2018; Cassar and Klein, 2019; Cohn et al., 2023).

two key dimensions: (i) whether risk-taking generates a positive externality for a third party, and (ii) whether risk is resolved by nature or by another person. In our *Control* treatment, risk affects only the stakeholder's own payoff and is resolved by nature (a random computer draw). The *Natural-Risk* treatment introduces a passive third party who benefits from the stakeholder's risk-taking, while risk remains resolved by nature. In *Social-Risk*, the structure is identical to Natural-Risk except that risk is resolved by the third party (who now becomes active) rather than by chance.[2]

We find a strong and significant effect of how risk is resolved on redistribution. In the aggregate, spectators are significantly less likely to accept inequality in the Social-Risk treatment compared to the other two conditions. The Gini coefficient in Social-Risk is between 15% and 20% lower than in Control and Natural-Risk, representing a substantial and economically meaningful reduction in inequality acceptance. This result highlights that impartial spectators are sensitive to the distinction between social and natural risk and, importantly, that they view those who expose themselves to social risk as more deserving of support.

An individual-level analysis reveals the sources for the aggregate results. We first apply a k-means clustering algorithm to identify redistribution patterns. This unsupervised method yields four main clusters, three of which correspond to well-documented fairness types (e.g., Cappelen et al., 2013): *Libertarians*, who never redistribute; *Egalitarians*, who always equalize payoffs; and *Choice Egalitarians*, who redistribute only when stakeholders made the same choices but experienced different outcomes. In addition to these, we identify a novel, fourth type: the *Insurer*. Insurers redistribute only when a stakeholder is "let down" — either by nature or by another player — and ends up with a low outcome. Unlike Choice Egalitarians, Insurers do not condition on ex-ante choices but instead focus on ex-post disadvantages. This new type is especially prevalent in the Social-Risk treatment relative to the Control and Natural-Risk treatments, a finding we confirm using a parametric mixture model that classifies individuals into fairness types.

Importantly, in our mixture model the Insurer type is well defined across all three treatments, since unlucky stakeholders can be compensated even when their choices do not affect (passive or active) third parties and regardless of how risk is resolved. Whether such a type actually emerges in such contexts is ultimately an empirical question. Indeed, our results show that the Insurer type would not have been identified as a distinct type in the cluster analysis had we only considered the Control treatment. This highlights that considering settings with a social component to the resolution of risk is instrumental to uncover the broader universe of fairness types.

[2] This setup resembles a standard trust game in which one player (the trustor) relies on another (the trustee) to act reciprocally. Our focus, however, is not on behavior within the trust game itself, but rather on how uninvolved spectators redistribute money between two players who may have been exposed to betrayal.

Regarding the distribution of fairness types, both the cluster analysis and the mixture models show that in the Control and Natural-Risk treatments, the dominant types are Choice Egalitarians and Libertarians, together accounting for more than 60% of spectators. In the Social-Risk treatment, in contrast, these groups become minorities, while the most common type is the Insurer (32%), followed by Egalitarians (21%). These results indicate that settings characterized by social risk lead significantly more individuals to favor redistribution, particularly when someone has been let down by another person. In a follow-up study, we show that this pattern cannot be explained by uncertainty or ambiguity about the nature of social risk-taking. Instead, we argue it is driven by the perceived violation of fair procedures and by expectations that trust should be reciprocated.

Our paper contributes to the empirical literature on redistribution and fairness ideals in two main ways. First, we extend the relatively limited body of research that investigates the acceptance of inequality when disparities in income and wealth arise from risk-taking, rather than from effort or luck. The seminal work by Cappelen et al. (2013) emphasizes that the decision to engage in risk is intrinsic to most economic transactions and market activities. This creates a tension between two moral viewpoints: the *ex-ante* view, which opposes redistribution of gains and losses resulting from risk-taking when initial opportunities were equal, and the *ex-post* view, which emphasizes final outcomes and considers it fair to mitigate inequalities that arise from such risks. Bortolotti et al. (2025) incorporate concerns about cheating into this framework, finding that suspicion of dishonesty amplifies the polarization of fairness judgments.

We extend this paradigm by incorporating a key feature of many risky choices: a strategic component, i.e., the involvement of a counterpart who can actively influence the final distribution of resources. Our results show that inequality acceptance declines significantly in the presence of *social risk*. Moreover, our individual-level analysis reveals a novel fairness type. This newly discovered type is primarily concerned with outcomes rather than initial opportunities and supports insurance for a specific group — those who accepted a risky option but were ultimately let down by another individual. This finding highlights the sensitivity of fairness ideals to social context, demonstrating that social risk is a far more powerful driver of redistributive preferences than mere externalities.[3] This is also supported by the finding that although redistribution is slightly higher in Natural-Risk than in Control, the difference is small and not statistically significant — suggesting that the presence of externalities alone plays only a minor role in shaping redistribution in our setting.[4]

[3]We use the term redistributive preferences in a broad sense. Much of the literature considers redistribution between players whose initial earnings are either exogenously assigned or stem from choices that do not affect others. By contrast, in our Natural-Risk and Social-Risk treatments, redistribution decisions may also reflect altruistic and welfare concerns. Importantly, however, these concerns are held constant across the two treatments.

[4]While we do not find strong effects of positive externalities from risk-taking, Lobeck and Støstad

Second, we complement the literature on market luck, particularly Yusof and Sartor (2025), who show that spectators are more accepting of inequality arising from market luck than from brute luck, even when both lie beyond workers' control and are unrelated to effort. Unlike in Yusof and Sartor (2025), we introduce meaningful agency on the side of the worker (third party in our terminology), hence making our setting more reflective of real-world contexts, such as business partnerships or informal agreements, where interpersonal trust plays a central role (which is usually not the case in competitive markets). In contrast to the greater tolerance for inequality under market luck, we find lower inequality acceptance when social risk is involved.

In addition, our study contributes to the literature on trust and betrayal aversion — the tendency for individuals to avoid risks when the uncertainty stems from another person rather than nature, even when expected payoffs are identical (Koehler and Gershoff, 2003; Bohnet and Zeckhauser, 2004; Bohnet et al., 2008; Aimone and Houser, 2012; Quercia, 2016). While trust is widely recognized as a cornerstone of economic performance, particularly in environments with incomplete contracts and weak legal enforcement (Knack and Keefer, 1997), betrayal aversion is believed to erode trust, strain business relationships, and impede economic development. For example, Bigoni et al. (2018) find that high levels of betrayal aversion and pessimistic beliefs help explain persistent cooperation gaps between Northern and Southern Italy. Yet, this whole literature on betrayal aversion has never addressed its potential implications for redistributive preferences and fairness ideals, even though the social risk generated by trust being betrayed may call for some compensation for it. In fact, our findings — specifically, the reduced acceptance of inequality and the greater prevalence of the Insurer type in the presence of social risk — highlight a potential counterforce when social risk puts individuals in a disadvantaged position. They suggest that individuals are more inclined to support policies that compensate those harmed by others' actions. By strengthening the social safety net for people who take interpersonal risks, such policies may encourage trust and foster greater willingness to engage in cooperative economic interactions.

The remainder of the paper is structured as follows. Section 2 describes the experimental design and procedures. Section 3 presents our main findings on the consequences of social risk for redistributive preferences. Section 4 provides supporting evidence for these findings from a follow-up study. Section 5 concludes.

(2025) show that beliefs about negative externalities of inequality — such as crime, polarization, and mistrust — can be an important driver of redistributive preferences. Note that Lobeck and Støstad (2025) are mainly concerned with the consequences of inequality (on crime, polarization and mistrust), while we are looking at the causes of inequality and how different causes have an impact on fairness types.

2 The Experiment

Our study design features two connected experiments in which different types of players — *stakeholders* and *spectators* — make decisions. We first present the details of the *stakeholder experiment.*

2.1 Stakeholder experiment

In the stakeholder experiment, participants were randomly assigned to one out of three between-subjects treatments. In each treatment, they had to choose between a *safe* option yielding a payoff of 100 tokens and a *risky* option yielding a payoff of either 0 or 300 tokens. We refer to these participants as stakeholders because they are both actively involved in the decision task and have their earnings at stake in the second, spectator experiment (see below). The treatments varied along two dimensions: (i) whether or not there were externalities for a third party, and (ii) how the risky option was resolved. Figure 1 summarizes the decision situation of the stakeholders in each of the three treatments, as detailed in the following.

Control treatment. In this treatment, stakeholders made five ordered choices between a safe option (100 tokens) and a risky option (0 or 300 tokens). The only difference across choices was the probability p that the high outcome (300 tokens) was realized, which ranged from 0% to 100% in increments of 25 percentage points. In all cases, the risky option was resolved by a random computer draw (i.e., risk of nature). For the corner probabilities of 0% and 100%, the decision reduced to a trivial choice between two certain amounts (100 vs. 0 and 100 vs. 300, respectively). For each stakeholder, a single probability level relevant for payment was randomly drawn (as discussed in the spectators' experiment).

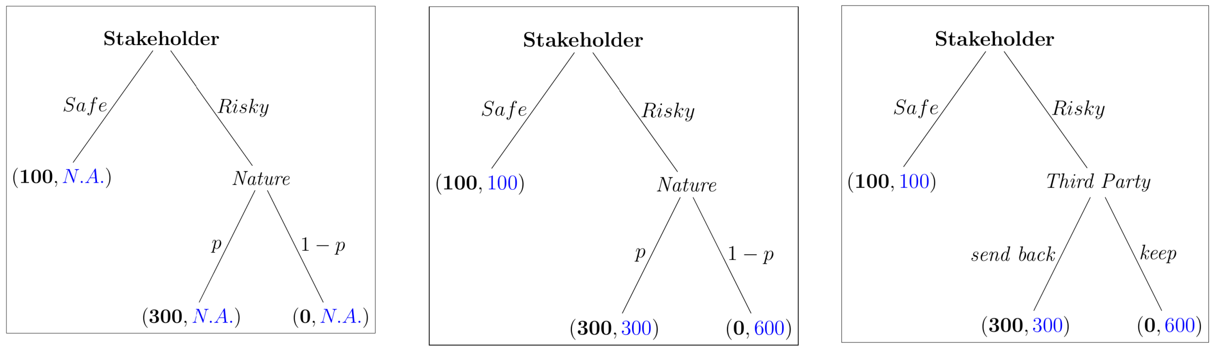
Natural-Risk treatment. This treatment is identical to the Control treatment except that the stakeholders' decisions now also affected a *passive* third party. The payoffs of the two players – the stakeholder and the third party – were as follows:

- If the stakeholder chose the safe option, both players earned 100 tokens for sure;
- If the stakeholder chose the risky option, then with probability p both players earned 300 tokens, and with probability $1-p$ the stakeholder earned 0 tokens and the third party earned 600 tokens.

As in the Control treatment, stakeholders made five ordered choices, one for each probability level $p = 0\%, 25\%, 50\%, 75\%, 100\%$, and risk was resolved by a random draw

Figure 1: Experimental games

(a) Control treatment (b) Natural-Risk treatment (c) Social-Risk treatment



Notes: Stakeholders' payoffs are indicated in bold. Third parties' payoffs (Natural-Risk and Social-Risk) are indicated in blue. All the amounts are expressed in tokens. Neutral labels were used in the experimental instructions (see the Appendix).

of nature. Thus, the only difference relative to the Control treatment is the presence of an externality on a (passive) third party. Consequently, from a stakeholder's self-interested perspective, the decision problem is unchanged and identical across the two treatments. Yet, from a social perspective it differs: in contrast to the Control treatment, in Natural-Risk, social efficiency (as measured by the sum of both players' payoffs) is always maximized when the risky option is chosen, since risk-taking creates positive externalities for the third party.

Social-Risk treatment. This treatment is identical to the Natural-Risk except that the outcome of the risky option was resolved by the third party — who became active — rather than by nature. That is, when the stakeholder chose the risky option, the third party decided whether to share the pie equally (*send back*) or take the entire pie for themselves (*keep*). Under the assumption of pure self-interest, standard game theory predicts that third parties always choose *keep* (yielding a payoff of 600), which by backward induction implies that stakeholders should select the safe option (100 instead of 0). Choosing the risky option in this game is commonly understood as a proxy for the stakeholder's trust, while choosing to send back reflects the third party's trustworthiness.

Unlike in the other two treatments, stakeholders in the Social-Risk treatment made only a single choice between the safe and the risky option.[5] In addition, they were asked to state their beliefs about the behavior of the third party. Specifically, they were asked to guess how many out of four third parties would choose send back. If their belief was correct, they received a 50-token bonus.[6] Note that our Social-Risk treatment thus not only

[5] In the other two treatments, they made five choices between risky and safe options, one for each probability level. See Section 4 for a robustness check in which stakeholders also make five decisions in the context of social risk.

[6] Third-party choices were elicited using the strategy method (Selten, 1967). That is, they had to

introduces social risk (relative to natural risk), but also a degree of uncertainty about the third party's behavior: stakeholders can form beliefs about the likelihood of outcomes, yet no objective probabilities are available. This feature is inherent to social risk-taking and a central element of our design. Whether such uncertainty affects redistributive behavior is itself an empirical question, which we address in two ways. First, using data from the three treatments above, we test whether redistribution varies with probability/belief levels. Second, we implement an additional treatment in a follow-up study that removes the uncertainty about third-party behavior. We discuss the results from both these analyses in Section 4.

2.2 Spectator experiment

In the spectator experiment, participants had to decide how to redistribute the payoffs within a pair of stakeholders.[7] Pairs were formed only among stakeholders from the same treatment (between-subjects design). For each redistribution decision, spectators were informed about each stakeholder's choice, the realized outcome, and the pair's total payoff (see Instructions and Figure E2 in the Appendix). The task of the spectators was to decide how to allocate the total earnings between the two stakeholders, in increments of 10 tokens, subject to the constraint that the full amount be distributed.

Each spectator made between 35 and 45 redistribution decisions, covering all possible outcome combinations within the assigned treatment. Only one of these decisions was payoff-relevant for a pair of stakeholders, but spectators were not told in advance which one. Decisions were organized into five blocks, each corresponding to a probability level p (Control and Natural-Risk) or a belief level (Social-Risk). At the start of each block, on-screen instructions indicated the relevant probability/belief level. For example, in Control and Natural-Risk, a block with $p = 25\%$ referred to stakeholder decisions under a 25% chance of the high outcome (300 tokens). The corresponding case in Social-Risk was a block in which stakeholders indicated that they believed that 25% (1 out of 4) of third parties would choose send back.

Importantly, spectators made redistribution choices only for situations in which both stakeholders faced the same probability level or held the same beliefs about the third parties' behavior.[8] For brevity, in the following we will refer to the probability level

indicate whether they would keep the money or send it back, with their decision implemented only if their paired stakeholder had chosen the risky option.

[7] Note that the focus of our paper is on fairness views regarding inequality arising from different forms of risk-taking and its resolution. We therefore restrict attention to redistribution between pairs of stakeholders and do not study redistribution choices between stakeholders and passive (Natural-Risk) or active (Social-Risk) third parties.

[8] While mixing stakeholders with different probabilities/beliefs is potentially interesting, it also makes the situation more complex and the results more difficult to interpret. For instance, the decision to trust might be considered selfish if the stakeholder is almost certain that the third party will send back, while

Table 1: Possible combinations of payoffs within stakeholder pairs

Scenario	Stakeholder 1	Stakeholder 2
1	Safe (100)	Safe (100)
2	Safe (100)	Risky (300)
3	Safe (100)	Risky (0)
4	Risky (0)	Safe (100)
5	Risky (0)	Risky (300)
6	Risky (0)	Risky (0)
7	Risky (300)	Safe (100)
8	Risky (300)	Risky (300)
9	Risky (300)	Risky (0)

of each block without always specifying that, in the Social-Risk treatment, this level corresponds to beliefs rather than objective probabilities.

In each block, spectators made a redistribution decision for every possible permutation of stakeholder payoffs. For intermediate probabilities (25%, 50%, and 75%), this resulted in nine possible combinations of payoffs, as illustrated in Table 1. The order in which spectators faced these scenarios was randomized at the individual level in the first block and then held constant across subsequent blocks. Each decision was presented on a separate screen, and at the end of each block, participants received a summary of their decisions, with the option to revise them. Revisions were very rare, occurring in about only 1% of all cases. In the results section, we therefore analyze only the final redistribution decisions.

In the Control and Natural-Risk treatments, the extreme probabilities $p = 0\%$ and $p = 100\%$ generated only four decision scenarios each, since by construction the high outcome (300) is impossible at p=0% and the low outcome (0) is impossible at p=100%. These treatments therefore comprised 35 decisions in total. By contrast, in the Social-Risk treatment no outcome can be ruled out with certainty, even under extreme beliefs, so all nine scenarios were presented in each block, yielding 45 decisions overall. To mitigate potential spillovers from this asymmetry (and because the extremes are least informative for our analysis), spectators first completed the three intermediate-probability blocks in randomized order, followed by the two extreme-probability blocks.

2.3 Experimental procedures

The two experiments were conducted sequentially: first the stakeholder experiment, followed by the spectator experiment. Details on the procedures of both are provided below.

it might signal other-regarding concerns if a stakeholder is rather pessimistic about the chances of getting any money back.

Stakeholders. Stakeholders were recruited via Amazon Mechanical Turk (MTurk, henceforth) using the behavioral research platform TurkPrime (Litman et al., 2017). For our experiment, we recruited a total of $n = 1,199$ online participants, of which $n = 719$ played as stakeholders and $n = 480$ played as third parties in the Natural-Risk and Social-Risk treatment. The stakeholders were equally divided across the three treatments.[9] Participation was restricted to U.S. participants with a high completion rate to minimize attrition. The stakeholders' average payment was about \$2, including a \$0.50 participation fee. The experiment took on average 8 minutes to complete (implying an average hourly rate of about \$15, comparable to many laboratory experiments and well above the platform's average hourly pay at the time of the experiment). Responses were collected via Qualtrics. Only participants who answered all control questions correctly were allowed to participate. After two incorrect attempts, participants were automatically dropped from the experiment and prevented from re-taking.

Stakeholders were informed that the study consisted of two sequential experiments. To minimize the scope for strategic behavior, and consistent with standard procedures in related work (Cappelen et al., 2013), we withheld details of the second experiment — redistribution by spectators — until after stakeholders had completed their choices. Once all stakeholder decisions were collected, the second experiment was conducted in the laboratory with spectators. Their redistribution decisions determined the stakeholders' final payoffs, which were paid out within one week, in line with standard MTurk procedures.

Spectators. We have a total of $n = 353$ participants playing in the role of spectators ($n = 116$ in Control, $n = 120$ in Natural-Risk, and $n = 117$ in Social-Risk).[10] All sessions were conducted at the Cologne Laboratory for Economic Research (CLER). Student participants from various disciplines were recruited using ORSEE (Greiner, 2015) and the experiment was programmed using z-Tree (Fischbacher, 2007). Upon arrival, participants were randomly assigned to a cubicle and no form of communication was allowed. A paper copy of the instructions was distributed (see the Instructions in the Appendix). Participants could only proceed to the decision-making part of the experiment if they answered all control questions correctly. Socio-demographic characteristics were collected at the end of the experiment. Spectators were paid a fixed amount of € 10 for the redistribution part, including a show-up fee of € 4. The average session lasted approximately 45 minutes.

[9]The slightly uneven numbers arises from one participant in the Natural-Risk treatment who did not complete the experiment. Consequently, the Social-Risk treatment included $n = 240$ stakeholders and $n = 240$ third parties, the Natural-Risk treatment $n = 239$ stakeholders and $n = 240$ third parties, and the Control treatment $n = 240$ stakeholders.

[10]The uneven numbers are due to a low shop-up rate in some of the sessions. In case we had fewer spectators than pairs of stakeholders, we randomly drew one of the spectators and applied their decision twice.

3 Results

The main question of this paper is how different sources of risk and the presence or absence of externalities affect inequality acceptance and redistribution behavior. Therefore, we focus our analysis on spectator choices, as elicited in the second experiment of our study. Stakeholders' choices, as elicited in the first experiment, in contrast, are not of particular interest for our purposes. However, the data are instrumental to the second experiment, as they are needed to ensure that spectators make choices that have real monetary consequences for some stakeholders. Therefore, we report the stakeholder choice results only in the Appendix. Our analyses mainly focus on choices from the blocks with intermediate probabilities (25%, 50%, and 75%), because only for intermediate probabilities there is a choice between a safe and risky option. For corner probabilities, stakeholders' decisions reduce to a binary dictator choice between two deterministic options. In such cases, there is no risk and nothing to insure. Thus, redistribution cannot be motivated by compensating risk or misplaced trust, but must instead reflect other motives. For example, with a 100% probability, the choice is between a certain payoff of (100, 100) and an equally certain payoff of (300, 300). Deviations from the dominant option are rare and likely reflect cognitive ability or effort rather than morally relevant preferences. In this sense, redistribution at corner probabilities takes on a completely different meaning, one that is unrelated to the source of risk.[11] Within these blocks, spectators ($n = 353$) made a total of 9,531 redistribution decisions. In 1,059 of these decisions, both stakeholders earned 0 tokens, and, hence, there was nothing to redistribute. In another 2,118 cases, both stakeholders had the same strictly positive income — either by choosing the safe option or by earning the same high payoff from the risky option. As one might expect, in almost all of these cases (97%), spectators did not redistribute any money from one stakeholder to the other. In the following, we therefore focus on the remaining 6,354 cases with initial inequality — those situations in which the payoffs of the two stakeholders differ.

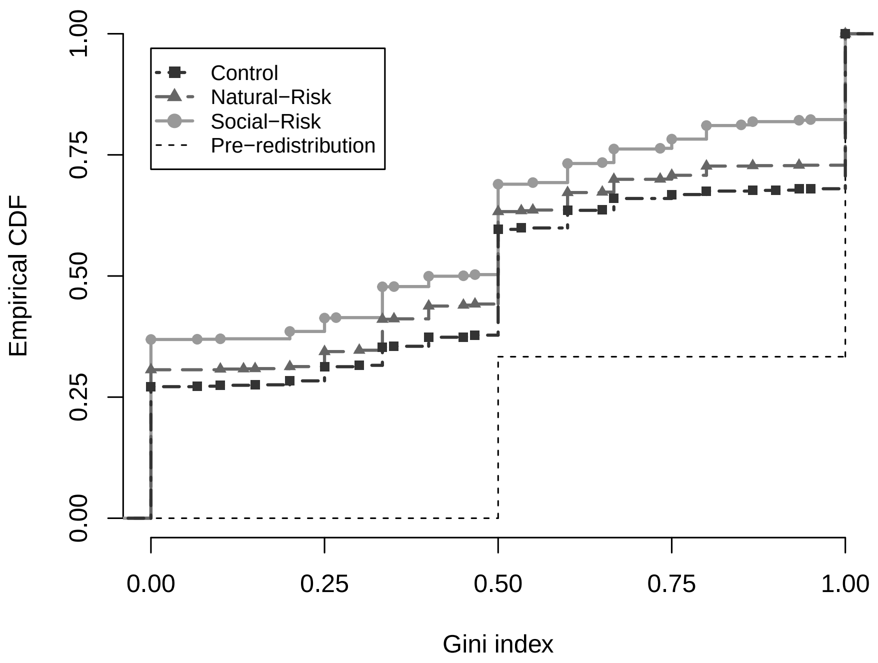
3.1 Aggregate analysis: Inequality across treatments

We start our analysis by comparing the overall level of implemented inequality across treatments. To this end, we compute the post-redistribution Gini index within each pair after the spectators' redistribution decisions. In our setting, the Gini index is simply defined as the absolute difference between the two stakeholders' final payoffs, normalized by the sum of the payoffs in the pair:

$$e_i = \frac{|\text{payoff of stakeholder 1} - \text{payoff of stakeholder 2}|}{\text{total payoff}} = |1 - 2y_i| \in [0, 1], \qquad (1)$$

[11] We discuss the results for corner probabilities in the Appendix A.

Figure 2: Distribution of post-redistribution Gini index across treatments



where stakeholder 1 is defined as the one with the higher pre-redistribution payoff, and ($y_i \in [0, 1]$) denotes the share of the total payoff that spectator i allocates to stakeholder 1 in a given redistribution decision. This measure equals 1 if no redistribution occurs (i.e., all payoff is left with the higher earner), and 0 if payoffs are fully equalized.

Note that, by construction, the pre-redistribution Gini index averaged over all decisions is identical across treatments (since the set of decisions is identical), equal to 0.83. In each treatment, one-third of scenarios with initial inequality had a pre-redistribution Gini of 0.50 (pairs in which one stakeholder chose risky and earned 300, and the other chose safe and earned 100), while the remaining scenarios had a Gini of 1 (pairs in which one stakeholder chose risky and earned 0, and the other earned either 100 or 300).

We compute the Gini index only over stakeholders' payoffs, excluding those of (active or passive) third parties. This choice is justified for two reasons: (i) third-party earnings cannot be consistently defined in the Control treatment, preventing meaningful comparisons across treatments; and (ii) spectators can modify only stakeholders' payoffs, but not those of third parties. Since our focus is on redistribution by spectators, it is natural to restrict attention to the dimensions they can actively influence.

Our results are summarized in Figure 2, which displays the cumulative distribution of the post-redistribution Gini index across treatments. As can be seen, we observe a pronounced reduction in inequality in all three treatments, with the average Gini index falling from 0.83 to 0.47. More strikingly, we find pronounced and significant differences in the level of inequality implemented across treatments (Kruskal-Wallis test, $p = 0.005$).[12] In particular, as indicated by the upwards-shifted CDFs, we observe more inequality-reducing redistribution when choices involve social rather than natural risk, as well as when externalities are present (rather than absent as in Control). As a result, the level of implemented inequality is highest in the Control treatment (Gini index = 0.52) and lowest in the Social-Risk treatment (Gini index = 0.41). The Natural-Risk treatment lies in between these two (Gini index = 0.48). Statistical analysis reveals that the Gini index is significantly lower in Social-Risk than in each of the other two treatments (Mann-Whitney U test, $p < 0.038$ in both pairwise comparisons), but that Control and Natural-Risk are not different (Mann-Whitney U test, $p = 0.286$).

These findings are confirmed by regression analysis. Model 1 in Table 2 reports the results of an OLS regression with the post-redistribution Gini index as the dependent variable and two dummy variables for the Natural-Risk and the Social-Risk treatments as regressors.[13] To control for the dependency of observations within subjects, we cluster standard errors at the individual level. The results reveal a statistically significant decrease in the implemented level of inequality in Social-Risk relative to the other two treatments. This indicates that the source of risk — natural versus social — matters for the extent to which individuals accept inequality. By contrast, the presence of externalities per se has only a small and statistically insignificant effect, as indicated by the Natural-Risk dummy.

To better understand the source of these results, we examine whether the observed treatment differences are driven by the fact that there are more cases in which people want to reduce inequality when social rather than natural risk is involved (*extensive margin*), or whether this effect is driven by the amount of inequality reduction conditional on reducing inequality at all (*intensive margin*).

To test for potential differences at the extensive margin, we simply count how often spectators increased, decreased, or did not change initial inequality. We find the highest instances of inequality reduction in the Social-Risk treatment (60% of the cases), followed by the Natural-Risk treatment (53%) and the Control treatment (45%). The fraction of cases without any change in initial earnings, in turn, is highest in Control (52%), followed by Natural-Risk (44%) and Social-Risk (36%).[14] To evaluate the significance of these results, in Model 2 of Table 2 we present the results of a logistic regression with a dummy variable that takes the value 1 if inequality was reduced (and 0 otherwise) as the

[12]When using non-parametric tests to compare treatment differences, we calculate individual averages by collapsing the data so that we have only one observation per individual.

[13]As a robustness check, we also estimate a fractional response regression with robust errors to account for the fact that the dependent variable ranges from 0 to 1. All results are quantitatively and qualitatively robust and available from the authors upon request.

[14]The fraction of cases in which inequality is increased, in contrast, is very similar across treatments (ranging between 3% and 4%).

Table 2: Redistribution across treatments

	Model 1	Model 2	Model 3
Natural-Risk tr (d)	-0.043	0.302*	0.012
	(0.033)	(0.172)	(0.030)
Social-Risk tr (d)	-0.112***	0.606***	0.006
	(0.050)	(0.165)	(0.033)
Constant	0.518***	-0.196	0.179***
	(0.023)	(0.123)	(0.021)
Post-estimation F-tests			
Natural-Risk - *Social-Risk* $= 0$	$p = 0.037$	$p = 0.060$	$p = 0.842$
Nr. observations	6354	6354	3345
Nr. clusters	353	353	303
R^2/log likelihood	0.014	-4347	0.000

Notes: Models 1 and 3: OLS regressions with the post-redistribution Gini index as dependent variable. Model 2: Logistic regression. The dependent variable is a dummy taking value 1 if the stakeholder decreased the initial inequality level in the pair and 0 otherwise. In Model 3 the sample is restricted to cases in which the inequality was decreased. In all models we cluster standard errors at the individual level. Symbols ***, **, and * indicate significance at the 1%, 5% and 10% level, respectively.

dependent variable. As before, we use two dummy variables, one for the Natural-Risk and one for the Social-Risk treatment, as explanatory variables, and cluster standard errors at the individual level. Our results show that spectators reduce inequality significantly more often in Natural-Risk and Social-Risk than in Control, and (weakly significantly) more often in Social-Risk than in Natural-Risk.

To test whether redistribution behavior differs across treatments also at the intensive margin, we re-run Model 1 (with the post redistribution Gini index as the dependent variable) but restrict the sample to cases in which inequality was reduced.[15] The results, reported in Model 3 of Table 2, show that the coefficients on the two treatment dummies are close to zero and statistically insignificant. This indicates that, conditional on redistributing, the extent of redistribution is very similar across treatments. We summarize our findings so far as follows:

Result 1 *Inequality is reduced significantly more when the resolution of risk happens in a strategic context by another person (social risk) than when it is resolved by nature. This effect is driven primarily by the likelihood that spectators engage in redistribution (extensive margin), rather than by the amount redistributed conditional on doing so (intensive margin). By contrast, the mere presence of externalities has only a small effect on redistribution behavior.*

[15] As for Model 1, we run a fractional response regression model with robust errors; all results are robust to this specification.

3.2 Individual-level analysis

To gain a better understanding of the drivers and mechanisms behind the aggregate results, we next analyze spectators' redistribution patterns in greater detail. In particular, we investigate whether differences in inequality reduction can be explained by the emergence of different behavioral types — each associated with a specific redistribution pattern over pairs of stakeholders.

To estimate types, we use a two-pronged approach combining cluster analysis and parametric mixture models — for a similar exercise see Fallucchi et al. (2019). First, we perform a k-means cluster analysis, which is an unsupervised classification method that partitions individuals into k distinct groups (clusters) based on their vector of choices such that the variance within clusters is minimized (see Appendix C for a detailed description of the method and procedures). The advantage of this approach is that it does not impose behavioral types ex-ante, but rather lets the data speak for itself. The disadvantage is that there is no a priori guarantee that clusters will be behaviorally meaningful. However, as we will show below, this is not a problem in our data, as meaningful types emerge, including the ones commonly identified in the literature. Second, informed by the cluster analysis results and following previous literature (Cappelen et al., 2013), we use a parametric mixture model to structurally estimate the prevalence of different behavioral types in each treatment. Importantly, the types are not defined ex-ante by the researchers, but are based on the results of the cluster analysis.

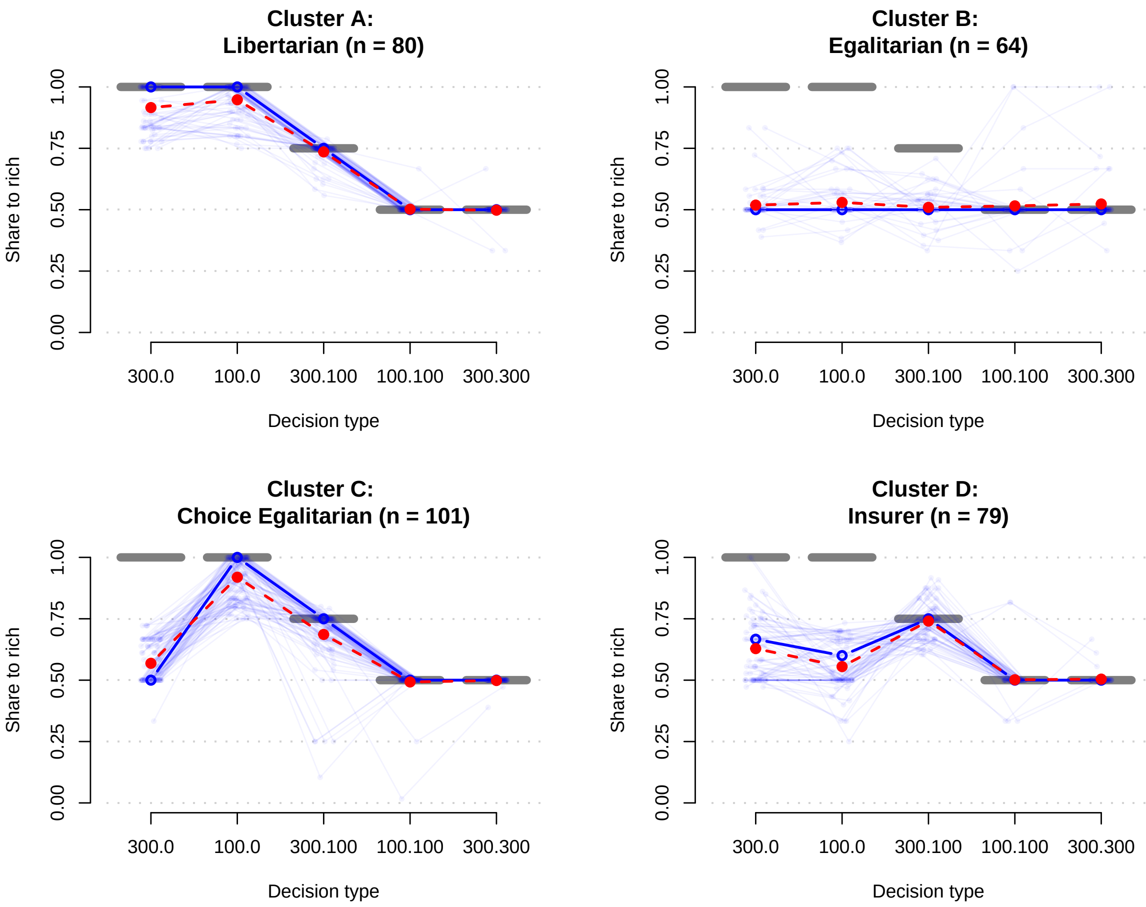
Cluster analysis. Figure 3 reports the main results of the k-means cluster analysis for $k = 5$ clusters.[16] The horizontal axis plots the five payoff pairs with strictly positive earnings (*300-0*, *100-0*, *300-100*, *100-100*, and *300-300*).[17] The label of each pair refers to the pre-redistribution allocation of resources across the two stakeholders[18] On the vertical axis, we report the share of resources allocated to the first stakeholder, i.e., the one with

[16]Appendix C describes the details of the k-means cluster analysis and the procedure used to select the most appropriate number of clusters, k. In the main text, we focus only on the four main clusters. The fifth cluster, reported in the Appendix, accounts for only 8% of spectators and exhibits more erratic patterns. We treat it as a residual group and exclude it from the mixture model.

[17]Note that in this section of the paper we include all pairs with strictly positive amounts to be distributed, including those without initial inequality. Even though there is almost no redistribution in these pairs, we decided to include them for comparability with previous papers that used them in their parametric estimation of fairness types (Cappelen et al., 2013). Since our cluster analysis is intended to inform the analysis of types to be used for the mixture model, we use the same set of observations in both analyses. Further note that each participant makes six decisions for each payoff pair involving strictly positive inequality, as we vary both the order of the stakeholders and the probability level. This increases the precision of our estimates and reduces the noise that may result from trembling in the redistribution process.

[18]Recall that for each pair with initial inequality, spectators were asked to make two decisions: one in which the richer player was first (e.g., 300-0) and one in which they were last (e.g., 0-300), as detailed in Table 1). We find a very high degree of consistency across these decisions: only in less than 15% of the cases the redistribution decision was not exactly the same. In Figure 3, we therefore show aggregated data across these two decisions. See Figure A1 in the Appendix for the full disaggregated analysis.

Figure 3: Redistribution behavior of the different clusters. k-means cluster analysis with ($k = 5$)



Notes: The x-axis shows the five payoff pairs with strictly positive earnings. Each pair reports the pre-redistribution payoffs of the richer and poorer stakeholder, respectively. The y-axis shows the final share of the total payoff allocated to the richer stakeholder after redistribution. Light-gray horizontal bars indicate this stakeholder's pre-redistribution share. Red dashed lines plot the average post-redistribution share based on the cluster centroid (the mean point in the cluster), while solid blue lines plot the share based on the cluster medoid (the most central observation). Shaded blue lines depict individual behavior.

the (weakly) higher ex-ante payoff. The light-grey horizontal bars represent their pre-redistribution share. This share is equal to 1 in pairs where the other stakeholder had 0 tokens and thus all resources were initially allocated to the first stakeholder (*300-0* and *100-0*), it is 0.75 in the *300-100* pair, and it is 0.50 in the pairs where resources were initially shared equally (*300-300* and *100-100*). The red dashed lines plot the average share of resources allocated to the first stakeholder after redistribution, calculated using the centroid of the cluster — the average point in the cluster. The blue solid lines report the same data using the medoid of the cluster — the most central subject in the cluster.[19] The shaded blue lines report behavior at the individual level.

As can be seen, in all clusters there is virtually no redistribution for pairs with no initial

[19]The medoid is the observation within a cluster whose average dissimilarity to all other members of the cluster is minimal—that is, it is the most central, representative data point.

inequality (*300-300* and *100-100*). In the following, we therefore focus our discussion of the clusters on the redistribution behavior in the pairs *with* initial inequality (*300-0*, *100-0*, and *300-100*), where clear differences emerge. Out of the four clusters shown in Figure 3, three (Clusters A to C) identify fairness types that are well known in the literature (Cappelen et al., 2013; Bortolotti et al., 2025). Importantly, if we restrict the analysis to the Control treatment, which is the closest to those in previous studies, the cluster analysis confirms the presence of only three types — the same as those identified in earlier papers. One cluster, Cluster D, is a novel type, however. In the following, we discuss each of the clusters in detail:

- Cluster A ($n = 80$) — Spectators in this cluster leave the distribution of resources largely unaffected. In fact, both the centroids and the medoids of the ex-post distribution of payoffs are almost identical to the ex-ante distribution of resources (light-grey horizontal bars). This type is commonly known in the literature as **Libertarian** and represents the second largest group in our data.

- Cluster B ($n = 64$) — This type of spectator always redistributes and shares resources equally among the two stakeholders, regardless of their choices. This means that they do not consider any kind of inequality to be fair. This pattern is typical of **Egalitarians** and is the fourth most common pattern in our data.

- Cluster C ($n = 101$) — The third type is an intermediate type that considers inequality unfair when it arises from differences in luck, but not when it arises from differences in risk-taking. Thus, spectators of this type redistribute resources equally in the case where both stakeholders made the same risky choice but ended up with different outcomes (*300-0*). Instead, when inequality results from different choices — i.e., when one stakeholder chose the safe option while the other one chose the risky option (pairs *100-0* and *300-100*) — they leave the initial distribution of payoffs unchanged. This type is commonly referred to as **Choice Egalitarians** and represents the largest group in our data.

- Cluster D ($n = 79$) — The fourth type is characterized by a redistribution pattern that offsets inequality only when one stakeholder was let down — either by nature or by another player — and ended up with 0 tokens (pairs *100-0* and *300-0*). If instead both stakeholders earn strictly positive payoffs (pair *300-100*), the initial distribution of payoffs is left unchanged. Thus, in contrast to Choice Egalitarians, the fairness ideal of this type is not shaped by ex-ante choices, but rather by the occurrence of bad luck (either by nature or another player). We refer to this new type as **Insurers**. They are the third largest group in our sample.

Result 2 *We identify a new fairness type not documented in previous studies: the Insurer, who consistently compensates individuals that take risks but incur losses. In addi-*

tion, we recover the three well-established types identified in earlier work: Egalitarians, Libertarians, and Choice Egalitarians.

Table 3 reports the fraction of types separately for each treatment. The results reveal significant differences in the distribution of types across treatments ($\chi^2(8) = 26.70$, $p = 0.001$). In particular, while in the Control and Natural-Risk treatment there is a clear prevalence of Libertarian and Choice Egalitarian types, in the Social-Risk treatment the two most frequent types are Insurers and Egalitarians. Consequently, when conducting pairwise comparisons across treatments, we find significant differences between the Social-Risk and the other two treatments ($\chi^2 - tests$, both $p < 0.016$), but not between Control and Natural-Risk ($\chi^2(4) = 3.45$, $p = 0.486$). The differences between treatments are thereby mainly driven by differences in the fraction of Insurer types that redistribute only when a stakeholder ended up with a payoff of zero. In the Social-Risk treatment, the proportion of Insurer types is 31.6%, which is almost twice as high as in the Control and Natural-Risk treatments (17.2% and 18.3%, respectively).[20]

Overall, these results show that when decisions involve social rather than natural risk, different fairness considerations emerge, with a stronger inclination to redistribute, consistent with our aggregate findings above.

Before turning to the choice model, we emphasize that our behavioral types are defined over action profiles across payoff pairs with different initial inequalities, and these profiles are held constant across treatments. This differs from the approach commonly used in the luck-versus-merit literature, where types are defined by how actions vary across treatments and sources of inequality (e.g., Almås et al., 2020). For example, a meritocratic spectator equalizes earnings when inequality arises from luck but not when it stems from merit. Applying such an approach in our setting would be considerably more complex and would generate a large number of types, since each treatment involves three relevant cases rather than one. In the final part of this section, we return to this issue by examining the sensitivity of our classification, focusing on how the distribution of types varies across treatments at the population level.

Choice model. In the previous subsection, we have shown that four meaningful behavioral types emerge from the data using a clustering analysis with no structure imposed ex-ante. Building on these results, we now estimate four types parametrically using a discrete-choice random utility model. Following previous literature (Cappelen et al., 2013), we assume that spectators are motivated solely by fairness views, as self-interest does not apply here (spectators were paid a flat fee for the task). Specifically, let X

[20] Pairwise comparisons reveal that the differences between treatments are significant between Control and Social-Risk ($\chi^2(1) = 6.52$, $p = 0.011$) and Natural-Risk and Social-Risk ($\chi^2(1) = 5.60$, $p = 0.018$), but not between Control and Natural-Risk ($\chi^2(1) = 0.05$, $p = 0.826$).

Table 3: k-means clustering classification by treatment

Cluster type	Control	Natural-Risk	Social-Risk	Cluster size
A) Libertarians	27.6%	24.2%	16.2%	80
B) Egalitarians	14.7%	19.2%	20.5%	64
C) Choice Egalitarians	37.1%	30.8%	18.0%	101
D) Insurers	17.2%	18.3%	31.6%	79
E) Others	3.4%	7.5%	13.7%	29
Total (%)	100%	100%	100%	
Nr. observations	116	120	117	353

denote the total income in the pair of stakeholders to which a spectator is assigned. The spectator's utility from allocating y to the first stakeholder and $X - y$ to the second is given by:

$$V(y; \cdot) = -\beta(y - F^k)^2/2X \tag{2}$$

where F^k denotes the fair amount allocated to the first stakeholder according to the spectator's type k, and where β is the weight attached to fairness. A spectator's utility is decreasing in the distance between the amount y and the fair amount F^k prescribed by their type k.

Spectators may differ along two dimensions: (i) the extent to which they care about fairness (β); and (ii) their type (F^k). Guided by the clustering analysis, we distinguish four behavioral types, presented below in ascending order of redistribution (and, correspondingly, by decreasing Gini index):

- ***Libertarians (based on cluster A)*** never support redistribution; if x is the income of the first stakeholder before redistribution, we have $F^{Libertarians} = x$, which yields the optimal choice $y = x$.

- ***Choice Egalitarians (based on cluster C)*** differentiate between ex-ante and ex-post inequality:

$$F^{ChoiceEgalitarians} = \begin{cases} X/2 & \text{if } C_1 = C_2 \\ x & \text{if } C_1 \neq C_2 \end{cases}$$

  where C_i takes value 1 if stakeholder i chooses the risky option and 0 if he/she chooses the safe option.

- ***Insurers (based on cluster D)*** equalize earnings only when a stakeholder who faced a bad draw — by nature or another player — is present in the pair:

$$F^{Insurers} = \begin{cases} X/2 & \text{if } x = 0 \text{ } or \text{ } X - x = 0 \\ x & otherwise \end{cases}$$

- ***Egalitarians (based on cluster B)*** always eliminate inequality within a pair and split the earnings equally: $F^{Egalitarians} = X/2$, which yields the optimal choice $y = X/2$.

Given a spectator's type k, we consider a discrete choice random utility model of the form:

$$U(y; \cdot) = V(y; \cdot) + \varepsilon_{iy} \qquad \text{for } y = 0, 10, \ldots, X \tag{3}$$

where ε_{iy} is assumed to be i.i.d extreme value distributed. To control for individual heterogeneity in noisy behavior, the β parameter in equation 2 is assumed to be log normally distributed with $\log(\beta) \sim \mathcal{N}(\zeta, \sigma^2)$. By denoting $L_{i,k}$ the individual likelihood conditional on being of type k, we can obtain the total likelihood of an individual by considering the finite mixture of types $L_i = \sum_k \lambda^k L_{i,k}$, where λ^k is the probability of being of type k.[21]

Figure 4: Estimation of behavioral types

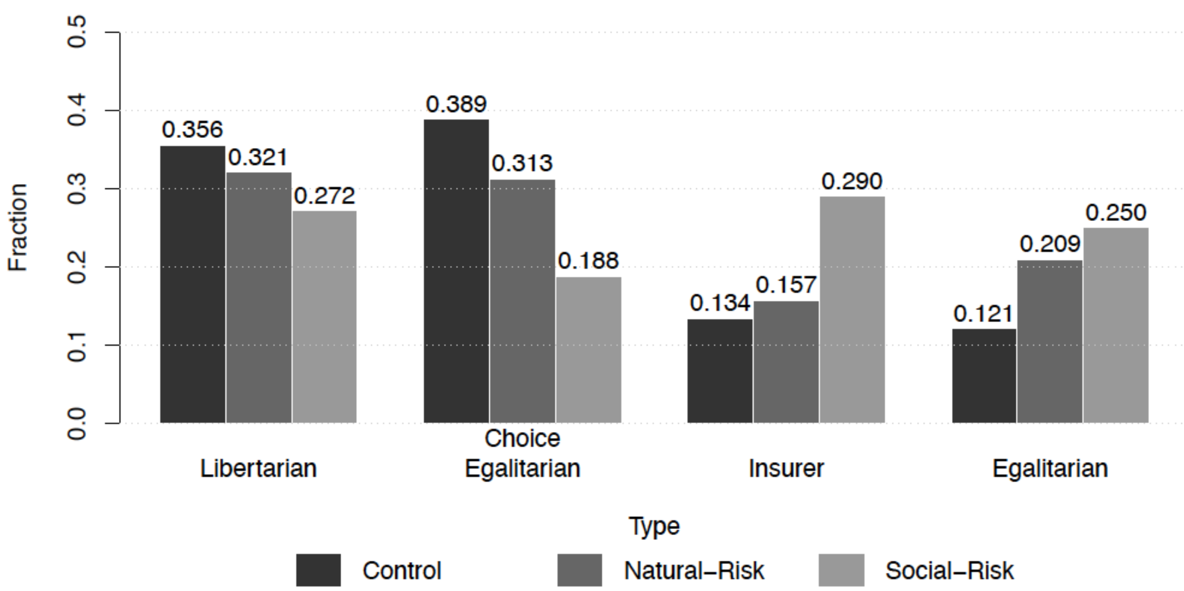


Figure 4 reports the estimated proportion of types, λ^k, separately for each treatment. The results are very much in line with our findings from the cluster analysis. They show

[21]For further details on the estimation strategy, see Appendix D.

a clear shift toward types favoring more redistribution in the Social-Risk compared to the Control and Natural-Risk treatments, where Libertarians and Choice Egalitarians are the majority. A series of likelihood ratio tests confirm that the distribution of types across treatments is different and statistically significant, except for the comparison of Control and Natural-Risk (Control vs. Natural-Risk, $\chi^2(5) = 3.727$; $p = 0.589$; Control vs. Social-Risk, $\chi^2(5) = 25.634$, $p < 0.001$; Natural-Risk vs. Social-Risk, $\chi^2(5) = 12.983$, $p = 0.024$).[22] We summarize our findings as follows:

Result 3 *The differences in inequality reduction between treatments are explained by differences in the distribution of types. When risk is resolved by nature (Control and Natural-Risk), the majority of spectators do not redistribute at all (Libertarians) or only when both stakeholders chose the risky option (Choice Egalitarians). In contrast, when risk is resolved by another person (Social-Risk), there is a marked shift toward Egalitarian and Insurer types, who also redistribute when stakeholders made different choices, especially when one stakeholder was let down by a third party.*

Sensitivity of behavioral types to social risk. Our behavioral types prescribe redistribution patterns over payoff pairs with different pre-redistribution allocations, and these prescriptions are treatment-invariant. For example, an Insurer is prescribed to equalize earnings in the 100–0 and 300–0 pairs, but not in the 300–100 pair, across all treatments. Up to this point, we have remained agnostic about the sensitivity of spectators' choices to social risk — that is, we have not imposed any assumption on how spectators' choices can follow the same or different types across treatments. Results from both the cluster analysis and the choice model, however, indicate that spectators do respond to our treatment manipulation. In other words, the same individual may change their pattern of actions when exposed to different treatments. For instance, a spectator classified as a Libertarian in the Control treatment may instead behave as a Choice Egalitarian in Natural Risk. This would suggest that the spectator views redistribution among stakeholders who make the same risky choice but face different outcomes as fair only when risk-taking generates positive externalities and welfare gains. We thus talk about *spectators' sensitivity* to specific features of risk whenever a change of type occurs.

A proper study of the spectators' sensitivity and the transitions between types would require a within-subject design. Nevertheless, under reasonable assumptions, we can provide population-level estimates of the share of spectators who are insensitive to social risk (i.e., who do not change type) and those who are sensitive. We begin by comparing

[22] The accuracy of our type classification is confirmed by the posterior probability of a given spectator to belong to a particular type. Overall, 80% of our spectators have a posterior probability greater than 90% of belonging to one of the four types. Furthermore, we find no systematic variation in the percentage of spectators with a posterior probability greater than 90% across treatments (78% in Control, 84% in Natural-Risk, and 78% in Social-Risk, $\chi^2(2) = 2.274$, $p = 0.321$).

the Control and Social-Risk treatments, under two assumptions: (i) if a spectator changes type when moving from Control to Social-Risk, the change always occurs in the direction of more redistribution; and (ii) transitions are possible only to the *next* type.

Assumption (i) implies that spectators classified as Egalitarians in Control must remain Egalitarians in Social-Risk. Conversely, spectators classified as Libertarians in Social-Risk must also have been Libertarians in Control. Assumption (ii) is somewhat stronger, as it imposes a structure on the transition from Control to Social-Risk, but is necessary to derive a unique solution. We order types by ascending levels of redistribution, as shown in Figure 4. Accordingly, a Libertarian in Control can either remain a Libertarian in Social-Risk or transition to a Choice Egalitarian. Likewise, Choice Egalitarians can become Insurers, and Insurers can become Egalitarians when moving from Control to Social-Risk.

Table 4 reports the estimated stability and sensitivity of types across treatments, under the assumptions discussed above. By *insensitive* we mean the share of spectators behaving in the same way across two treatments, while by *sensitive* we refer to those who transition into another type. All estimates are derived from the choice model results reported in Figure 4. When moving from Control to Social-Risk, the population is almost evenly split: about half of spectators are insensitive (50.2%) while the other half (49.8%) change type. Most of this sensitivity to social risk is accounted for by transitions into the Insurer category, with 28.5% of spectators moving into this type. Comparing Natural- to Social-Risk, we still find substantial sensitivity, with about one in four spectators (26.4%) changing type. Again, the bulk of this sensitivity comes from transitions into Insurers, with 17.4% of spectators moving into this category.

While exploratory and based on reasonable but simplifying assumptions, we consider this exercise as informative: it shows that most of the sensitivity to social risk is concentrated in the Insurer type, with a sizable share of Choice Egalitarians transitioning into Insurers when risk is resolved by a third party rather than by nature. This supports the idea that this novel type went unnoticed in earlier work because it is activated primarily by the way risk is resolved, and in particular by the possibility of being betrayed by a third party. By contrast, our cluster analysis and choice model took a more agnostic approach, defining types as redistribution patterns that remain constant across treatments. Taken together, both approaches highlight the importance of social risk in shaping fairness views and in revealing types of redistributive preferences that remained hidden in earlier work.

4 Discussion

We have so far examined the distinction between natural and social risk, and the extent to which impartial spectators support different redistributive norms in these contexts.

Table 4: Stability and sensitivity of types across treatments.

	Control → Social-Risk		Natural-Risk → Social-Risk	
	Insensitive	Sensitive	Insensitive	Sensitive
Libertarians	27.2	–	27.2	–
Choice Egalitarians	10.4	8.4	13.9	4.9
Insurers	0.5	28.5	11.6	17.4
Egalitarians	12.1	12.9	20.9	4.1
Total	50.2	49.8	73.6	26.4

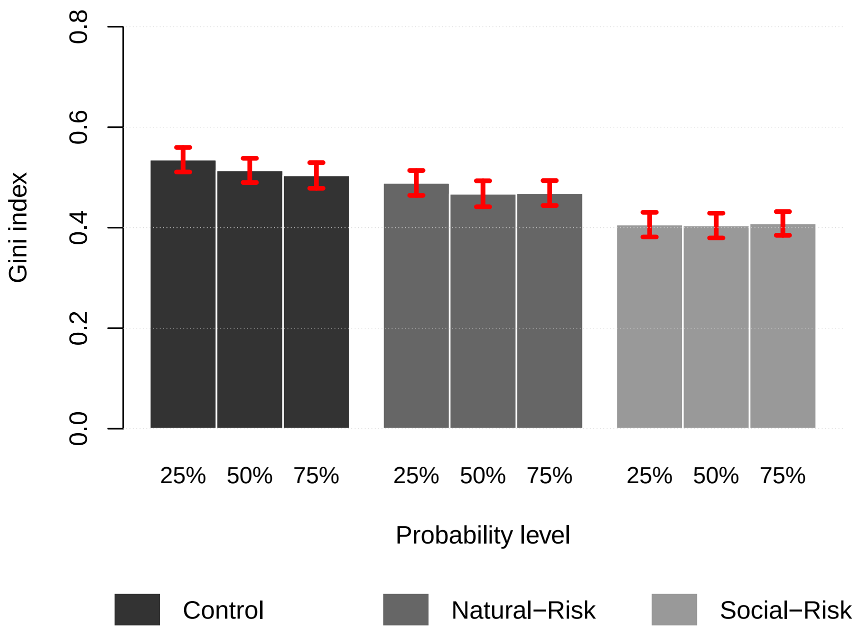
Notes: The table reports the estimated shares of spectators who are insensitive (i.e., retain the same type) or are sensitive (i.e., switch to another type) when moving across treatments. Estimates are derived from the choice model (Figure 4), under the assumptions that (i) type changes occur only in the direction of greater redistribution and (ii) transitions are possible only to the next type in the ordering shown in Figure 4. Values are percentages of the population. A dash (–) indicates that a transition is not possible under these assumptions.

While this distinction is central to most market interactions, risk-taking is a multifaceted phenomenon, and various factors may influence redistribution preferences. Indeed, in our setting, pre-redistribution payoffs are not merely the result of decisions taken in isolation and luck, but are closely linked to the other-regarding preferences of our stakeholders (in Natural-Risk and Social-Risk). This connection may give rise to redistribution preferences that are richer and more nuanced than those usually considered. We have so far examined the distinction between natural and social risk, and the extent to which impartial spectators support different redistributive norms in these contexts. While this distinction is central to most market interactions, risk-taking is a multifaceted phenomenon, and various factors may influence redistribution preferences. Indeed, in our setting, pre-redistribution payoffs are not merely the result of decisions taken in isolation and luck, but are closely linked to the other-regarding preferences of our stakeholders (in Natural-Risk and Social-Risk). This connection may give rise to redistribution preferences that are richer and more nuanced than those usually considered.

Beyond the role of externalities — captured by comparing Control and Natural-Risk (which are not significantly different, though) — a key dimension for redistribution is the likelihood of success when engaging in risk. Stakeholders who choose a risky option despite a low probability of success (e.g., 25%) might reasonably be seen as accountable for adverse outcomes, as such failure could have been anticipated. This could diminish both the perceived responsibility and the sense of betrayal in Social-Risk, leading spectators to reduce the amount of redistribution. However, in both Natural- and Social-Risk, the risky choice significantly enhances total welfare — a feature common in many real-world scenarios where entrepreneurs invest in innovation. In this light, an impartial spectator might instead favor compensating unlucky stakeholders who pursued risk despite low expected returns.

We can address whether one potential motivation for spectators dominates the other with our design as it comprises multiple probability levels. Therefore, we repeat our aggregate analysis from Section 3.1, but this time split up by belief/probability levels. The results are shown in Figure 5, which plots the average post-redistribution Gini index separately for each probability level and treatment. As can be seen, we find that the level of implemented inequality within a treatment does not systematically depend on the different belief/probability levels. Regression analyses, reported in Table A2 in Appendix A, confirm that beliefs/probabilities do not significantly affect the implemented inequality in any of the treatments.[23]

Figure 5: Post-redistribution Gini index by treatment and probability level (with SEM at the individual level)



Another important dimension of risk-taking concerns the behavior of the third party in the Social-Risk treatment. When social risk and trust are involved, individuals must generally rely on beliefs about the counterpart's actions rather than on precise probabilities. To assess whether this aspect influences spectators' redistribution choices and fairness views, we conducted a follow-up study with a modified Social-Risk treatment in which we eliminated the uncertainty associated with beliefs about the third party's behavior. Specifically, in the new treatment, called Social-Prob, stakeholders were paired with a group of four third parties. Stakeholders were then asked to make five decisions between

[23]Note that this result is also consistent with previous evidence, showing that changing the attractiveness of the risky option by varying the payoff of the safe option (holding the payoff and the probability of the risky option constant), has no effect on redistribution behavior (Cappelen et al., 2013; Bortolotti et al., 2025).

the safe and the risky option, conditional on $M \in \{0, 1, 2, 3, 4\}$ out of the four third parties choosing to send back money.[24] Only after stakeholders had made all their choices did they learn which scenario was relevant — how many out of the four third parties they were paired with actually chose to send money back. To determine their earnings, we then randomly matched them with one out of the four third parties. The design and procedures for the spectators were exactly the same as for the other treatments. In total, we collected data from $n = 120$ spectators, $n = 240$ stakeholders, and $n = 240$ third parties, none of whom had participated in our first study. As before, spectator decisions were elicited at the Cologne Laboratory for Economic Research (CLER) using student participants, while stakeholder and third parties decisions were elicited using MTurk.

To test for the effect of beliefs vs. probabilities in the Social-Risk treatment, we re-estimate the discrete choice model including the new treatment. We find that in Social-Prob, the share of Insurers is 22%, slightly lower than in Social-Risk (29%), but still about 40% higher than in the Natural-Risk treatment. When comparing the distribution of types across treatments, we find a statistically significant difference between Control and Social-Prob ($\chi^2(5) = 12.983, p = 0.002$), and Natural-Risk and Social-Prob ($\chi^2(5) = 12.722, p = 0.026$), but not between Social-Risk and Social-Prob ($\chi^2(5) = 6.606, p = 0.252$). Overall, these results show that while ambiguity may play some role in explaining the treatment differences in redistribution, it does not capture the full effect when moving from natural to social risk. We summarize these findings in our final result:

Result 4 *Ambiguity in risk-taking has some effect on redistributive behavior, but cannot explain the differences between natural versus social risk.*

5 Conclusion

This study examines how redistributive preferences and perceptions of fairness are shaped by the context in which inequality arises — particularly whether it results from natural or social risk, and whether externalities are present. Using a laboratory experiment with impartial spectators, we find that individuals exhibit significantly less tolerance for inequality stemming from social risk — where outcomes depend on another person's choice to reciprocate or betray — than from natural risk — where outcomes depend on brute luck. This aversion to inequality born of social interdependence is driven by a greater likelihood of redistribution, rather than more intensive transfers. The consideration of social risk has allowed us to identify a novel fairness type, which we term Insurer, who

[24] $M = 0$ implies no third party will *send back*, leading to a certain choice between 100 and 0 (or 100 and 600 for the third party). Conversely, $M = 4$ offers a certain choice between two fixed amounts. If the stakeholder chooses the risky option, intermediate values of $M = 1, 2, 3$ correspond to probabilities, p, of 25%, 50%, and 75%, respectively, for receiving 300. Our analysis focuses primarily on these intermediate probabilities.

consistently compensate individuals who take risks but incur losses. This type is especially prevalent when outcomes hinge on others' choices, suggesting that the interpersonal nature of risk shapes moral reasoning about fairness.

Our findings offer several important implications for economic theory and policy. First, they refine theories of distributive justice by highlighting how the nature of risk — specifically, whether it is social or natural — influences fairness judgments. While much of the literature emphasizes meritocratic principles, where inequality arising from voluntary choices under equal opportunities is deemed acceptable, our results point to a critical distinction: the source of risk matters. In particular, we show that the standard dichotomy of ex-ante versus ex-post fairness (e.g., Cappelen et al., 2013) is insufficient when risk is resolved in a strategic context, which is what we called social risk. Spectators appear to recognize the vulnerability inherent in trust-based interactions, treating betrayal not simply as bad luck, but as a violation of an implicit social contract that warrants intervention. Furthermore, in contrast to prior work showing that spectators often over-attribute responsibility even for outcomes clearly shaped by external factors (e.g., Cappelen et al., 2022a,b; Bhattacharya and Mollerstrom, 2022; Andre, 2025; Yusof and Sartor, 2025), we identify a domain — social risk — where actors are seen as less culpable for negative outcomes. The willingness to insure against betrayal, even when the risk is anticipated, points to a deeper societal appreciation for efforts that foster trust and cooperation.[25] Future models of social preferences and distributive justice should incorporate the strategic environment and relational dimension of risk-taking, moving beyond purely individualistic notions of responsibility.

Second, our findings have direct policy relevance, particularly in strategic contexts such as contracts, employment, and trade, where outcomes depend on the actions of multiple parties. The heightened preference for redistribution in trust-based environments points to public demand for social insurance mechanisms that address harms arising from social, as opposed to purely idiosyncratic or market-based, risks. This may be especially relevant in settings characterized by weak institutions and widespread corruption, where wealth is often perceived as stemming from selfishness rather than merit (Almås et al., 2022). In such environments, the prevalence of insurer-type preferences may translate into broader support for policies aimed at mitigating the fallout from opportunistic behavior, including enhanced consumer protection, compensation for fraud victims, and community-based precautionary measures against broken informal agreements. Designing institutions that not only deter breaches of trust but also provide remedies when they occur may be essential to fostering economic activity in interdependent settings.

[25] This view resonates with broader concerns about economic systems, where many believe that productivity gains should lead to shared prosperity rather than concentrated wealth among a powerful few. Violations of this expectation — particularly through selfishness or exploitation — are often perceived as morally objectionable.

In conclusion, our paper shows that the structure of economic risk — particularly the presence of social interdependence and the potential for betrayal — is a critical determinant of fairness judgments and redistributive preferences. To accurately capture public sentiment and design effective economic policies, it is not enough to consider choice and luck in isolation. The social context in which outcomes are generated profoundly shapes their perceived legitimacy. Recognizing these nuanced moral considerations is vital for fostering cooperation, building resilient economic systems, and promoting a distribution of resources that commands broad social acceptance.

References

Aimone, J. and Houser, D. (2012). What you don't know won't hurt you: A laboratory analysis of betrayal aversion. *Experimental Economics*, 15(4):571–588.

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

Almås, I., Cappelen, A. W., Sørensen, E. Ø., and Tungodden, B. (2022). Global evidence on the selfish rich inequality hypothesis. *Proceedings of the National Academy of Sciences*, 119(3):e2109690119.

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.

Amasino, D. R., Pace, D. D., and van der Weele, J. J. (2024). Fair shares and selective attention. *American Economic Journal: Microeconomics*, 16(4):259–90.

Andre, P. (2025). Shallow meritocracy. *Review of Economic Studies*, 92(2):772–807.

Barr, A., Miller, L., and Ubeda, P. (2016). Moral consequences of becoming unemployed. *Proceedings of the National Academy of Sciences*, 113(17):4676–4681.

Bauer, M., Blattman, C., Chytilová, J., Henrich, J., Miguel, E., and Mitts, T. (2016). Can war foster cooperation? *Journal of Economic Perspectives*, 30(3):249–74.

Bhattacharya, P. and Mollerstrom, J. (2022). Lucky to work. Technical Report 22–46, George Mason University.

Bigoni, M., Bortolotti, S., Casari, M., and Gambetta, D. (2018). At the Root of the North–South Cooperation Gap in Italy: Preferences or Beliefs? *Economic Journal*, 129(619):1139–1152.

Bohnet, I., Greig, F., Herrmann, B., and Zeckhouser, R. (2008). Betrayal aversion: evidence from brazil, china, oman, switzerland, turkey, and the united states. *American Economic Review*, 98(1):294–310.

Bohnet, I. and Zeckhauser, R. (2004). Trust, risk and betrayal. *Journal of Economic Behavior & Organization*, 55(4):467–484.

Bortolotti, S., Soraperra, I., Sutter, M., and Zoller, C. (2025). Too Lucky to Be True: Fairness Views under the Shadow of Cheating. *Review of Economics and Statistics*, 107(3):771–785.

Cappelen, A. W., de Haan, T., and Tungodden, B. (2024). Fairness and limited information: Are people bayesian meritocrats? *Journal of Public Economics*, 233:105097.

Cappelen, A. W., Hole, A. D., Sørensen, E. Ø., and Tungodden, B. (2007). The Pluralism of Fairness Ideals: An Experimental Approach. *American Economic Review*, 97(3):818–827.

Cappelen, A. W., Konow, J., Sørensen, E. Ø., and Tungodden, B. (2013). Just Luck: An Experimental Study of Risk-Taking and Fairness. *American Economic Review*, 103(4):1398–1413.

Cappelen, A. W., Mollerstrom, J., Reme, B.-A., and Tungodden, B. (2022a). A meritocratic origin of egalitarian behaviour. *The Economic Journal*, 132(646):2101–2117.

Cappelen, A. W., Fest, S., Sørensen, E. Ø., and Tungodden, B. (2022b). Choice and personal responsibility: What is a morally relevant choice? *Review of Economics and Statistics*, 104(5):1110–1119.

Cassar, L. and Klein, A. H. (2019). A Matter of Perspective: How Experience Shapes Preferences for Redistribution. *Management Science*, 65(11):5050–5064.

Chetty, R., Hendren, N., Kline, P., Saez, E., and Turner, N. (2014). Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility. *American Economic Review*, 104(5):141–47.

Cohn, A., Lasse J., J., Marko, K., and Paul, S. (2023). Wealthy americans and redistribution: The role of fairness preferences. *Journal of Public Economics*, 225:104977.

Dohmen, T., Meyer, F., and Walkowitz, G. (2017). Satisfaction of Basic Needs as a Distributive Principle: Experimental Evidence on Prevalence and Relative Importance. Discussion paper.

Erkal, N., Gangadharan, L., and Nikiforakis, N. (2011). Relative Earnings and Giving in a Real-Effort Experiment. *American Economic Review*, 101(7):3330–48.

Fallucchi, F., Luccasen, R. A., and Turocy, T. L. (2019). Identifying discrete behavioural types: a re-analysis of public goods game contributions by hierarchical clustering. *Journal of the Economic Science Association*, (5):238–254.

Faravelli, M. (2007). How Context Matters: A Survey Based Experiment on Distributive Justice. *Journal of Public Economics*, 91(7–8):1399–1422.

Fehr, E., Epper, T., and Senn, J. (2024). Social preferences and redistributive politics. *Review of Economics and Statistics*, pages 1–45.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

Fong, C. (2001). Social Preferences, Self-interest, and the Demand for Redistribution. *Journal of Public Economics*, 82(2):225–246.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association*, 1(1):114–125.

Harrs, S. and Sterba, M.-B. (2025). Fairness and support for welfare policies. Mimeo.

Knack, S. and Keefer, P. (1997). Does social capital have an economic payoff? a cross-country investigation. *Quarterly Journal of Economics*, 112(4):1251–1288.

Koehler, J. J. and Gershoff, A. D. (2003). Betrayal aversion: When agents of protection become agents of harm. *Organizational Behavior and Human Decision Processes*, 90(2):244–261.

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

Konow, J. (2003). Which is the fairest one of all? a positive analysis of justice theories. *Journal of Economic Literature*, 41(4):1186–1237.

Litman, L., Robinson, J., and Abberbock, T. (2017). Turkprime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, 49(2):433–442.

Lobeck, M. and Støstad, M. N. (2025). The consequences of inequality: Beliefs and redistributive preferences. Mimeo.

Piketty, T. (2014). *Capital in the Twenty-First Century.* Harvard University Press.

Quercia, S. (2016). Eliciting and measuring betrayal aversion using the BDM mechanism. *Journal of the Economic Science Association*, 2(1):48–59.

Roth, C. and Wohlfart, J. (2018). Experienced inequality and preferences for redistribution. *Journal of Public Economics*, 167:251–262.

Selten, R. (1967). Die strategiemethode zur erforschung des eingeschränkt rationalen verhaltens im rahmen eines oligopolexperimentes. In *Beiträge zur experimentellen Wirtschaftsforschung*, pages 136–168. Sauermann, H.

Stantcheva, S. (2021). Understanding tax policy: How do people reason? *Quarterly Journal of Economics*, 136:2309–2369.

Train, K. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press.

Yusof, J. and Sartor, S. (2025). Market luck: Skill-biased inequality and redistributive preferences. Mimeo.

Appendix for Online Publication

A Additional Tables and Figures

Table A1: Implemented inequality (Gini index) by treatment and probability level

	Control			Natural-Risk			Social-Risk		
Pair	0%	25% - 75%	100%	0%	25% - 75%	100%	0%	25% - 75%	100%
Pairs without initial inequality									
100-100	0.03	0.03	0.03	0.05	0.04	0.05	0.02	0.02	0.03
300-300	-	0.04	0.05	-	0.04	0.04	0.03	0.02	0.02
Pairs with initial inequality									
300-0	-	0.39	-	-	0.37	-	0.43	0.33	0.29
100-0	0.86	0.72	-	0.71	0.65	-	0.63	0.49	0.52
300-100	-	0.44	0.56	-	0.41	0.50	0.38	0.40	0.46

Notes: "-" indicate cases that were ruled out by design. Higher numbers correspond to higher levels of implemented inequality.

Table A1 reports the implemented level of inequality (as measured by the Gini index) for each treatment and payoff pair, broken down by the two corner probabilities (0% and 100%) and the three intermediate probabilities (25%, 50%, and 75%). As noted in the main text, for the payoff pairs without initial inequality (*100-100* and *300-300*), the Gini index remains close to zero across all treatments and probability levels.

When comparing inequality between intermediate and corner probabilities, we observe some consistent patterns in Control and Natural-Risk. For instance, in the case where one stakeholder chose the safe option while the other chose the risky option and was unlucky (*100-0*), redistribution is lower (i.e., inequality is higher) when the risky option was a dominated choice, i.e., when the chance of receiving the high payoff was equal to 0%. Likewise, redistribution is also lower at the corner probabilities when the safe option was dominated, i.e., when the chance of receiving the high payoff was 100

For the Social-Risk treatment, we find that when both stakeholders decided to trust but only one was rewarded while the other was betrayed (*300-0*), redistribution is higher (and hence inequality lower) the more optimistic both stakeholders were about the trustworthiness of the (active) third party. When the two stakeholders chose different options, we observe that in cases where one was betrayed (*100-0*), redistribution toward the unlucky stakeholder is lower the less optimistic they were about the trustworthiness of the third party. In contrast, in the case where one played safe while the other chose to trust and was rewarded (*300-100*), spectators are less likely to redistribute the more optimistic both stakeholders were about the the trustworthiness of the third party.

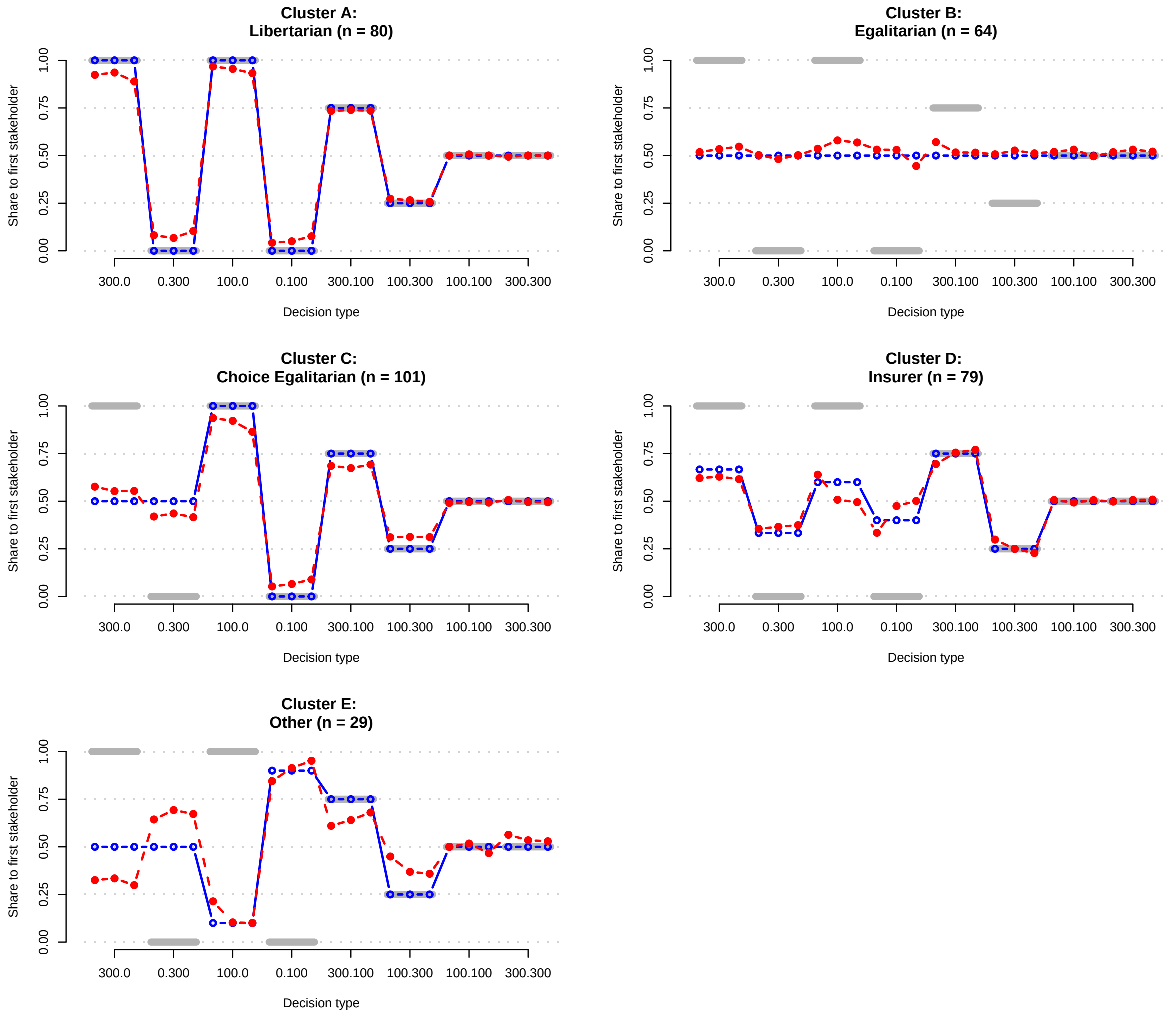
Overall, these results suggest that spectators are indeed sensitive with regard to extreme probabilities. In particular, they appear willing, at least to some extent, to punish those stakeholders whose choices can be considered as a mistake, for example, when choosing dominated options in the Control and Natural-Risk treatments, or when deciding (not) to trust when expectations about the trustworthiness of the trustee are low (high).

Table A2: Implemented inequality (Gini index) and probability levels

	Gini index			
	Control Model 1	Natural-Risk Model 2	Social-Risk Model 3	Pooled Model 4
Probability 50% (d)	-0.021	-0.022	-0.002	-0.021
	(0.013)	(0.015)	(0.015)	(0.013)
Probability 75% (d)	-0.031*	-0.020	0.002	-0.031*
	(0.017)	(0.015)	(0.015)	(0.017)
Natural-Risk (d)				-0.046
				(0.035)
Social-Risk (d)				-0.129***
				(0.035)
Natural-Risk × Probability 50%				-0.000
				(0.020)
Social-Risk (d) × Probability 50%				-0.019
				(0.020)
Natural-Risk × Probability 75%				-0.011
				(0.023)
Social-Risk (d) × Probability 75%				-0.034
				(0.022)
Constant	0.535***	0.489***	0.406***	0.535***
	(0.025)	(0.025)	(0.025)	(0.024)
Post-estimation F-tests				
Probability 50% - Probability 75% $= 0$	$p = .526$	$p = .899$	$p = .768$	$p = .524$
Nr. observations	2088	2160	2106	6354
Nr. clusters	116	120	117	353
R^2	0.000	0.000	0.000	0.016

Notes: OLS regressions with standard errors clustered at the individual level. Symbols ***, **, and * indicate significance at the 1%, 5% and 10% level, respectively.

Figure A1: k-means clustering - centroid and medoid of each cluster ($k = 5$)

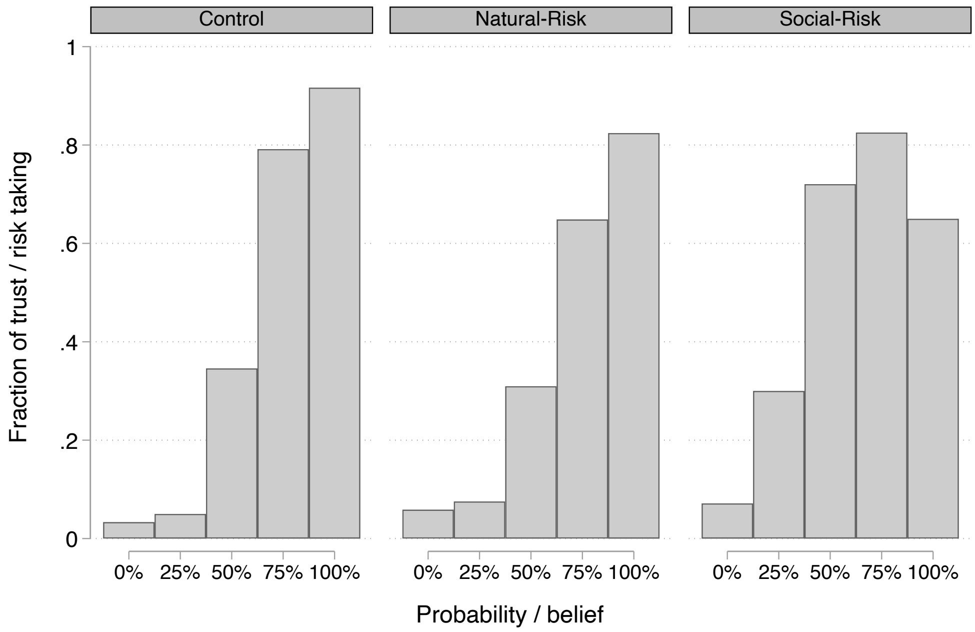


Notes: The x-axis shows the five payoff pairs with strictly positive earnings. The two numbers in each pair indicate the pre-redistribution payoffs of the two stakeholders, respectively. The y-axis shows the final share of the total payoff assigned to the first stakeholder after redistribution. The light-grey horizontal bars represent the share of resources of this stakeholder before redistribution. The red lines report the average share of resources allocated to the first stakeholder after redistribution, calculated using the centroid of the cluster, i.e., the average point in the cluster. The blue lines report the same data using the medoid of the cluster, i.e., the most central participant in the cluster. For each payoff pair, three dots are shown, corresponding to the three probability levels 25%, 50%, and 75%.

B Stakeholders' risk-taking and trusting behavior

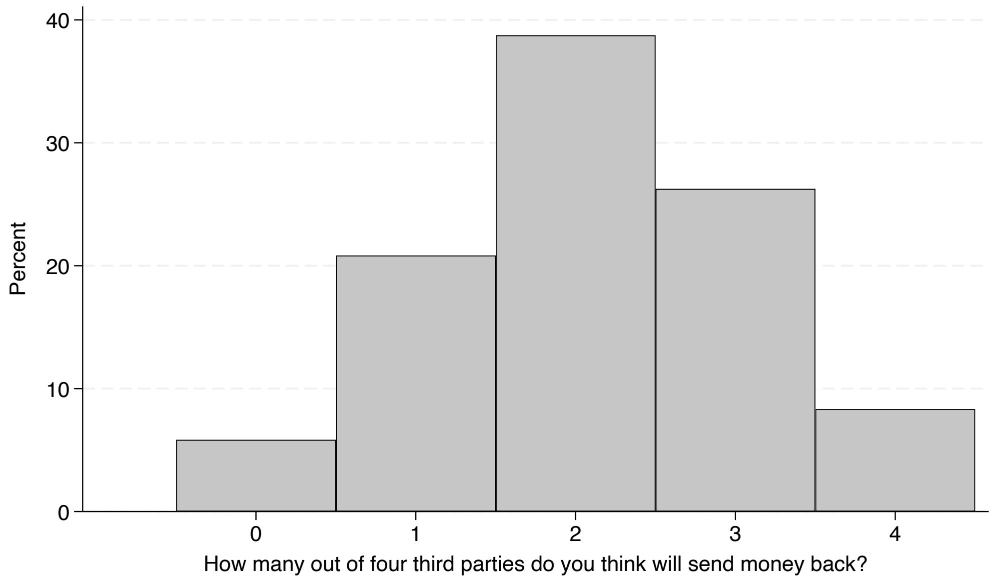
In this section, we describe the behavior of stakeholders in our online experiment. Our main results are summarized by Figure B1. It shows, separately for each treatment and each probability/belief level, the share of stakeholders choosing the risky option. In all three treatments, risk taking increases with the (perceived) probability of receiving the high outcome: the higher the probability, the larger the share of stakeholders selecting the risky option. Interestingly, the patterns for the Control and Natural-Risk treatment look quite similar, suggesting that efficiency concerns play only a minor role for the stakeholders' choices. In Social-Risk, in contrast, we observe a somewhat higher fraction of risk taking, in particular for intermediate levels of beliefs.

Figure B1: Stakeholders' choices by treatment and (expected) level of success



Recall that while the stakeholders' decision in the Control and Natural-Risk treatment were elicited using the strategy method — each stakeholder had to make five decisions, one for each probability level — stakeholders in the Social-Risk treatment made only a single decision on whether to trust. In addition, we elicited their beliefs about the trustworthiness of the (active) third party. Specifically, we asked them how many out of four participants in the role of the third party they think will send money back. The distribution of these beliefs is shown in Figure B2. The largest fraction of stakeholders (39%) believed that half of the third parties would chose to send back. In comparison, only 6% expected that no third party would be trustworthy, and only 8% expected all would be. On average, stakeholders believed that 52.5% (2.1 out of 4) of third parties would send back money, which is quite close to the actual share of third parties choosing to send money back (58%).

Figure B2: Distribution of stakeholders' beliefs in the Social-Risk treatment about the trustworthiness of third-parties



C Detailed procedure for the k-means cluster analysis

The k-means cluster analysis was performed with R using the "kmeans" function with Euclidean distance and 100,000 random initial values for the centroids of the clusters. The data used in the clustering are the 24-dimensional individual vectors of ex-post redistribution, standardized by the total amount to be redistributed.

To select the most appropriate number of clusters, k, we relied on a combination of different techniques. We started by using the elbow method, and since the results from this method were not conclusive, we complemented it with additional evidence —specifically, the silhouette method and an assessment of the robustness of specific clusters when increasing the size of k. Figure C1 shows the within-cluster variance as a function of the number of clusters k (elbow method). No clear discontinuity emerges from this analysis, which suggests that $k = 4$ or $k = 5$ may be the most appropriate number of clusters.

Since we cannot reach a clear-cut conclusion simply using the elbow method, we compute the silhouette distribution for $k = 4$ and $k = 5$. For each participant, the silhouette is computed as follows. First, we compute the distance between the participant's vector and the vectors of all other participants in their cluster. Second, the same process is repeated, but this time the participant's vector of choices is compared with the vector of choices of each of the participants in the next closest cluster. The individual silhouette index s_i in $[-1, 1]$ is calculated comparing these two measures. An s_i close to one means that the participant is closer to the other members in their own cluster than to the participants in the neighboring cluster. An s_i close to minus one means that the participant is closer to the neighboring cluster than their own cluster. Figure C2 reports the distribution of the silhouette index in each cluster for both $k = 4$ and $k = 5$. While the overall silhouette index performs well for both $k = 4$ and $k = 5$, we find that using five clusters provides a better classification of participants into homogeneous groups. In particular, the number of participants with negative silhouette values is lower when using five clusters. To further corroborate this choice, we repeated the cluster analysis for levels of k up to 10. The meaningful clusters identified in for $k = 5$ are robust to the addition of more clusters, suggesting that we have uncovered stable patterns.[26]

[26] Results for $k > 5$ are available upon request from the authors.

Figure C1: Within cluster variance (k-means cluster analysis)

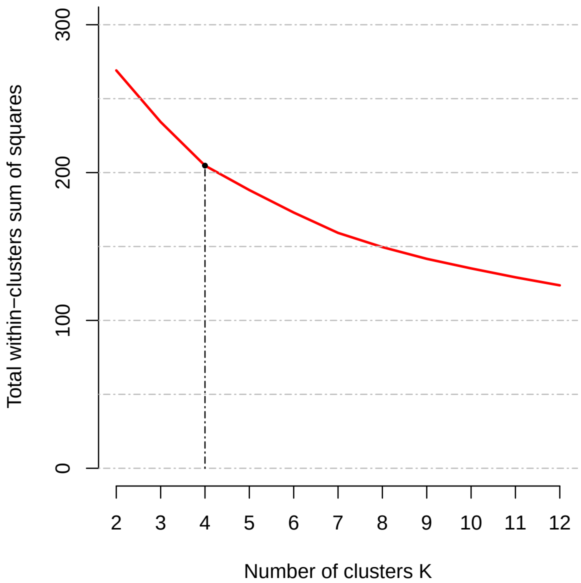
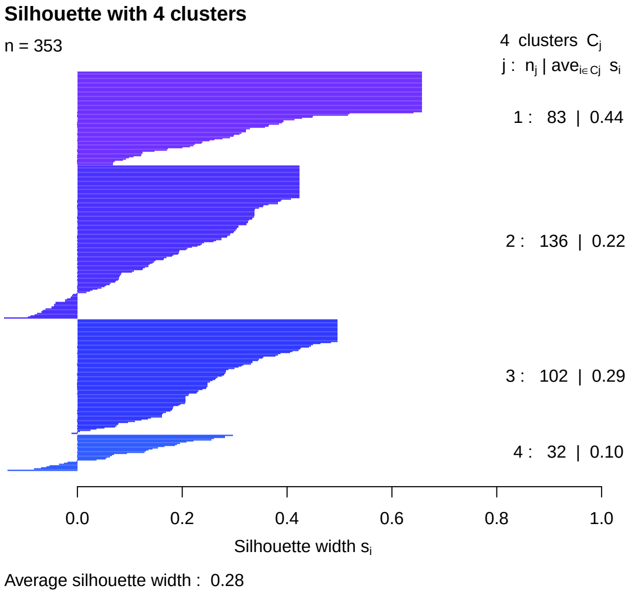
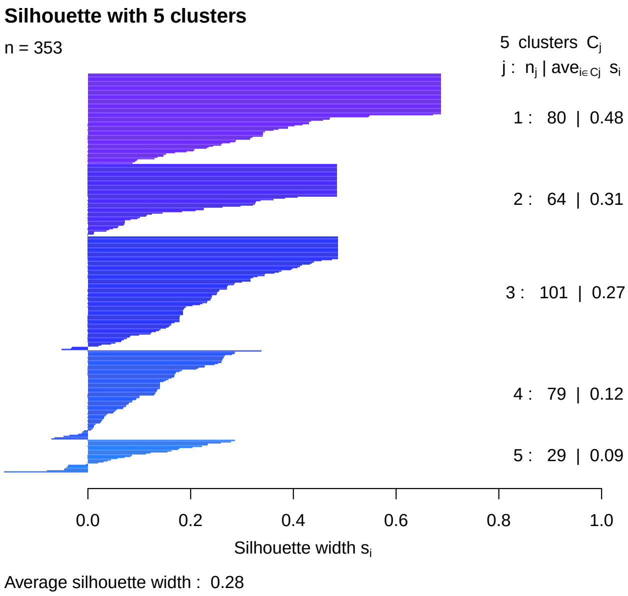


Figure C2: k-means clustering - silhouette plot ($k = 4$)

(a) $k = 4$



(b) $k = 5$



Notes: The figure reports the distribution of the individual silhouette index s_i in $[-1, 1]$. An s_i close to one indicates that a participant is more similar to other participants in their own cluster than to those in a neighboring cluster. Conversely, an s_i close to minus one indicates that a participant is closer to participants in a neighboring cluster than to those in their own cluster.

D Details on the estimation procedure

In this Appendix, we provide further details about the estimation of fairness views (based on Cappelen et al. 2007, 2013; Bortolotti et al. 2025).

Given the random utility model in equation (3) and under the assumption that ε_{iy} is i.i.d. extreme value distributed and that $\log(\beta)$ is $\mathcal{N}(\zeta, \sigma^2)$, we can write the likelihood contribution of a spectator i conditional on fairness view k as follows:

$$L_{i,k}(\zeta, \sigma) = \int_0^{\infty} \left(\prod_{j=1}^{j_i} \frac{e^{V(y_{ij}; F^k, \beta, \cdot)}}{\sum_{s \in \mathcal{Y}_{ij}} e^{V(s; F^k, \beta, \cdot)}} \right) f(\beta; \zeta, \sigma) d\beta \tag{D.1}$$

where $f(\cdot)$ is the density function of the log normal distribution and y_{ij} is the allocation chosen by spectator i from the choice set $\mathcal{Y}_{ij} = \{0, 10, \ldots, X_{ij}\}$ that spectator i faces in the redistribution decision j.

To calculate the total likelihood contribution of spectator i, we take the weighted sum of the conditional likelihood $L_{i,k}$

$$L_i(\lambda^L, \lambda^E, \lambda^{CE}, \lambda^I, \zeta, \sigma) = \sum_{k \in \{L, E, CE, I\}} \lambda^k L_{i,k} \tag{D.2}$$

where λ^k is the population share of spectators with fairness view $k \in \{L, E, CE, I\}$. k^L corresponds to *Libertarians* view, k^E corresponds to *Egalitarians* view, k^{CE} corresponds to *Choice Egalitarian*, and k^I corresponds to *Insurer* view. Finally, the total log-likelihood is obtained by taking the sum of the log of the total likelihood contributions of each spectator.

Parameters are estimated by maximum simulated likelihood with integrations perfomed using 100 Halton draws for each observation (Train, 2009). One population share and its standard error are calculated residually. The estimation is performed in R using the BFGS method with mle2 function (bbmle package). Table D1 reports the estimation results.

Table D1: Estimation of types

	Control Model 1	*Natural-Risk* Model 2	*Social-Risk* Model 3	*Social-Prob* Model 4
$\lambda^{Libertarians}$	0.356	0.321	0.272	0.352
	(0.047)	(0.046)	(0.044)	(0.047)
$\lambda^{ChoiceEgalitarians}$	0.389	0.313	0.188	0.263
	(0.049)	(0.047)	(0.039)	(0.046)
$\lambda^{Insurers}$	0.134	0.157	0.290	0.222
	(0.035)	(0.037)	(0.047)	(0.042)
$\lambda^{Egalitarians}$	0.121	0.209	0.250	0.162
	(0.033)	(0.040)	(0.044)	(0.036)
ζ	6.963	6.844	5.968	5.784
	(0.233)	(0.281)	(0.326)	(0.339)
σ	4.268	4.145	3.340	2.959
	(0.338)	(0.319)	(0.271)	(0.238)
logLik	-5214.866	-5838.841	-6033.670	-6317.867
Degrees of freedom	5	5	5	5

Notes: The likelihood is maximized in R using the BFGS method with mle2 function (bbmle package). One population share and its standard error are calculated residually. Numerical integration is perfomed using 100 halton draws for each observation (Train, 2009). Models 1 to 4 report estimates separately by treatment: *Control*, *Natural-Risk*, *Social-Risk*, and *Social-Prob*. The log-likelihood of the pooled models are the following: (i) pooling *Social-Risk* and *Social-Prob*: -12354.840; (ii) pooling *Social-Risk* and *Control*: -11261.354; (iii) pooling *Social-Risk* and *Natural-Risk*: -11879.003; (iv) pooling *Social-Prob* and *Control*: -11542.438; (v) pooling *Social-Prob* and *Natural-Risk*: -12163.068; and (vi) pooling *Control* and *Natural-Risk*: -11055.570.

E Experimental instructions for spectators (*translated from German*)

The following instructions are for the Natural-Risk treatment. The instructions for the other treatments are very similar and available upon request.

Welcome to this experiment. The purpose of this study is to investigate how people make decisions in certain contexts. From now until the end of the study, any communication with other participants is not allowed. If you have a question, please raise your hand. One of the experimenters will then come to your desk and answer your question. Upon completion of the study, you will receive a payment of €10, including €4 show-up fee.

Overview. In this study, you will be presented with a total of 35 decisions, one after the other. In each decision, your task is to decide how to redistribute money between an Orange and a Blue player. One of these decisions will have real monetary consequences for two individuals that we recruited via an international online marketplace to conduct an assignment. In the following, we will first explain in detail the task we gave to the individuals who participated in the online assignment. After that, we will provide you with the information about your task.

Online Study. A few days ago we recruited participants via an international online marketplace to conduct an assignment. For their participation, participants received a fix payment of $0.50. On top of that, participants could earn an additional amount of money (see below).

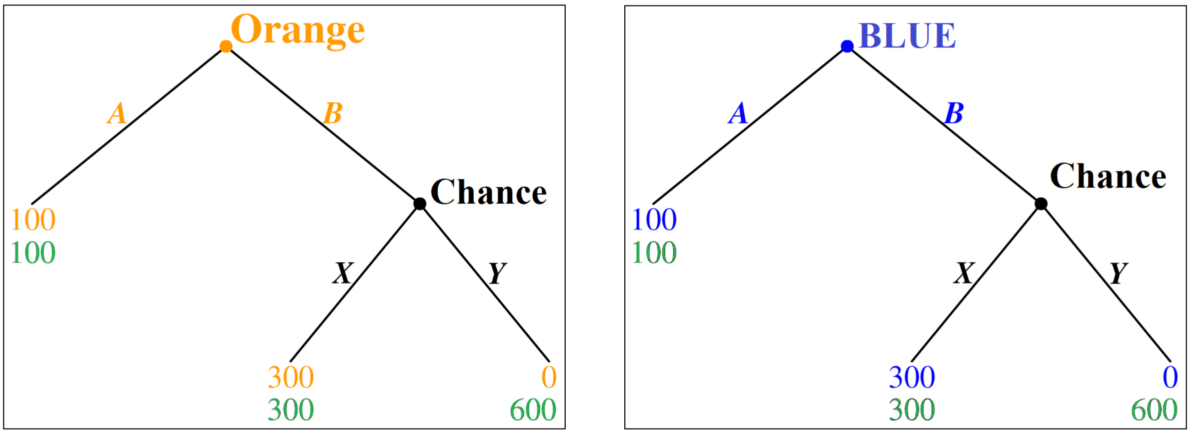
In the study there were two pairs of players. The first pair consisted of an ORANGE and a GREEN player. The second pair consisted of a BLUE and a GREEN player. Please note that ORANGE and BLUE interacted with two different GREEN players. The decision situation faced by the ORANGE, the BLUE, and the two GREEN players is depicted in Figure E1.

Please consider first the pair on the left side of Figure E1.

At first, ORANGE could choose between two options, Option A and Option B.

- **Option A.** If ORANGE chose Option A, then ORANGE and GREEN each earned 100 tokens with certainty. In this case, GREEN had no decision to make.
- **Option B.** If ORANGE chose Option B, then a random draw determined how earnings are allocated. There were two possible outcomes:
  - **Option X:** ORANGE and Green each earned 300 tokens;

Figure E1: Pairs and options



– **Option Y:** ORANGE earned 0 tokens and Green earned 600 tokens.

The second pair with the BLUE and the other GREEN player were facing the exact same decision situation as the first pair (see right side of Figure E1). BLUE could choose between Option A and Option B, and in case BLUE chose Option B, a random draw decided between Option X and Option Y.

In total, ORANGE and BLUE had to make five decisions (see Table E1). Option A was the same for all five decisions. Option B differed in each of the five decision. In particular, the probability that Option X was randomly drawn increased from decision to decision in steps of 25%. That is, in the first decision, if Option B was chosen, Option X had a 0% chance of being randomly drawn and Option Y had a 100% chance of being randomly drawn. In the second decision, Option X had a 25% chance of being randomly drawn and Option Y had a 75% chance of being randomly drawn. In the third decision, Option X had a 50% chance of being randomly drawn and Option Y had a 50% chance of being randomly drawn, and so on. Finally, in the fifth decision, Option X had a 100% chance of being randomly drawn and Option Y had a 0% chance of being randomly drawn.

All amounts in the study were initially expressed in tokens. At the end of the study, tokens were exchanged at the rate of \$1=150 tokens. Please notice that the amount of money at stake is above the average earnings compared to similar tasks at the same online marketplace. Participants were allowed to take part in the study only if they correctly answered a set of control questions.

After completing the study, we randomly formed pairs of one ORANGE and one BLUE player. After that, participants were told that a third person would be informed about the rules and the outcome of the study, and that this person would be given the opportunity to redistribute the earnings between the ORANGE and the BLUE player, and thus determine how much money they would be paid for the online study.

Table E1: Online decisions

Decision	Option A	Option B	
		Option X	Option Y
1	Orange/Blue:100, Green:100	**0% chance** Orange/Blue:300, Green:300	**100% chance** Orange/Blue:0, Green:600
2	Orange/Blue:100, Green:100	**25% chance** Orange/Blue:300, Green:300	**75% chance** Orange/Blue:0, Green:600
3	Orange/Blue:100, Green:100	**50% chance** Orange/Blue:300, Green:300	**50% chance** Orange/Blue:0, Green:600
4	Orange/Blue:100, Green:100	**75% chance** Orange/Blue:300, Green:300	**25% chance** Orange/Blue:0, Green:600
5	Orange/Blue:100, Green:100	**100% chance** Orange/Blue:300, Green:300	**0% chance** Orange/Blue:0, Green:600

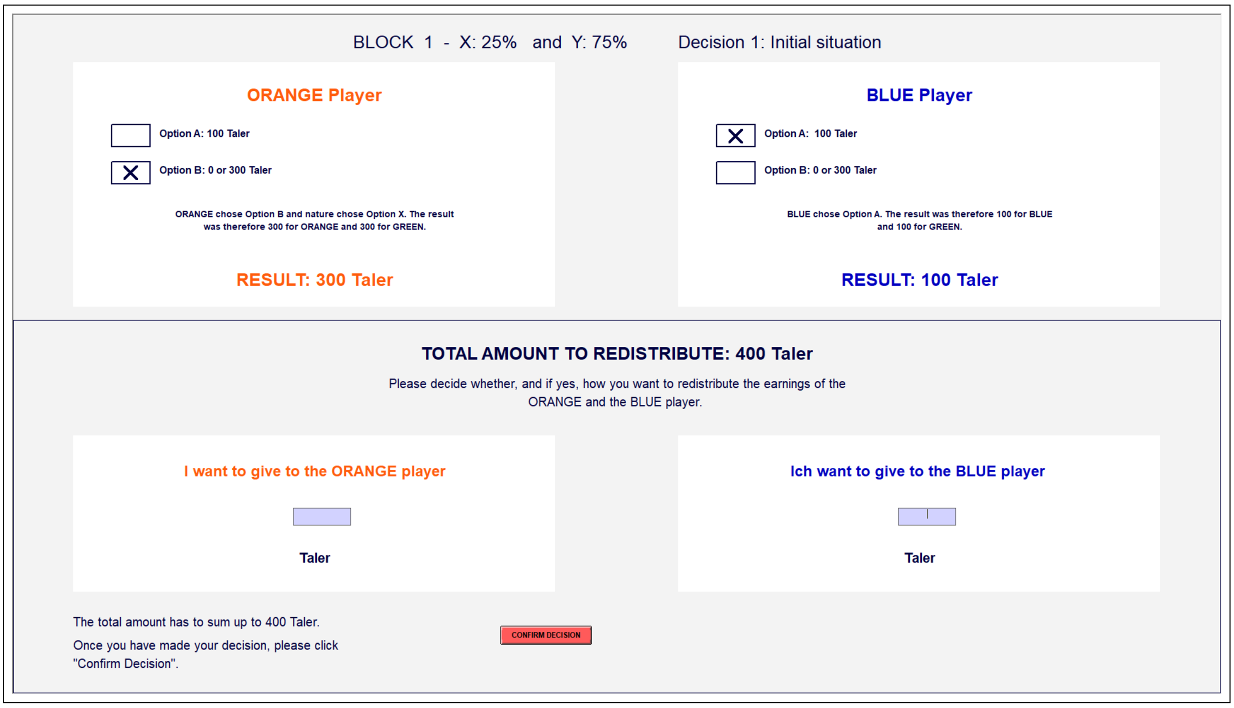
Your Task. You are the third person and you can now choose whether to redistribute the tokens between the ORANGE and the BLUE player. Please notice that you can only redistribute money between ORANGE and BLUE, but not any tokens received by GREEN. The participants of the online study will receive the payment that you choose for them within the next few days. The participants will not receive any additional information (apart from their final earnings). Your decision is completely anonymous.

Figure E2 shows a sample decision screen. In the upper part of the screen you can see the initial situation for the ORANGE and the BLUE player. For each player, you can see whether they chose Option A or Option B. In case any of the players chose Option B, you can also see which option (Option X or Option Y) was implemented by the random draw. You will further see how many tokens each of the players received.

In the example in Figure E2, ORANGE chose Option B and the random draw implemented Option X. Therefore, ORANGE and GREEN received 300 tokens each. BLUE, in contrast, chose Option A. As a result, he and GREEN received 100 tokens each.

In the central part of the screen you can see the sum of the tokens received by the ORANGE and the BLUE player. In the example, the sum of tokens is 400. **Your task is to decide whether and how to redistribute the total amount of tokens between ORANGE and BLUE.** You can choose any positive amount in steps of 10 tokens, as long as you redistribute all tokens. In our example, the sum of what you give to ORANGE and BLUE must be exactly 400 tokens.

Figure E2: Sample screen shot



Decisions and blocks. You have to make 35 decisions–one for each possible case– from which one will be payoff-relevant. This means that one decision will have actual monetary consequences for two individuals who completed the online assignment. You won't know in advance which decision will be relevant for the earnings of the participants of the online study. Please make a considerate decision in each situation.

The 35 decisions will be divided into **5 blocks**. The blocks differ with regard to the chance that Option X is implemented by the random draw (in case ORANGE or BLUE chose Option B). For instance, the block **X: 25%, Y: 75%** combines all decision situations in which the chance of Option X being implemented by the random draw was 25%, and the chance of Option Y being implemented by the random draw was 75%. The block **X: 50%, Y: 50%**, in contrast, combines all decision situations in which the chance of Option X being implemented by the random draw was 50%, and the chance of Option Y being implemented by the random draw was 50%. Please notice that the chance of Option X being implemented is the same for both ORANGE and BLUE and that it is the same for all decision within this block. You will be informed about the relevant probability at the beginning of each block. At the end of each block, you will see a summary of your decisions and you will have the opportunity to revise them.

Before we start, we ask you to answer a few control questions. These questions should ensure, that all participants have understood the instructions. After that, the experiment starts.

F Experimental instructions for stakeholders (*MTurk*)

The following instructions are for the first movers in the Natural-Risk treatment. The instructions for the other treatments are very similar and available upon request.

Welcome!

Thank you for participating in our HIT.

This HIT comprises two stages. If you complete the HIT, you will earn a fixed amount of \$0.50 plus a bonus that depends on your and others' choices. All earnings are expressed in tokens that will be converted into real money at the end of the study (\$1=150 tokens). The study will take about 8 minutes to complete (including the time for reading the instructions). You will receive a code to collect your payment via MTurk upon completion.

If you want to participate, please enter your MTurk worker ID and proceed to the instructions for Stage 1.

$\Longrightarrow$ ——— *new screen* ——— $\Longleftarrow$

Stage 1

In Stage 1, you will be paired with another worker on Mturk. In the following, we will call this other Mturker your 'Partner'. In Stage 1 you will face five decisions. In each decision, you have to choose between two options: **Option A** and **Option B** (see table below).

Decision	Option A	Option B	
1	You:100, Partner:100	0% chance You:300, Partner:300	100% chance You:0, Partner:600
2	You:100, Partner:100	25% chance You:300, Partner:300	75% chance You:0, Partner:600
3	You:100, Partner:100	50% chance You:300, Partner:300	50% chance You:0, Partner:600
4	You:100, Partner:100	75% chance You:300, Partner:300	25% chance You:0, Partner:600
5	You:100, Partner:100	100% chance You:300, Partner:300	0% chance You:0, Partner:600

- **Option A**: Option A is the same for all five decisions and pays you and your partner 100 tokens each for sure.

- **Option B**: Option B is risky and there are two possible outcomes:
  - You and your partner earn 300 tokens each, or
  - You earn 0 tokens and your partner earns 600 tokens.

The realized outcome depends on a random draw and the chances of reaching one of the two outcomes change in each of the five decisions. The chance of you and your partner earning 300 tokens each is 0% in decision 1, 25% in decision 2, 50% in decision 3, 75% in decision 4, and 100% in decision 5. At the same time, the chance of you earning 0 tokens and your partner earning 600 tokens is 100% in decision 1, 75% in decision 2, 50% in decision 3, 25% in decision 4, and 0% in decision 5.

After you have made a choice for each decision, the computer will randomly select one decision that will be relevant for Stage 2 and your and your partner's earnings. If you choose Option B for a given decision, the computer will resolve the lottery.

If you have understand these instructions, please answer the control questions below.

Question 1: Suppose you chose **Option A** in **decision 3**. What is the outcome of this decision?

- ☐ The outcome is: You:100, Partner:100
- ☐ The outcome is: You:200, Partner:200
- ☐ There is a 50% chance that the outcome is: You:300, Partner:300 and a 50% chance that the outcome is: You:0, Partner:600
- ☐ There is no bonus for sure

Question 2: Suppose you chose **Option B** in **decision 3**. What is the outcome of this decision?

- ☐ The outcome is: You:100, Partner:100
- ☐ The outcome is: You:200, Partner:200
- ☐ There is a 50% chance that the outcome is: You:300, Partner:300 and a 50% chance that the outcome is: You:0, Partner:600
- ☐ There is no bonus for sure

⟹ ——— *new screen* ——— ⟸

Stage 2

Thank you for completing Stage 1 of the study. We will now explain Stage 2. In Stage 2 you will be randomly matched with another worker on Mturk, who has completed the exact same HIT in the exact same role as you.

A third person will be informed about the assignment and about your choice and the other worker's choice. In case you or the other worker chose Option B, the third person is also informed about the outcome of the random draw. This third person will therefore be informed about how many tokens you and the other Mturker earned in Stage 1.

The third person will then be given the opportunity to redistribute tokens between you and the other worker. That is, the third person can either:

- Transfer any token amount from you to the other Mturker,
- transfer any token amount from the other Mturker to you, or
- leave the token earnings from you and the other Mturker from Stage 1 unchanged.

The redistribution done by the third person will determine your bonus for the present assignment. You will receive your bonus within one week from the completion of the assignment.

In the following, we kindly ask you to answer the following questions about Stage 2:

Consider two Mturkers – worker 1 and worker 2 – who completed the the exact same HIT in the exact same role as you.

Question 1:

Suppose in decision 3 **worker 1** selected **Option A** - and his and his partner's outcome was 100 tokens. Suppose further that **worker 2** chose **Option B** and the outcome of the random draw was 0 tokens for himself and 600 tokens for his partner. The third person now has the opportunity to redistribute tokens between worker 1 and worker 2. How do you think the tokens will be redistributed? (Note: The distributed tokens must sum up to 100 tokens.)

- Amount of tokens worker 1 will receive: [*blank*]
- Amount of tokens worker 2 will receive: [*blank*]

- Total [*automatically filled*]

Question 2:

Suppose **worker 1** selected **Option A** and his and his partner's outcome was 100 tokens. Suppose further that **worker 2** chose **Option B** and the outcome of the random draw was 300 tokens for himself and 300 tokens for his partner. The third person now has the opportunity to redistribute tokens between worker 1 and worker 2. How do you think the tokens will be redistributed? (Note: The distributed tokens must sum up to 400 tokens.)

- Amount of tokens worker 1 will receive: [*blank*]
- Amount of tokens worker 2 will receive: [*blank*]
- Total [*automatically filled*]

Question 3:

Suppose **worker 1** selected **Option B** and the outcome of the random draw was 0 tokens for himself and 600 tokens for his partner. Suppose further that **worker 2** chose **Option B** and the outcome of the random draw was 300 tokens for himself and 300 tokens for his partner. The third person now has the opportunity to redistribute tokens between worker 1 and worker 2. How do you think the tokens will be redistributed? (Note: The distributed tokens must sum up to 300 tokens.)

- Amount of tokens worker 1 will receive: [*blank*]
- Amount of tokens worker 2 will receive: [*blank*]
- Total [*automatically filled*]

$\Longrightarrow$ ——— *new screen* ——— $\Longleftarrow$

Validation code

We thank you for your time spent taking this survey. Your response has been recorded.
Your MTurk completion code is <*code here*>

IMPORTANT: you need to enter this code on MTurk to collect your payments. After that, you may close the browser, window or tab