Current Inequality, Future Prospects, and Trust: Evidence from a Laboratory Experiment*

Marco Stimolo[1] and Alessandro Stringhi[2]

[1]University of Siena, Italy; Email: marco.stimolo@unisi.it
[1]Prague University of Economics and Business, Czech Republic; Email: alessandro.stringhi@vse.cz

Abstract

We report a pre-registered experiment on how reducing current inequality and improving future prospects affect trust and trustworthiness in interactions with disadvantaged counterparts. In a Trust Mini Game played over two periods, middle-endowment senders interact with either low-endowment ("Poor") or high-endowment ("Rich") receivers. Participants are assigned to one of three conditions: (i) a High-inequality Control; (ii) Low inequality, which halves the endowment gap; or (iii) Rank reversal, which preserves within-period inequality while introducing a stochastic reversal of receivers' rank across periods. The two treatments are calibrated to imply the same expected intertemporal endowment inequality. For the first-period behavior, before any rank reversal is realized, Fehr and Schmidt's inequity-aversion models and our original extension allowing future prospects to influence the perception of current inequality, both predict stronger effects under Low inequality than under Rank reversal. Consistently, Low inequality increases senders' trust and expectations of reciprocation in poor-receiver matches, and poor receivers are correspondingly more trustworthy. Rank reversal, by contrast, yields estimates close to zero.

JEL classification: C92, D31, D63.

Keywords: Interpersonal trust, current economic inequality, future prospects.

*This project has been co-funded by the European Regional Development Fund through the project "Center for Inequality and Open Society," no. CZ.02.01.01/00/23_025/0008690. We are grateful to Jan Kolin and Nina Muráňová for their assistance with data collection. We also thank Moti Michaeli, Pierpaolo Parrotta, Paolo Pin, Alberto Dalmazzo, Luigi Mittone, Francesco Parisi, and Antonio Filippin for helpful comments and suggestions. Finally, we thank participants at the Officina Dati seminar at the University of Siena (2024), the BEEN Workshop at the University of Turin (2024), the ASFEE Conference in Grenoble (2024), and the SEET Conference in Sevilla (2026). The usual disclaimers apply.

1 Introduction

The attitude to entrust unknown individuals in one-shot interactions (generalized trust Tabellini, 2008) is widely viewed as a key determinant of economic performance and social cooperation (e.g., Algan and Cahuc, 2013). This has motivated sustained interest in its economic determinants, among which inequality features prominently (Stringhi, 2026). A large literature shows that wider income gaps are associated with weaker social cohesion, lower trust, and less cooperative behavior (e.g., Alesina and La Ferrara, 2002; Ananyev and Guriev, 2019; D'Amato et al., 2021). These effects may be particularly important in interactions involving disadvantaged individuals, because being worse off than one's counterpart may generate stronger resentment than being better off generates unease. Trust may therefore be especially fragile when one party is worse off than the other, suggesting that reducing inequality may matter not only for distributive reasons, but also for sustaining social cooperation.

Yet relative disadvantage can be viewed through at least two distinct lenses. One focuses on current inequality, that is, on the actual resource gap between interacting individuals. The other focuses on future prospects, that is, on the disadvantaged individual's chance of moving to a better condition over time. These two features need not coincide. This raises a question for which comparative causal evidence remains limited: in trust interactions with disadvantaged counterparts, do individuals respond more strongly to a narrower current resource gap or to better future prospects?

This distinction is relevant in contexts where disadvantage is currently present, but institutions are in place to expand opportunities for future improvement. For instance, access to public education can expand disadvantaged individuals' opportunities to improve their future economic conditions, even if it does not alter their current economic circumstances. Empirical evidence from school-construction reforms and randomized reductions in the cost of secondary schooling shows that expanding educational access can raise later educational attainment and labor-market outcomes (Duflo, 2001; Duflo et al., 2021). A similar logic applies when disadvantaged children gain access to better schools through admission lotteries, or when low-income families are given opportunities to move to better neighborhoods through randomized housing-mobility programs. In such cases, institutions expand the set of feasible future trajectories without necessarily eliminating current deprivation at the time the opportunity is granted (Deming et al., 2014; Chetty et al., 2016; Bergman et al., 2024). These interventions therefore modify relative disadvantage in an intertemporal sense: they enlarge the scope for improving one's own position across periods, and may reduce intertemporal inequality even when within-period inequality is unchanged (Jäntti and Jenkins, 2015). At the same time, improved future prospects need not render current relative disadvantage behaviorally irrelevant. In particular, individuals may continue to evaluate themselves relative to their current relative condition (Cullen et al., 2006; Pop-Eleches and Urquiola, 2013). Consequently, it is not obvious that expanding *ex ante* opportunities without reducing current inequality will have beneficial effects on trust and cooperation.

Identifying these effects in field data is difficult, because inequality, trust, and the institutional environment are jointly shaped by broader economic and social conditions, and because current disparities and future prospects often change together. As a result, it is hard to isolate whether trust-based interactions respond more strongly to a smaller current resource gap or to better *ex ante* future prospects. We therefore conduct a pre-registered laboratory experiment that cleanly isolates and compares these two behavioral channels in interactions involving disadvantaged counterparts, abstracting from other features of real-world environments.

We employ a Trust Mini Game with heterogeneous endowments, capturing a stylized interaction between a middle-endowment ("Middle Class") sender and either a low-endowment ("Poor") or a high-endowment ("Rich") receiver. Because both players make dichotomous choices, the design provides a tractable and transparent setting for identifying treatment effects on trust and trustworthiness. Each round is a one-shot, anonymous interaction with random matching. The game is played over two independent periods, implemented as two blocks of ten rounds each; the second block is purely instrumental to generate variation in future prospects.

We implement a between-subject design featuring a *High Inequality* control condition (T0) and two treatments. In the *Low Inequality* treatment (T1), the endowment gap between the sender and the receiver is half as large as in the control. In the *Rank Reversal* treatment (T2), by contrast, within-period endowment inequality is the same as in T0, but receivers' relative positions may reverse in the second period: a "Poor" receiver has a 50% probability of becoming "Rich" in period 2, and *vice versa*. This reshuffling reduces the expected dispersion of endowments across the two periods relative to T0, while leaving current inequality unchanged in each period. Importantly, T1 and T2 are calibrated to imply the same expected dispersion of endowments over the two periods of the game. We therefore focus on first-period behavior, where the design allows us to compare the behavioral effects of a narrower current resource gap with those of better *ex ante* prospects under unchanged current inequality, before any rank reversal is realized. Because our main interest is in interactions involving disadvantaged counterparts, our analysis focuses on matches with a "Poor" receiver. Matches with a "Rich" receiver serve primarily as a benchmark, and the game is calibrated so that the inequality manipulations in T1 and T2 are not expected to affect their behavior.

This simplified environment allows us to derive sharp, testable predictions from the inequity-aversion framework of Fehr and Schmidt (1999). The intuition is that, under the Fehr–Schmidt benchmark, individuals evaluate inequality in the realized payoff allocation generated by the current interaction. Accordingly, the *Low Inequality* treatment (T1), which reduces the current sender–receiver gap respect to the *High Inequality* T0, should make "Poor" receivers more willing to reciprocate and should therefore increase Senders' expectations of reciprocation and willingness to trust. By contrast, the *Rank Reversal* treatment (T2) introduces prospects of improving "Poor" receivers' condition, while leaving current inequality unchanged. Therefore the Fehr–Schmidt benchmark is not suited to derive predictions in our one-shot anonymous setting.

Inspired by extensions of the Fehr–Schmidt framework that account for the behavioral relevance of reductions in *ex-ante* inequality (Saito, 2013; Cappelen et al., 2013), we introduce a forward-looking version of the inequity-aversion model, in which future prospects shape the perception of current inequality. We adapt the Fehr–Schmidt model to an intertemporal environment with one-shot anonymous interactions that allows expected future relative conditions to enter the reference point used to evaluate current payoff inequality. This extension is central to derive clean testable predictions for the *Rank Reversal* treatment. Under this formulation, in the *Rank Reversal* treatment (T2), a "Poor" receiver may perceive period 1 disadvantage as less persistent. Hence, she may experience a smaller disutility from disadvantageous inequality relative to the *High Inequality* Control (T0). This may increase poor receivers' willingness to reciprocate and, through Senders' expectations, increase trust. The framework predicts that T2 should lie between the other two conditions: if future prospects are ignored, T2 coincides with T0; instead, if they are fully incorporated, T2 coincides with T1.

Consistent with our conceptual framework and design calibration, statistically significant treatment effects are concentrated in matches with a "Poor" receiver. In the main random-effects Probit specifications, *Low Inequality* increases senders' willingness to trust poor receivers by about 14.1 percentage points (p.p.) relative to the *High Inequality* Control and by about 18.4 p.p. relative to *Rank Reversal*. In the corresponding linear panel regressions, senders' first-order beliefs move in the same direction: expected cooperation by poor receivers rises by about 10 p.p. under *Low Inequality*, whereas *Rank Reversal* leaves beliefs essentially unchanged relative to the Control. Poor receivers' behavior mirrors this pattern. In the Probit specifications, *Low Inequality* increases poor receivers' probability of reciprocate the sender's trust by about 19.7 p.p. relative to the Control and by about 21 p.p. relative to *Rank Reversal*. In contrast, *Rank Reversal* yields estimates close to zero relative to the Control. In matches with "Rich" receivers, trust, beliefs, and trustworthiness are essentially stable across conditions. The main mechanism appears to operate through an increase in poor receivers' trustworthiness under *Low Inequality*, which in turn raises senders' expectations of reciprocation and, consequently, their willingness to trust.

Overall, reducing within-period inequality under the *Low Inequality* treatment strengthens trust and trustworthiness in interactions with disadvantaged counterparts, whereas improving future prospects under stochastic *Rank Reversal* yields estimates close to zero. Our experiment separates a current-inequality channel from a future prospects channel. The former operates through the payoff disparities individuals face at the time of interaction; the latter operates through the possibility that current disadvantage may be attenuated in the future. The results indicate that trust and trustworthiness respond primarily to current inequality. In this environment, reducing the present endowment gap strengthens cooperation, whereas equalizing expected conditions over time through stochastic rank reversal does not generate comparable gains.

This study contributes to the experimental literature on inequality, mobility, and social behavior by providing direct causal evidence on how lower current inequality and improved *ex ante* mobility prospects shape trust and trustworthiness in interactions involving disadvantaged counterparts. A growing body of laboratory work examines the prospect of upward mobility in political-economy settings, showing that mobility can reduce support for redistribution, especially when fiscal policy is persistent across periods (Agranov and Palfrey, 2020). Other experiments study mobility in broader social environments, finding that it can reduce antisocial behavior (Gangadharan et al., 2021), possibly by alleviating frustration arising from limited economic opportunities, and that temporary role reversal may induce social sympathy and thereby increase cooperation in public-goods settings (Lange et al., 2021). Our experiment shifts the focus to the behavioral determinants of trust and trustworthiness under inequality through a deliberately abstract design. It does not reproduce the institutional complexity of redistribution or mobility policies, but instead isolates two core behavioral channels: a reduction in the current resource gap faced by individuals in the interaction, and an improvement in initially disadvantaged receivers' expected future position. Moreover, by concentrating on first-period behavior before rank reversal is realized, we limit the role of emotional responses — such as disappointment, anger, relief, or positive surprise — that may arise once individuals actually move up or down. This allows us to compare the effect of a smaller current resource gap with that of improved *ex ante* prospects before realized reversal introduces additional emotional or history-dependent responses.

Experimental work on inequality and trust largely relies on the standard trust game, in which trusting typically involves a continuous transfer from the Sender to the Receiver (e.g., Anderson et al., 2006; Hargreaves Heap et al., 2013; D'Amato et al., 2021). Because this transfer changes both players' resource positions before the Receiver responds, the act of trusting mechanically affects the *ex post* distribution of payoffs. This makes it harder to distinguish trust as an expectation of reciprocation from a willingness to redistribute resources. Our methodological contribution is to use a Trust Mini Game together with a parametrization that sharpens the interpretation of behavior. In our setting, the Sender's decision is primarily a strategic response to expected reciprocation, and the Receiver's decision is a binary choice to honor or exploit trust. In addition, because the Sender's endowment is held fixed across treatments and trusting does not mechanically compress initial inequality, differences in trust and trustworthiness can be interpreted more directly as responses to the inequality environment itself.

The rest of the paper is organized as follows. Section 2 illustrates the features of the experimental design together with the behavioral expectations. Section 3 show the our results. Section 4 concludes.

2 Experimental Design

This section presents the experimental design and explains how it identifies the distinct behavioral effects of a smaller current endowment gap and improved ex ante mobility prospects on trust and trustworthiness. It then derives the corresponding predictions from the inequity-aversion framework of Fehr and Schmidt (1999) and from a parsimonious forward-looking extension in which improved prospects in the future enter the evaluation of current payoff inequality.

2.1 Design

Our experimental setup consists of a Trust Mini Game designed to capture a stylized interaction between a "Middle Class" sender and a receiver who is either "Poor" or "Rich". The game is played over two independent periods. Participants are randomly and anonymously matched with different counterparts across periods. Since both periods are independently payoff-relevant, each interaction can be treated as one-shot. The second period is instrumental in introducing an intertemporal dimension of inequality. In our setting, this dimension arises through the possibility that receivers' relative conditions change across periods.

We implement a between-subject design with three conditions: (i) a *High Inequality* control (T0), featuring a substantial endowment gap between senders and receivers; (ii) a *Low Inequality* treatment (T1), in which the sender–receiver endowment gap is half as large as in the Control; and (iii) a *Rank Reversal* treatment (T2), which preserves the same within-period endowment gap as in T0 but allows receivers' relative ranks to reverse across periods: the "Poor" receiver has a 50% probability of becoming "Rich" in period 2, and *vice versa*. As a result, T1 and T2 imply the same expected intertemporal endowment gap across the two periods.

Focusing on first-period behavior allows us to compare the effects of two distinct ways of mitigating inequality before any rank change is realized. Relative to T0, T1 reduces the current sender–receiver endowment gap. By contrast, T2 leaves current inequality unchanged but improves initially "Poor" receivers' prospects through the possibility of becoming "Rich" in the second period.

Rules of the Trust Mini Game. The stage game between the Sender and the Receiver is illustrated in Figure 1. The Sender is endowed with $w = 6$ Experimental Currency Units (ECUs), and this endowment is held fixed across periods and treatments. The Receiver may instead have either a lower or a higher endowment than the Sender. We label these two receiver types "Poor" and "Rich", respectively, and receivers are evenly split between the two types. The Sender therefore occupies an intermediate position, which we label "Middle Class".

The Receiver's endowment varies across treatments ($T \in \{T0, T1, T2\}$). In the *High Inequality* control (T0) and in the *Rank Reversal* treatment (T2), the "Poor" receiver is endowed with $w_P^{T0} = w_P^{T2} = 2$ and the "Rich" receiver with $w_R^{T0} = w_R^{T2} = 10$. In the *Low Inequality* treatment (T1), the endowment gap is halved: $w_P^{T1} = 4$ and $w_R^{T1} = 8$. In all treatments, receiver endowments remain symmetric around the Sender's endowment, so that $|w - w_P^T| = |w - w_R^T|$.

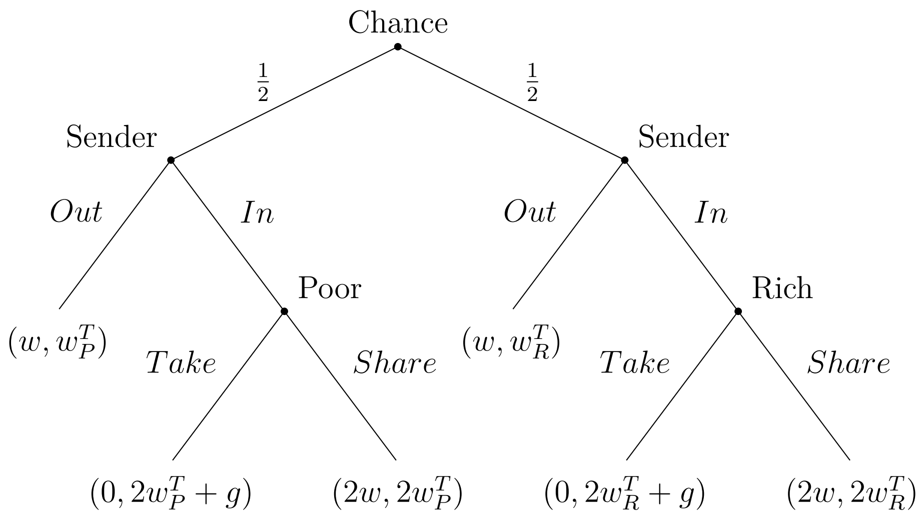

Figure 1: Game Form of the Stage Game with $T \in \{T0,\ T1,\ T2.\}$

As illustrated in Figure 1, the Sender chooses between *Out* and *In*, knowing the endowment of the Receiver. If the Sender chooses *Out*, the game ends and both players keep their initial endowment. If the Sender chooses *In*, the Sender entrusts the Receiver and the game continues. The Receiver then chooses between *Share* and *Take*. Under *Share*, the Receiver reciprocates the Sender's trust, and both players receive a payoff equal to twice their initial endowment. Under *Take*, the Receiver does not reciprocate the Sender's trust: the Receiver's endowment is doubled and increased by a bonus of $g = 2$ ECUs, while the Sender gets nothing in return.

2.2 Treatments

We employ a between-subject design with three conditions to identify the distinct behavioral effects of two ways of mitigating inequality: reducing the current endowment gap and improving ex ante future prospects through stochastic rank reversal across periods. These two mechanisms may affect receivers' willingness to reciprocate and, consequently, "Middle Class" senders' expectations of reciprocation and willingness to trust. The three conditions are as follows:

High Inequality (T0): In this condition, endowment inequality is high in both periods. The "Poor" receiver is endowed with 2 ECUs, the "Rich" receiver with 10 ECUs, and the "Middle Class" Sender with 6 ECUs. Thus, the distance between the Sender's endowment and that of either receiver type is symmetric and equal to ± 4 ECUs.

Low Inequality (T1): In this treatment, endowment inequality is lower in both periods than in the Control condition. The "Poor" receiver is endowed with 4 ECUs, the "Rich" receiver with 8 ECUs, and the "Middle Class" Sender with 6 ECUs. Thus, the distance between the Sender's endowment and that of either receiver type remains symmetric, and it is equal to ± 2 ECUs.

Rank Reversal (T2): In this treatment, within-period endowment inequality is the same as in the Control condition: in each period, the "Poor" receiver is endowed with 2 ECUs, the "Rich" receiver with 10 ECUs, and the "Middle Class" Sender with 6 ECUs. Receivers' relative ranks may, however, reverse across periods: a "Poor" receiver has a 50% probability of becoming "Rich" in Period 2, and vice versa. This reshuffling implies the same expected endowment dispersion across the two periods as in T1.

Table 1 summarizes the relevant dimensions manipulated in each treatment conditions.

Table 1: Summary of treatment conditions

	High Inequality (T0)	Low Inequality (T1)	Rank Reversal (T2)
Sender endowment	6	6	6
Poor receiver endowment	2	4	2
Rich receiver endowment	10	8	10
Expected Poor/Rich receiver endowment (across periods)	2/10	4/8	4/8
Sender-receiver gap (within period)	±4	±2	±4
Expected sender–receiver gap (across periods)	±4	±2	±2

2.3 Treatment Effects

Our pre-registered theoretical predictions[1] concern exclusively "Poor" receiver matches in period 1, before any rank reversal is realized. While both treatments T1 and T2 reduce inequality relative to the control T0, they do so differently: T1 compresses current endowment gaps, whereas T2 leaves within-period inequality unchanged but improves future prospects through stochastic rank reversal. This contrast yields sharp hypotheses about how these two distinct mechanism to reduce inequality affect trustworthiness and trust. We derive these predictions using the inequity-aversion framework of Fehr and Schmidt (1999) and a newly introduced parsimonious extension that takes into account prospects of future rank-reversal in one-shot anonymous interaction.

2.3.1 Current inequality

We begin with the inequity-aversion model of Fehr and Schmidt (1999), which provides a tractable framework for deriving predictions about how trust and trustworthiness respond to current payoff inequality.

We assume that receivers exhibit inequity aversion with an utility function defined by:

$$u_r(z;\theta_r^I,\theta_r^S) = \pi_r(z) - \theta_r^I\Big(\pi_s(z) - \pi_r(z)\Big)^+ - \theta_r^S\Big(\pi_r(z) - \pi_s(z)\Big)^+, \tag{1}$$

where π_i is the material payoff of player $i \in \{s, r\}$ and θ_i^I and θ_i^S are parameters that describe the aversion to inequality of player i. In the specific, $\theta_r^I \in [0, 1)$ and $\theta_r^S \in [0, 1)$ measure the receiver's aversion to disadvantageous and advantageous inequality, respectively. For simplicity, we assume that $\theta_r^I \geq \theta_r^S$, in line with the intuition of Fehr and Schmidt (1999) that being behind breeds stronger resentment than being ahead generates unease.

For a "Poor" receiver, the choice between *Share* and *Take* depends on the comparison between the material payoffs at the two terminal nodes, the envy generated by remaining behind the Sender under (*In*, *Share*) and the disutility associated with ending ahead under (*In*, *Take*):

[1] https://www.socialscienceregistry.org/trials/17352

$$u_P(In, Share) = 2w_P - \theta_P^I \left(2w - 2w_P\right)^+ , \quad \text{and} \quad u_P(In, Take) = 2w_P + g - \theta_P^S \left(2w_P + g\right)^+ .$$

Comparing the "Poor" receiver's utility at the two terminal nodes yields the following threshold for advantageous-inequality aversion:

$$\hat{\theta}_P^S = \frac{g + 2\theta_P^I(w - w_P)}{g + 2w_P}. \tag{2}$$

This threshold can be interpreted as the minimum level of advantageous-inequality aversion required for a "Poor" receiver to choose *Share*. Hence, the receiver reciprocates if and only if $\theta_P^S \geq \hat{\theta}_P^S$. Since $\hat{\theta}_P^S$ increases with the endowment gap $(w - w_P)$, higher inequality makes reciprocation less likely. Conversely, a lower level of inequality reduces the threshold and therefore increases the set of "Poor" receivers who choose *Share*.

For the Sender paired with a Poor Receiver, the problem is mostly strategical, as we assume the Sender to be selfish, risk-neutral, rational and to believe in the Receiver's rationality and inequity aversion.[2] Then, a Sender will trust the Poor Receiver, by playing *In*, if she thinks the Receiver will play *Share* with a probability greater than one half, $\alpha_S^{Sh} \geq \frac{1}{2}$. Following the theoretical analysis of Stringhi (2026), the value of α_S^{Sh} is inversely propositional to the threshold $\hat{\theta}_P^S$, which, in turn, is increasing in the inequality between the two players. As the Sender anticipates the Receiver behavior, this translate in lower first-order beliefs, and consequently, lower frequency of *In* being played as the inequality grows. Then, second-beliefs of the Receiver, β_P^{Sh} being the expectation over α_P^{Sh}, follow the same decreasing pattern.

The preferences for equality of the Receiver, together with the strategic reasoning of the Sender allow us to formulate the first hypothesis on the treatment effect:

Hypothesis 1. *Under the hypothesis of preferences for equality a la Fehr and Schmidt (1999), in matches with "Poor" receivers, the frequency of Share, In, and first- and second-order beliefs about the frequency of Share are higher in T1 compared to T0.*

This first result is due to the fact that in the *Low Inequality* treatment T1, the current inequality is lower compared to the other two treatments. Given that interactions are one-shot and anonymous, the two periods of play are strategically independent. Hence, the prospects for future improvement, present in T2, cannot be taken into account in the Fehr and Schmidt (1999) preferences.

2.3.2 Relevance of future prospects

In the Fehr–Schmidt benchmark, inequity aversion is evaluated over the realized material payoffs generated by the terminal history of the current interaction. This captures the case in which agents assess their relative payoff position using only the material consequences of their current actions. The Rank Reversal treatment introduces a distinct dimension of inequality: a "Poor" receiver who is currently disadvantaged may occupy the high-endowment position in the future, thereby reducing the expected (disadvantageous) inequality. Motivated by the idea that future prospects can affect how current disadvantage is perceived, we introduce an extension of Fehr–Schmidt framework, in which prospects of future endowment changes enter the reference

[2] This assumption — referred to as role-dependent inequity aversion — serves to clarify the results and the behavioral mechanisms at play. Assuming a selfish trustor does not meaningfully affect the analysis: her ranking over outcomes — $(In, Share) \succ (Out) \succ (In, Take)$ — remains unchanged for any level of inequity aversion. Her decision remains driven by strategic reasoning about the trustee's behavior, requiring a higher belief threshold (a probability premium) to justify trusting, since playing *In* may lead to greater inequality than *Out*.

point used to evaluate current payoff inequality. Under this new formulation, current relative disadvantage is perceived as less persistent, so to generate a smaller disutility.

Our new formulation captures situations in which institutions expand future opportunities without immediately eliminating current disadvantage. For example, disadvantaged students who gain access to better schools through admission lotteries may remain economically disadvantaged at the time of access, but their current position may be perceived as less persistent because the lottery changes their expected future trajectory. Similarly, families receiving access to housing-mobility programs may still face current resource gaps, yet the prospect of moving to a higher-opportunity environment can alter how enduring their relative disadvantage appears. In both cases, the opportunity does not directly change the current gap; rather, it changes the reference point against which current disadvantage is evaluated. More generally, individuals may react less negatively to current inequality when present disadvantage is perceived as less persistent and therefore less pronounced.

Formally, we capture this idea by modifying the inequality component of Fehr–Schmidt utility. For the current terminal history, the material consequences remains the payoffs generated by the current action profile. The inequality term, however, is computed taking into account the average inequality generated in all the possible future contingencies at the same terminal history. In our specific case, the future contingencies are represented by the chance of rank reversal in the second period in the *Rank Reversal* treatment.[3]

To keep the analysis aligned with the core mechanism of our experiment, we introduce a simple formulation: expected future endowment changes enter the reference point used to evaluate the inequality generated by current actions. Rank-reversal prospects affect behavior not as future payoffs per se but by changing how current payoff differences are perceived. The following specification implements this idea by modifying only the inequality terms in Fehr–Schmidt utility. Instead of using only the current payoff gap, the model uses a time discounted, probability-weighted average of current and possible future gaps. We formalize this idea through a forward-looking extension of Fehr–Schmidt preferences:

$$
\begin{aligned}
u_r^0(z) = \pi_r^0(z) - \theta_r^I \Big(\frac{1}{N} \sum_{t=0}^{T} \delta^t \sum_{p_n^t \in \Omega} p_n^t \big(\pi_{s,n}^t(z) - \pi_{r,n}^t(z) \big) \Big)^+ \\
- \theta_r^S \Big(\frac{1}{N} \sum_{t=0}^{T} \delta^t \sum_{p_n^t \in \Omega} p_n^t \big(\pi_{r,n}^t(z) - \pi_{s,n}^t(z) \big) \Big)^+, \quad \frac{1}{N} = \frac{1}{\sum_{t=0}^{T} \delta^t}
\end{aligned}
\tag{3}
$$

The material payoff component $\pi_r(z)$ remains the payoff generated by the current terminal history z. The departure from the Fehr–Schmidt benchmark lies only in the inequality terms. For each period t, the inner summation computes the expected inequality across the possible future states $p_n^t \in \Omega$, evaluated at the same terminal history z. These expected gaps are then aggregated over the current and future periods using the inter-temporal discount factor $\delta \in [0, 1]$. The normalization term $N = \sum_{t=0}^{T} \delta^t$ ensures that inequity is evaluated as a weighted average payoff gap over the relevant horizon. Hence, δ measures the extent to which future rank prospects enter the receiver's current assessment of inequality. When $\delta = 0$, the model

[3] This formulation is inspired by Saito (2013)'s model of inequality-averse preferences under risk, but it departs from a direct application of Saito's formal representation. Saito evaluates lotteries over payoff allocations by combining inequality in expected allocations and expected inequality in realized allocations. Moreover, Saito (2013) is not an inter-temporal model. In our setting, an inter-temporal extension of Saito would require a dynamic specification of the full plans of action (including beliefs about future play) of both players in all the future contingencies. This makes the problem hard to tract analytically. Also, Saito's model doesn't fully apply to our setting, since we have random matching between periods, therefore the future inequality cannot be evaluated with respect to the initial counterpart.

collapses to the Fehr–Schmidt benchmark.

We now derive the treatment effect predicted in the *Rank Reversal* condition (T2) using our new utility function. Consider a Poor receiver in period 1, before rank reversal is realized. Conditional on the Sender choosing *In*, the receiver compares the utilities at the two terminal histories (*In*, *Share*) and (*In*, *Take*). Substituting the corresponding stage-game payoffs into equation (3) yields the utility for *Share* and *Take* when future endowment positions enter the inequality term.

$$\begin{aligned} u_P(\mathit{In}, \mathit{Share}) &= 2w_P - \theta_P^I \Big(\frac{1}{1+\delta} \cdot \Big[(2w - 2w_P) + \delta(1-p)(2w - 2w_P) + \delta p (2w - 2w_R) \Big] \Big)^+ \\ &= 2w_P - \theta_P^I \Big(\frac{1}{1+\delta} \cdot \Big[2w(1+\delta) - 2(1+\delta - \delta p) w_P - 2\delta p w_R \Big] \Big)^+, \\ u_P(\mathit{In}, \mathit{Take}) &= 2w_P + g - \theta_P^S \Big(\frac{1}{1+\delta} \cdot \Big[(2w_P + g) + \delta(1-p)(2w_P + g) + \delta p (2w_R + g) \Big] \Big)^+ \\ &= 2w_P + g - \theta_P^S \Big(\frac{1}{1+\delta} \cdot \Big[2(1+\delta - \delta p) w_P + 2\delta p w_R + (1+\delta) g \Big] \Big)^+. \end{aligned}$$

Comparing the receiver's utility at the two terminal histories yields the threshold level of advantageous-inequality aversion required for for a "Poor" receiver to play *Share*:

$$\hat{\theta}_P^S(\delta, p) = \frac{(1+\delta)g + 2\theta_P^I \left[(1+\delta)w - (1+\delta-\delta p)w_P - \delta p w_R \right]}{(1+\delta)g + 2(1+\delta-\delta p)w_P + 2\delta p w_R}. \tag{4}$$

Thus, a Poor receiver chooses *Share* rather than *Take* if and only if $\theta_P^S \geq \hat{\theta}_P^S(\delta, p)$. The threshold is monotonically decreasing in the weight attached to future endowment prospect δ:

$$\frac{\partial \hat{\theta}_P^S(\delta, p)}{\partial \delta} = - \frac{2p(w_R - w_P) \left[g + \theta_P^I (g + 2w) \right]}{\left[(1+\delta)g + 2(1+\delta-\delta p)w_P + 2\delta p w_R \right]^2} \leq 0.$$

Consider the special case in which future endowment position is incorporated without discounting, $\delta = 1$. Since in the *Rank Reversal* treatment a Poor receiver becomes Rich in the second period with probability $p = 1/2$, the relevant endowment benchmark is

$$\bar{w}_P = \frac{w_P + \left(\frac{1}{2} w_P + \frac{1}{2} w_R \right)}{2}. \tag{5}$$

Substituting this benchmark into the utility expressions gives

$$u_P(\mathit{In}, \mathit{Share}) = 2w_P - \theta_P^I \left(2w - 2\bar{w}_P \right)^+, \quad \text{and} \quad u_P(\mathit{In}, \mathit{Take}) = 2w_P + g - \theta_P^S \left(2\bar{w}_P + g \right)^+.$$

Under our calibration, $\bar{w}_P < w$. Thus, the relevant comparison under *Share* remains one of disadvantage. The corresponding threshold is

$$\hat{\theta}_P^S = \frac{g + 2\theta_P^I (w - \bar{w}_P)}{g + 2\bar{w}_P}.$$

By design, $w_P^{T1} = \bar{w}_P$. Hence, when future endowment position is incorporated without discounting, the Poor receiver's sharing threshold in T2 coincides with the threshold in T1. Hence, the behavior of senders and receiver should be the same in the two conditions. Instead for $\delta = 0$, meaning that the Poor receive doesn't take into consideration the future prospects, we obtain that in T2 a Poor receiver should behave as in T0, as the utility function collapse to the usual Fehr–Schmidt utility. These considerations allows us to formulate the second behavioral hypothesis relative to the behavior in T2.

Hypothesis 2. *Under preferences for equality that incorporate future prospects, in matches with "Poor" receivers, the frequency of Share, In, and first- and second-order beliefs about the frequency of Share in T2 lies between T0 and T1. As $\delta \to 0$, behavior and beliefs in T2 converge to T0; as $\delta \to 1$, behavior and beliefs in T2 converge to T1.*

2.3.3 Behavior of Rich receivers

Rich receivers play an ancillary role as their presence makes rank reversal possible in T2: a currently Poor receiver may become Rich in the second period, and a currently Rich receiver may become Poor. Consistently, in our design, their behavioral incentives are calibrated to remain invariant across treatments. Hence, under both the Fehr–Schmidt benchmark and our forward-looking extension, the Rich receiver's decision is characterized by the same threshold for advantageous-inequality aversion in all treatments:[4]

$$\hat{\theta}_R^S = \frac{g}{g + 2w}. \tag{6}$$

The threshold is independent of the Rich receiver's endowment. Therefore, the model predicts no systematic treatment effect on Rich receivers' willingness to choose *Share*. This feature of the calibration provides a useful benchmark to test the relevance of inequity-aversion preferences in explaining the behavioral variation across treatment observed in our experiment. If the behavior of Rich receivers is constant across treatment conditions, this entails that other preferences that are not treatment invariant (e.g., loss aversion, aversion to intertemporal fluctuations, or risk aversion) are not prominent in describing the behavior of the subjects in our experiment.

2.4 Design choices

The choice of the Trust Mini Game and the calibration of the Sender's payoffs are deliberate. They are designed to make the Sender's decision primarily a strategic one, rather than a direct expression of the Sender's own inequity aversion. In utility terms, the terminal history (*In*, *Take*) is always less attractive to the Sender than (*Out*), while (*In*, *Share*) yields higher utility than (*Out*) whenever the Receiver is expected to reciprocate with sufficiently high probability. Hence, the Sender chooses *In* only if the expected utility from trusting exceeds the utility of opting out. In this sense, choosing *In* can be interpreted as a behavioral measure of trust, namely a decision based on the expectation that the Receiver will behave in a trustworthy manner.

To preserve comparability across endowment levels and treatments, the payoff gain from the Receiver's non-cooperative action *Take* is fixed at $g = 2$. This prevents treatment differences in receivers' behavior from being mechanically driven by variation in the material return to non-cooperation. As a result, behavioral differences across treatments can be more cleanly interpreted as responses to the underlying inequality environment rather than to changing earnings opportunities.

An important design choice in experiments on mobility and rank reversal concerns the mechanism through which future positions are determined. Several studies implement rank reversal through rules that involve merit, effort, or deservingness considerations (e.g., Gangadharan et al., 2021; Lange et al., 2021). In our experiment, by contrast, rank reversal is purely stochastic. Although this choice abstracts from important real-world features of mobility, it is central to our identification strategy. Our stochastic implementation of rank reversal is intended to manipulate the perceived persistence of current disadvantage while avoiding confounds related to effort, merit, or deservingness. This choice is motivated by evidence that the behavioral relevance of ranks depends on their persistence. Martinangeli and Windsteiger (2021) show

[4] The derivation of this threshold under both utility specifications is reported in Appendix A.1.

that rank-based concerns are weaker when ranks are randomly reassigned and stronger when ranks are fixed, suggesting that non-persistent rank positions may carry less behavioral weight. In our design, *Rank Reversal* therefore provides a clean test of whether reducing the expected persistence of disadvantage, without reducing current inequality, affects trust and trustworthiness.

This experiment compares the behavioral effect of reducing current inequality with the effect of improving expected future condition, while holding expected intertemporal endowments fixed. This requires the endowments in the *Low Inequality* treatment to coincide with the expected endowments in the *Rank Reversal* treatment, as in Formula 5. In our design, this equality is obtained by assigning an exogenous probability of rank reversal equal to $\frac{1}{2}$. If rank reversal depended on endogenous factors such as effort, ability, or performance, subjects' perceived probabilities of mobility could differ across individuals and need not equal $\frac{1}{2}$. Moreover, such mechanisms would introduce additional considerations of deservingness, responsibility, and perceived control. This would compromise the calibration and make it harder to interpret differences between T1 and T2 as responses to current inequality versus future rank prospects.

Random matching and independent payment across periods are also important features of the design. They allow us to capture the idea that future rank prospects may affect the perception of current inequality even when the current interaction is one-shot. Thus, mobility prospects need not operate through an ongoing relationship with the same counterpart; they may instead affect the reference point used to evaluate the current payoff gap. At the same time, it ensures that the two periods of the game are strategically independent. This independence is important for the theoretical interpretation of the treatment. If we paid the outcome with the same players in both periods, evaluating inequality would require assessing the total inequality generated by their actions over the entire two-period interaction. This, in turn, would require a full specification of dynamic strategies, beliefs, and plan of play. Random rematching and independence of payments between rounds avoid this complication. It allows us to keep the material consequences of the current action tied to the current terminal history, while allowing future prospects to enter only through the reference point used to evaluate current inequality, as captured by our intertemporal utility formulation.

Taken together, these design choices allow us to cleanly identify how alternative inequality-reducing mechanisms — reducing current inequality *vs* improving future prospects — affect trust and trustworthiness in interactions with disadvantaged counterparts.

2.5 Procedure

As preregistered, we recruited 240 participants from the subject pool of the Prague University of Economics and Business via ORSEE (Greiner, 2004). The pool consists primarily of students in business and economics-related disciplines. The experiment was programmed in oTree (Chen et al., 2016) and conducted at the Laboratoř Experimentální Ekonomie (LEE). The study was approved by the Ethics Committee of the University of Siena, and all data were collected and processed in accordance with the General Data Protection Regulation (GDPR 2016/679).

The experiment was conducted in 12 sessions in December 2025. Participants signed up for available session time slots before treatment assignment, and sessions were subsequently randomly assigned to treatments. To achieve a balanced design, the 12 sessions were evenly allocated across treatments, with four sessions per treatment.

Average earnings were 318 CZK, including a 150 CZK show-up fee, and sessions lasted approximately 60 minutes on average. Payments were made in cash at an exchange rate of 1 ECU = 10 CZK.

After participants were seated at their computer terminals, research assistants read the instructions aloud and provided clarifications when needed. The instructions were in English, reflecting the international composition of the student body at the Prague University of Economics and Business. Each participant also received a printed copy of the instructions.

At the beginning of the experimental session, participants were randomly assigned to one of two roles, Sender or Receiver, and remained in that role for the entire session. Receivers were further randomly assigned to one of two endowment conditions, "Poor" or "Rich".

In each session, participants took part in a Trust Mini Game played over two independent periods, with each period implemented as a block of 10 rounds, for a total of 20 rounds. The first block corresponded to period 1, and the second block corresponded to period 2. The repetition was done to improve the consistency of the results (Hey, 2001). In every round, Senders were randomly and anonymously matched with a Receiver, whose status was either "Poor" or "Rich". Interactions were anonymous, and both players were informed of their matched counterpart's endowment level.

In treatment T2, rank-reversal took place after the 10 rounds of the first block. Participants were informed if their endowment changed or not before proceeding with the second block.

In each round, participants completed two tasks displayed on a single screen: a decision task and a belief-elicitation task. This procedure was the same throughout the experiment.

In the decision task, Senders chose between *In* and *Out*, while Receivers chose between *Share* and *Take*. Receivers were informed of their realized status before choosing between *Share* and *Take*. Senders were informed of the current status of their matched receiver before choosing between *In* and *Out*.

In the belief-elicitation task, Senders reported their first-order beliefs about the behavior of the receiver type with whom they were matched. Specifically, a Sender matched with a "Poor" ("Rich") Receiver indicated how many out of five "Poor" ("Rich") Receivers she expected to choose *Share*. Receivers reported their second-order beliefs by indicating what they believed their matched Sender expected for Receivers of their own type. In both cases, beliefs were elicited on a six-point scale ranging from 0/5 to 5/5.

For payment, one round was randomly selected from each of the two blocks. Participants also earned 1 ECU for each correct belief reported in each rounds.

3 Results

3.1 Descriptive statistics

We begin by presenting descriptive evidence on trust, trustworthiness, and beliefs across treatments. Because our main predictions concern period-1 matches involving "Poor" receivers, we focus on these first and then briefly consider "Rich"-receiver matches as a benchmark.[5]

Poor-receiver matches. Figure 2 shows a clear descriptive pattern: trust, trustworthiness, and senders' first-order beliefs are all higher under *Low Inequality* (T1) than under *High Inequality* (T0), while *Rank Reversal* (T2) remains much closer to T0. The magnitudes are economically meaningful. The probability that senders choose *In* rises from about 0.42 in T0 to about 0.56 in T1, while remaining around 0.38 in T2. Senders' first-order beliefs follow the same pattern: on average, senders expect about 39% of matched "Poor" receivers to choose *Share* in T0, 38% in T2, and 49% in T1. Consistently, the probability that "Poor" receivers choose *Share* rises from about 0.34 in T0 to about 0.50 in T1, while remaining around 0.35 in T2. "Poor" receivers' second-order beliefs are less tightly aligned with behavior: the pattern is broadly similar, but noisier overall, and differences across treatments are much less pronounced.

[5]For a descriptive summary of the main outcomes by match type and treatment, see Table A1 in the Appendix.

Overall, the descriptive evidence suggests that the main differences in poor-receiver matches are driven by T1, whereas T2 displays little visible change relative to the control.[6]

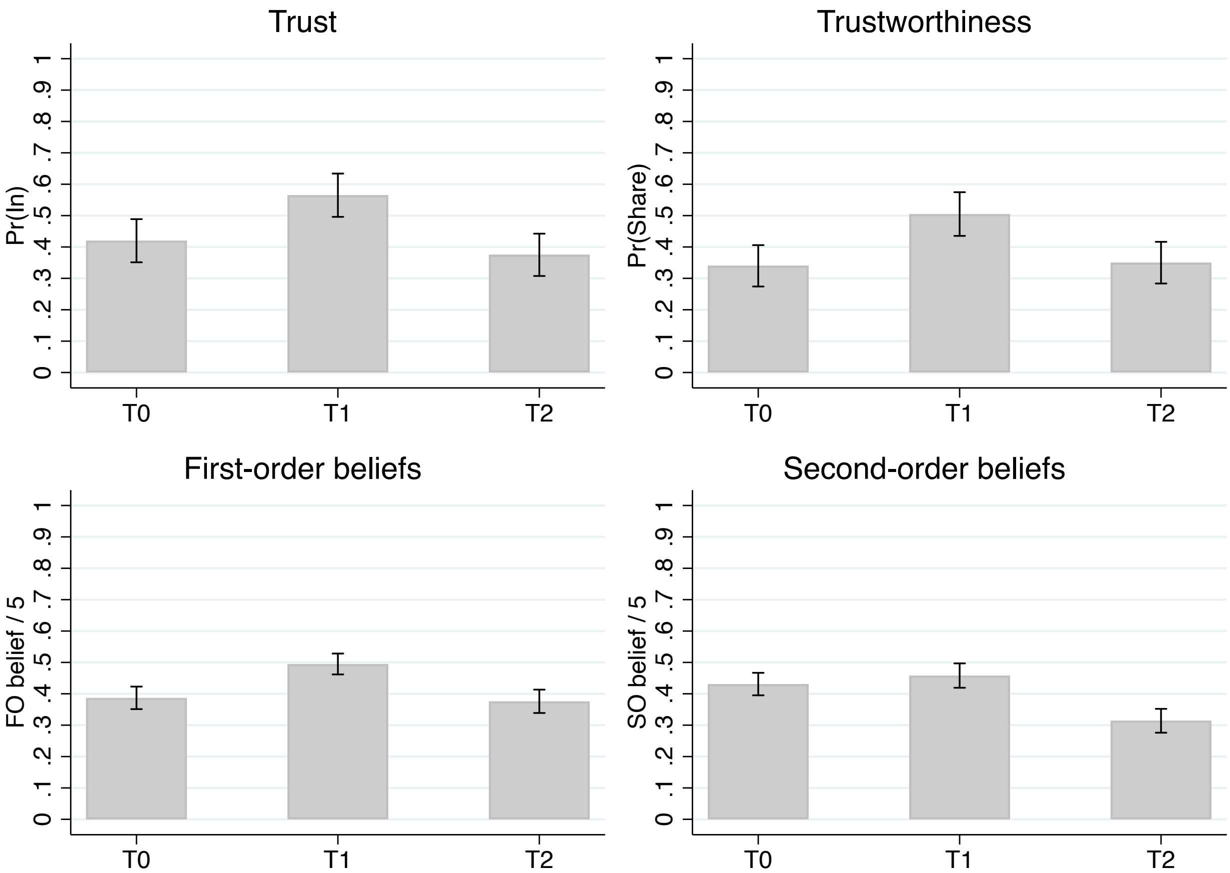

Figure 2: **Poor-receiver matches**: Trust, trustworthiness, and first- and second-order beliefs across treatments. Bars report average outcomes across the 10 rounds in the first block; whiskers indicate 95% confidence intervals. Treatments are ordered T0 = High Inequality, T1 = Low Inequality, and T2 = Rank Reversal.

Rich-receiver matches. We next consider period-1 matches involving "Rich" receivers, which serve as a benchmark. Relative to poor-receiver matches, trust, trustworthiness, and both first- and second-order beliefs are higher overall. Figure 3 shows, however, that these outcomes vary little across treatments, with no clear descriptive pattern of treatment effects. This suggests that the main treatment differences are concentrated in poor-receiver matches, while rich-receiver matches remain broadly stable across conditions.

[6]See Section A.2 of the Appendix for additional evidence at both the session level (Figures A1 and A2) and the round level (Figures A3 and A4), confirming that the reported patterns are robust across levels of aggregation.

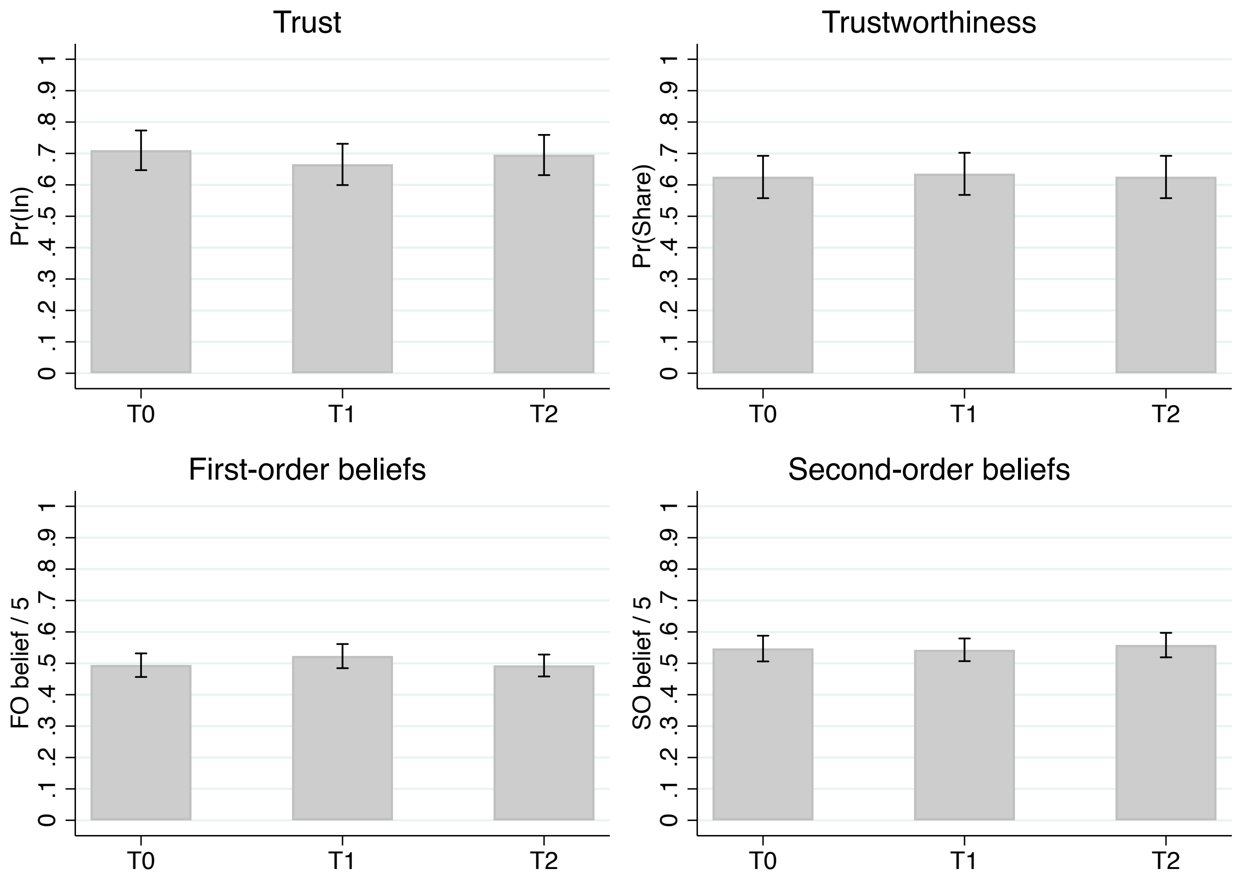

Figure 3: Rich-receiver matches: Trust, trustworthiness, and first- and second-order beliefs across treatments. Bars report average outcomes across the 10 rounds in the first block; whiskers indicate 95% confidence intervals. Treatments are ordered T0 = High Inequality, T1 = Low Inequality, and T2 = Rank Reversal.

We next turn to the regression analysis to examine these patterns more formally, especially in poor-receiver matches.

3.2 Econometric analysis

This section reports the results from the pre-registered econometric analysis of behavior and beliefs in rounds 1–10. We estimate the model separately for each treatment comparison: Control vs. Low Inequality, Control vs. Rank Reversal, and Rank Reversal vs. Low Inequality. In each comparison, the treatment indicator $Treat_j$ equals one for the second treatment named in the comparison and zero for the first. Sender and receiver choices are estimated using random-effects Probit models, while belief outcomes using linear panel models. All specifications include round fixed effects, controls for age and gender, and standard errors clustered at the session level.

Senders. Let In_{jr} be an indicator equal to one if sender j chooses *In* in round r, and let $PoorPartner_{jr}$ indicate whether the matched receiver is "Poor". We estimate

$$\begin{aligned} In_{jr} = \alpha + \beta_1 PoorPartner_{jr} + \beta_2 Treat_j \\ + \beta_3 \Big(PoorPartner_{jr} \times Treat_j \Big) + \boldsymbol{\gamma}' \mathbf{X}_j + \lambda_r + \varepsilon_{jr}. \end{aligned} \tag{7}$$

The coefficient β_2 captures the treatment effect for matches with a rich receiver, while β_3 indicates whether this effect differs between poor- and rich-partner matches. Accordingly, $\beta_2 + \beta_3$ captures the treatment effect for matches with a poor receiver. The same specification is used for senders' first-order beliefs, estimated with a linear panel model. Here, we plot the average marginal effects for the probability of choosing *In* and for first-order beliefs. Full estimation results are reported in Table A2 and A3 in Appendix.

Figure 4 reports the estimated average marginal treatment effects for senders matched with poor receivers, for both the probability of choosing *In* and first-order beliefs.

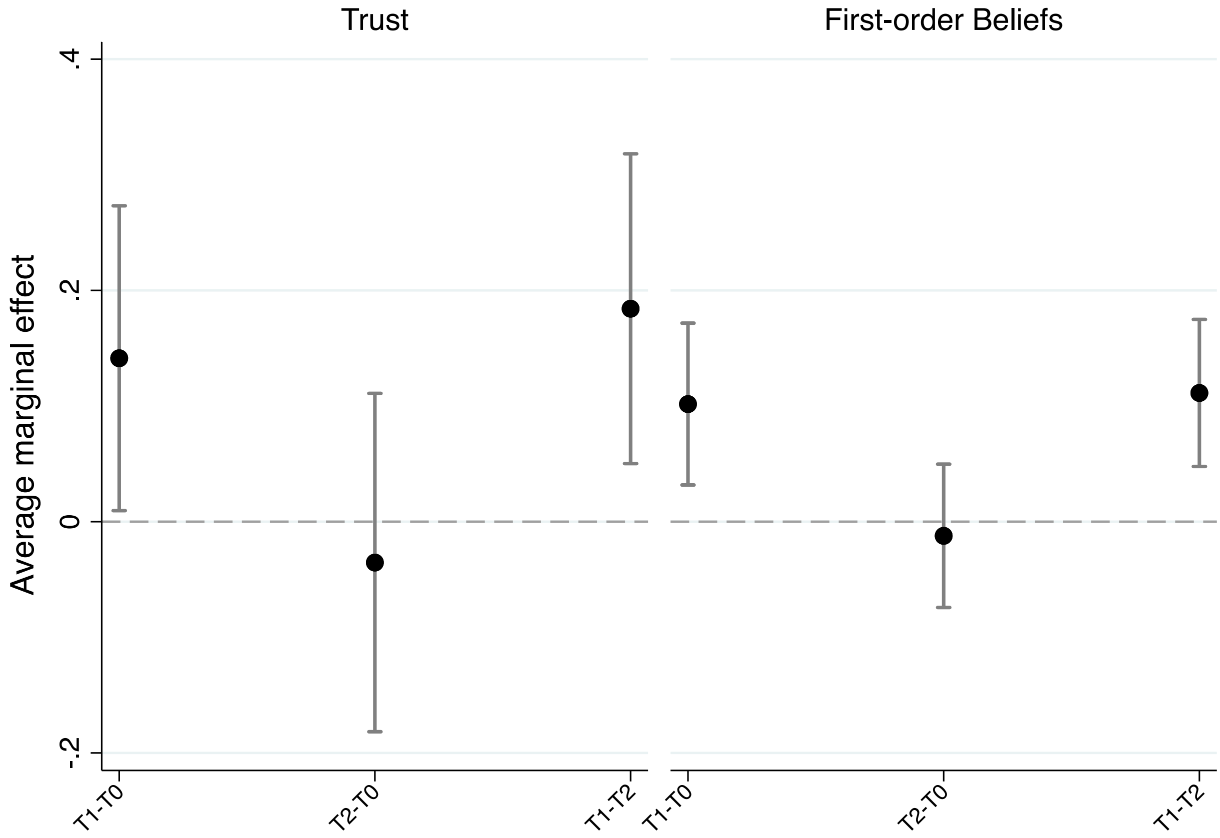

Figure 4: Marginal treatment effects for senders (poor-receiver matches). The left panel shows trust; the right panel shows first-order beliefs. Points denote estimates and whiskers 95% confidence intervals.

The figure shows that Low Inequality increases both trust and first-order beliefs in poor matches relative to both High Inequality Control and Rank Reversal, whereas Rank Reversal does not differ significantly from High Inequality Control. The effects are economically meaningful: under Low Inequality, the probability that senders choose *In* rises by about 14 p.p. relative to Control and by about 18 p.p. relative to Rank Reversal. First-order beliefs also increase by about 10-11 percentage points relative to High Inequality Control and Rank Reversal, respectively.[7]

Figure 5 reports the estimated average marginal treatment effects for senders matched with rich receivers.[8] Effects on both the probability of choosing *In* and first-order beliefs are small and statistically indistinguishable from zero across all pairwise comparisons.

[7]First-order beliefs are rescaled to the unit interval by dividing by five the reported number of expected

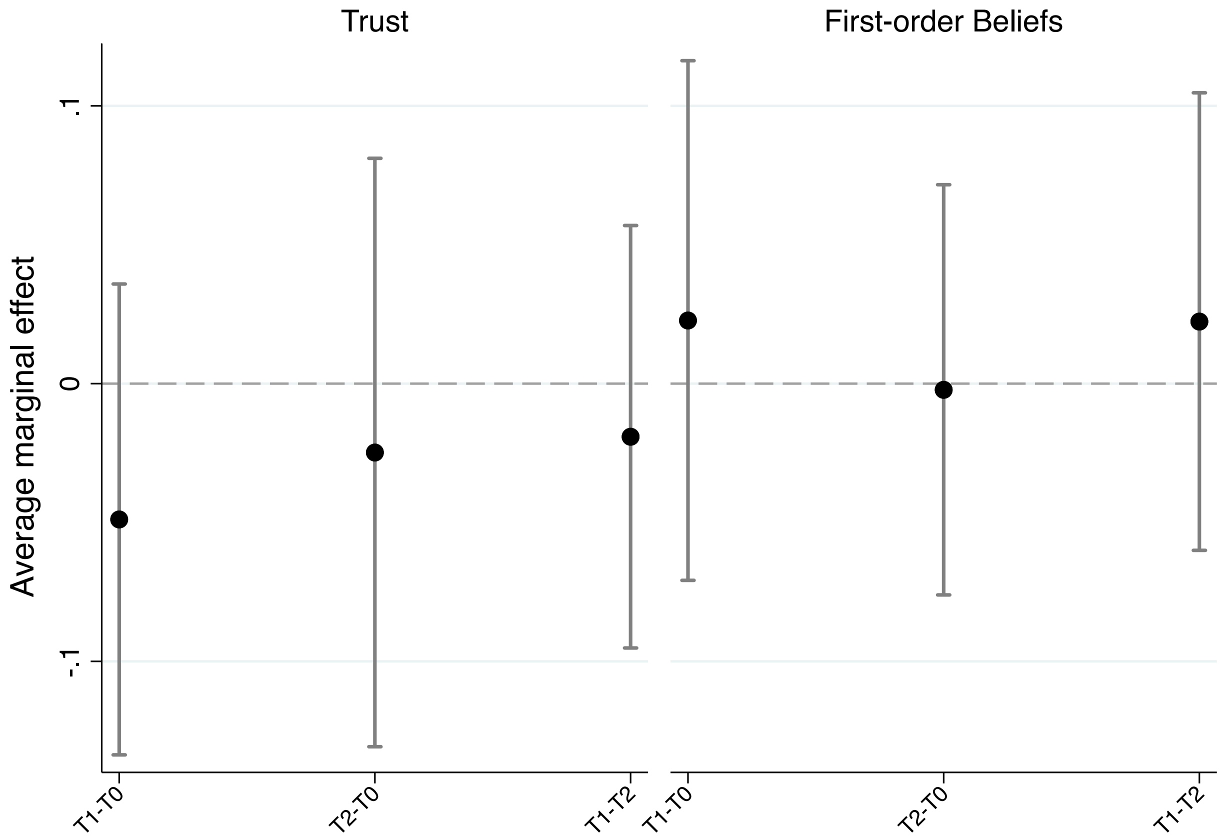

Figure 5: Marginal treatment effects for senders (rich-receiver matches). The left panel shows trust; the right panel shows first-order beliefs. Points denote estimates and whiskers 95% confidence intervals.

Receivers. Let $Share_{ir}$ be an indicator equal to one if receiver i chooses *Share* in round r. For each pairwise treatment comparison, we estimate the model separately for poor and rich receivers:[9]

$$Share_{ir} = \alpha_g + \delta_g Treat_i + \boldsymbol{\gamma}_g' \mathbf{X}_i + \eta_r + \varepsilon_{ir}, \tag{8}$$

where $g \in \{poor, rich\}$ indexes receiver type. The coefficient δ_g captures the treatment effect for receiver type g. The same specification is used for receivers' second-order beliefs, estimated with a linear panel model. In Figure 6, we plot the average marginal effects on the probability of choosing *Share* and for second-order beliefs (full estimation results are in Table A4 and A5 in Appendix).

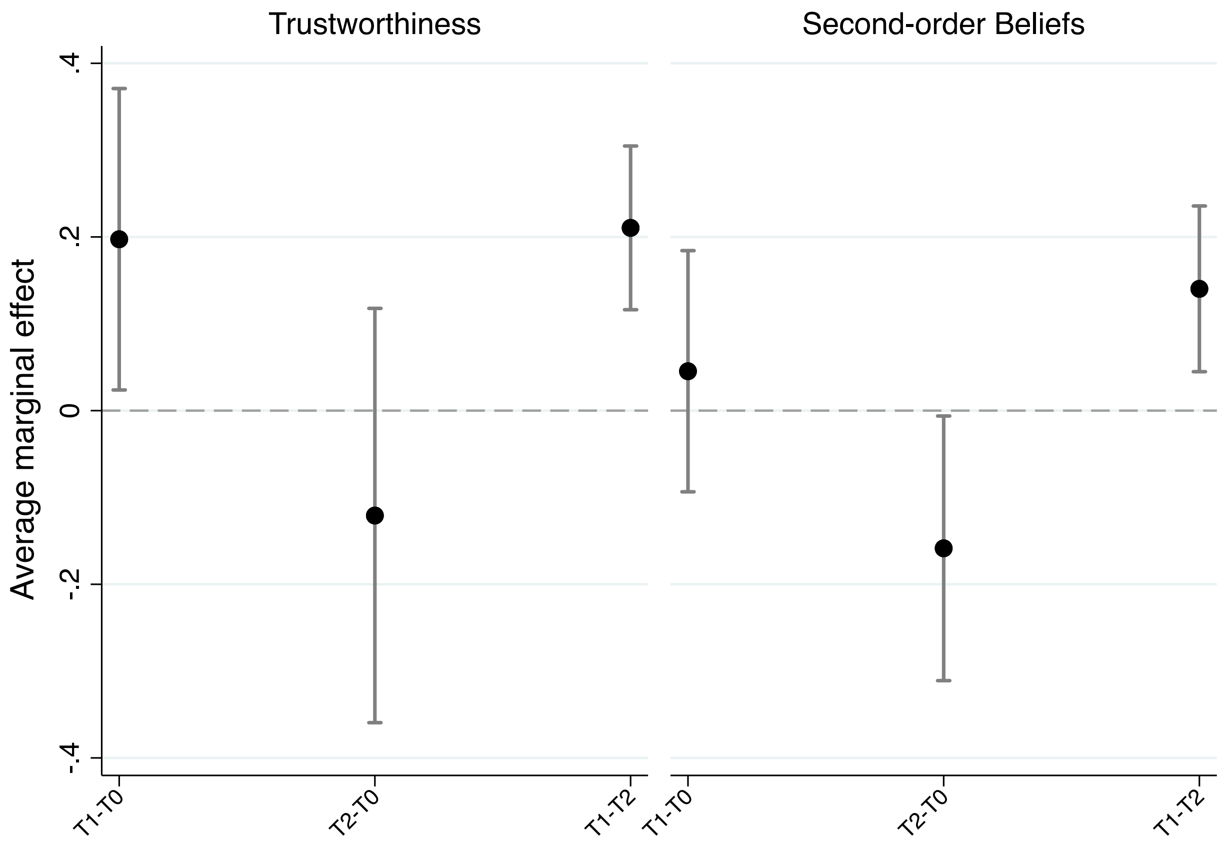

Figure 6: Average marginal treatment effects for poor receivers. The left panel shows sharing and the right panel second-order beliefs. Points denote estimates and whiskers 95% confidence intervals.

The figure shows that Low Inequality increases the probability that poor receivers choose *Share* by about 20 p.p. relative to High Inequality Control and by about 21 p.p. relative to Rank Reversal.. The Rank Reversal treatment, instead, does not differ significantly from Control. Second-order beliefs are noisier and less closely aligned with behavior. Low Inequality raises

cooperators in each specific group ("Poor" or "Rich").

[8]For estimation results see Table A2 and A3 in Appendix

[9]We estimate the receiver model separately by receiver type because "Poor" and "Rich" receivers interact only with "Middle Class" senders. Treatment effects are therefore naturally defined within receiver type, and the split specification keeps each regression tied to a homogeneous decision environment. Consistent estimation results from pooled specifications with treatment-by-receiver-type interactions are in Table A6 in Section A.4 of the Appendix.

beliefs, but this increase is precisely estimated only relative to Rank Reversal. Rank Reversal, in turn, lowers receivers' beliefs relative to Control.

Figure 7 reports the estimated average marginal treatment effects for rich receivers.

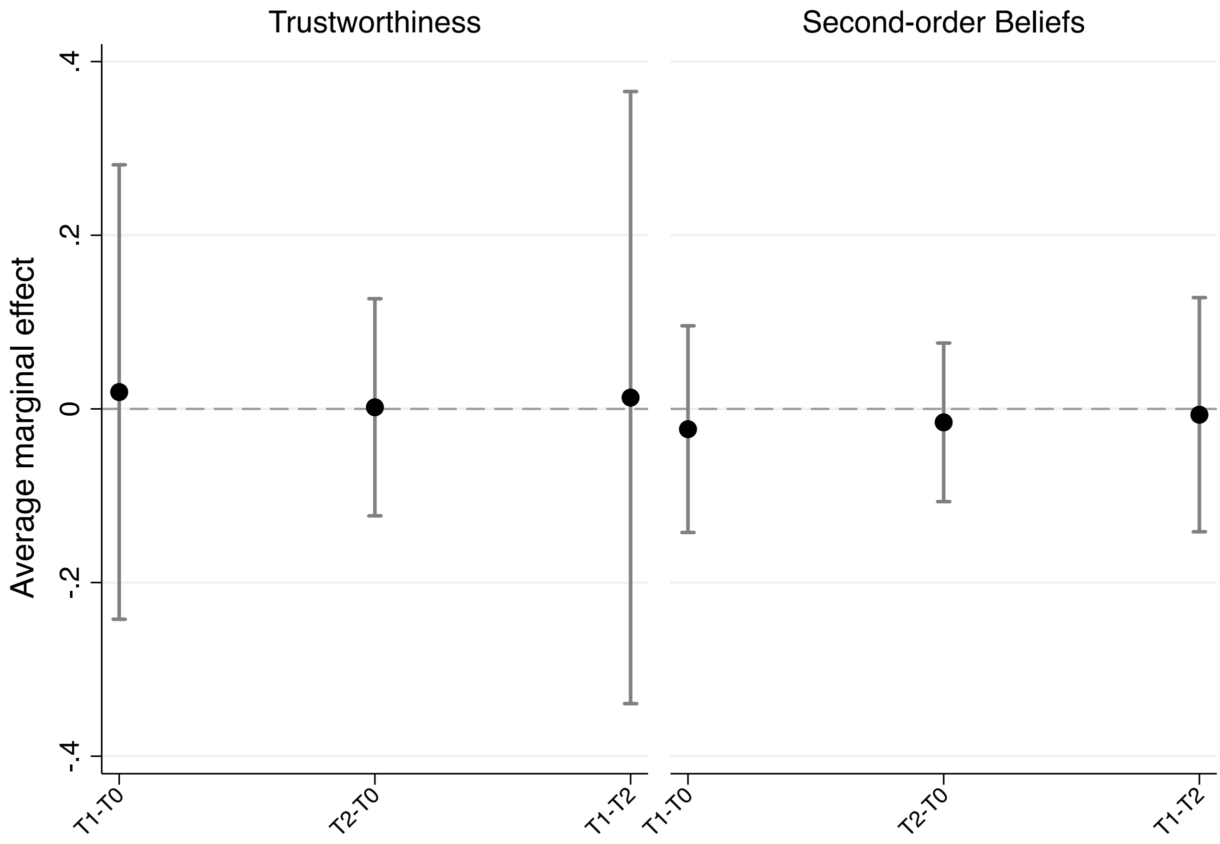

Figure 7: Average marginal treatment effects for rich receivers. The left panel shows sharing and the right panel second-order beliefs. Points denote estimates and whiskers 95% confidence intervals.

The figure shows no meaningful variation across treatments. The estimated effects on both sharing and second-order beliefs are small and statistically indistinguishable from zero across all pairwise comparisons.

Overall, results show that *Low Inequality*, but not *Rank Reversal*, strengthens trust and trustworthiness toward disadvantaged counterparts. This effect appears to operate through higher trustworthiness among poor receivers, which increases senders' expectations of reciprocation and hence their willingness to trust.[10]

4 Conclusion

This paper compares two stylized environments in a pre-registered experiment based on a Trust Mini Game played over two independent periods. One environment features a smaller

[10]Because participants observe counterpart choices after each round, behavior may partly reflect learning from previous interactions. We therefore estimate the treatment effect conditional on feedback from the previous round. Figures A5–A8 in the Appendix report average marginal treatment effects from specifications that control for the counterpart's choice in the previous round. he estimates reproduce the same pattern of treatment differences documented in the main analysis.

current endowment gap, implemented in the *Low Inequality* treatment. The other features better *ex ante* prospects under unchanged current inequality, implemented in the *Rank Reversal* treatment, where a "Poor" receiver may become "Rich" in period 2, and *vice versa*. The two treatments are calibrated to imply the same expected dispersion of endowments across the two periods of the game. We therefore focus on whether first-period behavior responds more strongly to a smaller current gap or to better prospects *ex ante*, before any rank reversal is realized.

Our benchmark predictions concern period-1 matches involving "Poor" receivers. To formulate the treatment predictions, we take the Fehr–Schmidt framework as a benchmark and then propose a forward-looking extension in which future rank prospects enter the reference point used to evaluate current payoff inequality, while current material payoffs remain unchanged. Under the standard Fehr–Schmidt logic, behavior is driven by current realized inequality: *Low Inequality* should therefore generate more trust and trustworthiness than both the *High Inequality* Control and *Rank Reversal*, because only *Low Inequality* reduces the sender–receiver gap in the current interaction. Under the forward-looking extension, by contrast, *Rank Reversal* may also affect behavior even though current inequality is unchanged, because the "Poor" receiver's period-1 disadvantage may be perceived as less persistent when future rank reversal is possible. The extension therefore nests the benchmark case in which future prospects are ignored and provides a simple prediction for the intermediate case: if future prospects matter, trust and trustworthiness should be weakly above the *High Inequality* Control, where current disadvantage is persistent, but weakly below *Low Inequality*, where the current sender–receiver gap is directly reduced.

The evidence is most consistent with the view that behavior in this setting responds primarily to current inequality. Relative to the *High Inequality* Control, *Low Inequality* increases trust toward "Poor" receivers, whereas *Rank Reversal* yields effects close to zero. Senders' first-order beliefs move in the same direction: *Low Inequality* raises expectations about poor receivers' cooperation, while *Rank Reversal* leaves these expectations essentially unchanged. Poor receivers' behavior mirrors this pattern. Under *Low Inequality*, they are more likely to reciprocate, whereas *Rank Reversal* does not produce a comparable increase in trustworthiness. Receivers' second-order beliefs are less precisely aligned with behavior and therefore play a more limited role in the interpretation. By contrast, matches with "Rich" receivers show little evidence of treatment effects on trust, trustworthiness, or beliefs. Taken together, these findings suggest that a smaller current resource gap makes poor receivers more willing to cooperate and, consistent with higher expected reciprocation, also makes senders more willing to trust.

More broadly, the findings speak to environments in which institutions expand future trajectories without removing current disadvantage at the time of interaction. In such settings, better *ex ante* prospects may be insufficient to generate the same gains in trust and social cooperation as a direct reduction in the present economic gap. An important agenda for future research is to examine this comparison in environments where mobility prospects arise through qualitatively different processes rather than through an exogenous stochastic reversal of positions. In particular, introducing real-effort tasks, performance-based mobility, or endogenous institutional choices over the distribution of resources would allow the analysis to speak to settings in which future positions are linked to responsibility, selection, or political choice, and therefore to different dimensions of mobility, inequality, and fairness.

References

Agranov, M. and T. R. Palfrey (2020). The effects of income mobility and tax persistence on income redistribution and inequality. *European Economic Review 123*, 103372.

Alesina, A. and E. La Ferrara (2002). Who trusts others? *Journal of public economics 85*(2), 207–234.

Algan, Y. and P. Cahuc (2013). Trust and growth. *Annu. Rev. Econ. 5*(1), 521–549.

Ananyev, M. and S. Guriev (2019). Effect of income on trust: Evidence from the 2009 economic crisis in russia. *The Economic Journal 129*(619), 1082–1118.

Anderson, L. R., J. M. Mellor, and J. Milyo (2006). Induced heterogeneity in trust experiments. *Experimental Economics 9*(3), 223–235.

Bergman, P., R. Chetty, S. DeLuca, N. Hendren, L. F. Katz, and C. Palmer (2024). Creating moves to opportunity: Experimental evidence on barriers to neighborhood choice. *American Economic Review 114*(5), 1281–1337.

Cappelen, A. W., J. Konow, E. Ø. Sørensen, and B. Tungodden (2013). Just luck: An experimental study of risk-taking and fairness. *American Economic Review 103*(4), 1398–1413.

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

Chetty, R., N. Hendren, and L. F. Katz (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review 106*(4), 855–902.

Cullen, J. B., B. A. Jacob, and S. D. Levitt (2006). The effect of school choice on participants: Evidence from randomized lotteries. *Econometrica 74*(5), 1191–1230.

D'Amato, M., N. O'Higgins, and M. Stimolo (2021). On inequality, growth and trust: some evidence from the lab. *Journal of Institutional Economics*, 1–16.

Deming, D. J., J. S. Hastings, T. J. Kane, and D. O. Staiger (2014). School choice, school quality, and postsecondary attainment. *American Economic Review 104*(3), 991–1013.

Duflo, E. (2001). Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment. *American Economic Review 91*(4), 795–813.

Duflo, E., P. Dupas, and M. Kremer (2021, June). The impact of free secondary education: Experimental evidence from ghana. NBER Working Paper 28937, National Bureau of Economic Research.

Fehr, E. and K. M. Schmidt (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics 114*(3), 817–868.

Gangadharan, L., P. J. Grossman, and J. Vecci (2021). Moving on up: The impact of income mobility on antisocial behaviour. *European Economic Review 134*, 103686.

Greiner, B. (2004). The online recruitment system orsee 2.0-a guide for the organization of experiments in economics. Technical report.

Hargreaves Heap, S. P., J. H. Tan, and D. J. Zizzo (2013). Trust, inequality and the market. *Theory and Decision 74*(3), 311–333.

Hey, J. D. (2001). Does repetition improve consistency? *Experimental economics 4*, 5–54.

Jäntti, M. and S. P. Jenkins (2015). Income mobility. In *Handbook of income distribution*, Volume 2, pp. 807–935. Elsevier.

Lange, A., J. Schmitz, and C. Schwirplies (2021). Inequality, role reversal and cooperation in multiple group membership settings. *Experimental economics*, 1–43.

Martinangeli, A. F. and L. Windsteiger (2021). Last word not yet spoken: a reinvestigation of last place aversion with aversion to rank reversals. *Experimental Economics 24*(3), 800–820.

Pop-Eleches, C. and M. Urquiola (2013). Going to a better school: Effects and behavioral responses. *American Economic Review 103*(4), 1289–1324.

Saito, K. (2013). Social preferences under risk: Equality of opportunity versus equality of outcome. *American Economic Review 103*(7), 3084–3101.

Stringhi, A. (2026). Guilt and fairness. *Games and Economic Behavior 158*, 191–218.

Tabellini, G. (2008). The scope of cooperation: Values and incentives. *The Quarterly Journal of Economics 123*(3), 905–950.

A Appendix

A.1 Rich receivers' sharing threshold

This appendix section derives the sharing threshold for Rich receivers under the two utility specifications used in the paper. Consider a Rich receiver with endowment $w_R > w$. Conditional on the Sender choosing *In*, the two relevant terminal histories are

$$(\textit{In}, \textit{Share}): \quad (\pi_s, \pi_R) = (2w, 2w_R),$$

and

$$(\textit{In}, \textit{Take}): \quad (\pi_s, \pi_R) = (0, 2w_R + g).$$

Fehr–Schmidt benchmark. Under the Fehr–Schmidt utility in equation (1), the Rich receiver is ahead of the Sender at both terminal histories. Hence only the advantageous-inequality term is active. The utility from sharing is

$$u_R(\textit{In}, \textit{Share}) = 2w_R - \theta_R^S(2w_R - 2w),$$

whereas the utility from taking is

$$u_R(\textit{In}, \textit{Take}) = 2w_R + g - \theta_R^S(2w_R + g).$$

The Rich receiver chooses *Share* if and only if

$$2w_R - \theta_R^S(2w_R - 2w) \geq 2w_R + g - \theta_R^S(2w_R + g).$$

Rearranging gives

$$\theta_R^S\Big[(2w_R + g) - (2w_R - 2w)\Big] \geq g,$$

or equivalently

$$\theta_R^S(g + 2w) \geq g.$$

Thus the threshold level of advantageous-inequality aversion required for sharing is

$$\hat{\theta}_R^S = \frac{g}{g + 2w}.$$

Importantly, the threshold does not depend on the Rich receiver's endowment w_R.

Forward-looking specification. We now consider the forward-looking utility in equation (3). In the Rank Reversal treatment, let p denote the probability that a currently Rich receiver occupies the Poor position in the following period. Define the forward-looking endowment benchmark for a currently Rich receiver as

$$\widetilde{w}_R(\delta, p) = \frac{(1 + \delta - \delta p)w_R + \delta p w_P}{1 + \delta}.$$

Under our calibration, $\widetilde{w}_R(\delta, p) > w$ for all $\delta \in [0, 1]$. Hence, the Rich receiver is ahead in the relevant averaged comparison term under both (*In*, *Share*) and (*In*, *Take*), so only the advantageous-inequality term is active.

The utility from sharing is therefore

$$u_R^0(\textit{In}, \textit{Share}) = 2w_R - \theta_R^S\left(2\widetilde{w}_R(\delta, p) - 2w\right),$$

whereas the utility from taking is

$$u_R^0(In, Take) = 2w_R + g - \theta_R^S \left(2\widetilde{w}_R(\delta, p) + g\right).$$

The Rich receiver chooses *Share* if and only if

$$2w_R - \theta_R^S \left(2\widetilde{w}_R(\delta, p) - 2w\right) \geq 2w_R + g - \theta_R^S \left(2\widetilde{w}_R(\delta, p) + g\right).$$

Rearranging yields

$$\theta_R^S \left[2\widetilde{w}_R(\delta, p) + g - \left(2\widetilde{w}_R(\delta, p) - 2w\right)\right] \geq g.$$

The forward-looking benchmark cancels out, leaving

$$\theta_R^S(g + 2w) \geq g.$$

Therefore, under the forward-looking specification as well,

$$\hat{\theta}_R^S = \frac{g}{g + 2w}.$$

Thus, both utility specifications imply the same sharing threshold for Rich receivers. The threshold is independent of w_R, and under the forward-looking specification it is also independent of δ and p. This invariance is the calibration property used in the design: treatment variation is predicted to affect Poor-receiver matches, while Rich-receiver incentives are held fixed.

A.2 Descriptive Analysis

Table A1: Descriptive statistics by match type and treatment

Outcome	T0	T1	T2	N
Poor-receiver matches				
Trust	0.420 (0.359)	0.565 (0.310)	0.375 (0.291)	40
Trustworthiness	0.340 (0.341)	0.505 (0.356)	0.350 (0.336)	20
First-order beliefs	0.387 (0.181)	0.495 (0.180)	0.376 (0.201)	40
Second-order beliefs	0.431 (0.197)	0.458 (0.206)	0.314 (0.185)	20
Rich-receiver matches				
Trust	0.710 (0.314)	0.665 (0.352)	0.695 (0.272)	40
Trustworthiness	0.625 (0.352)	0.635 (0.408)	0.625 (0.410)	20
First-order beliefs	0.494 (0.186)	0.523 (0.219)	0.493 (0.187)	40
Second-order beliefs	0.547 (0.232)	0.543 (0.217)	0.558 (0.236)	20

Entries report subject-level averages computed over the 10 rounds of the first block. Standard deviations are in parentheses. N denotes the number of subjects in each treatment cell.

Here, we provide additional descriptive evidence at the session level for period-1 matches involving "Poor" receivers. Because treatments were assigned at the session level, these figures complement the pooled descriptive analysis in the main text by showing whether the same patterns are visible across sessions. In particular, they allow us to assess whether the differences reported in the main text are broadly reflected in session averages rather than being driven solely by pooled individual observations. Following a similar logic, we also provide evidence of behavioral differences round-by-round.

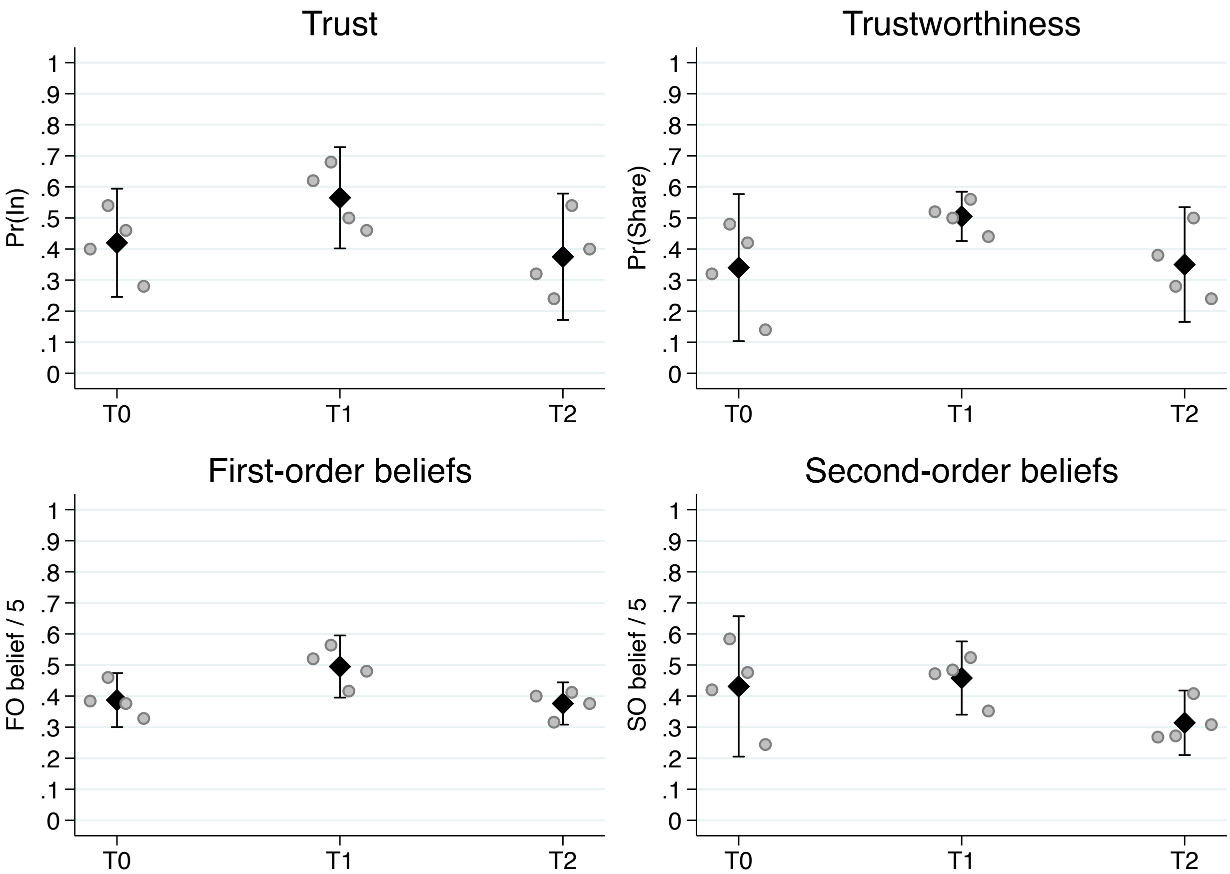

Figure A1: Session-level evidence for poor-receiver matches: Trust, trustworthiness, and first- and second-order beliefs across treatments. Grey circles denote session averages; black diamonds denote treatment means; whiskers indicate 95% confidence intervals across sessions. Outcomes are averaged across the 10 rounds of the first block. Treatments are ordered T0 = High Inequality, T1 = Low Inequality, and T2 = Rank Reversal.

Figure A1 shows that the main descriptive patterns reported in the text are also visible at the session level. For trust, trustworthiness, and senders' first-order beliefs, session means tend to be higher under *Low Inequality* (T1) than under *High Inequality* (T0), while *Rank Reversal* (T2) remains broadly similar to T0. Poor receivers' second-order beliefs appear noisier, with less pronounced differences across treatments. Taken together, the figure is consistent with the interpretation that the main descriptive differences in period-1 poor-receiver matches are associated with T1 rather than T2.

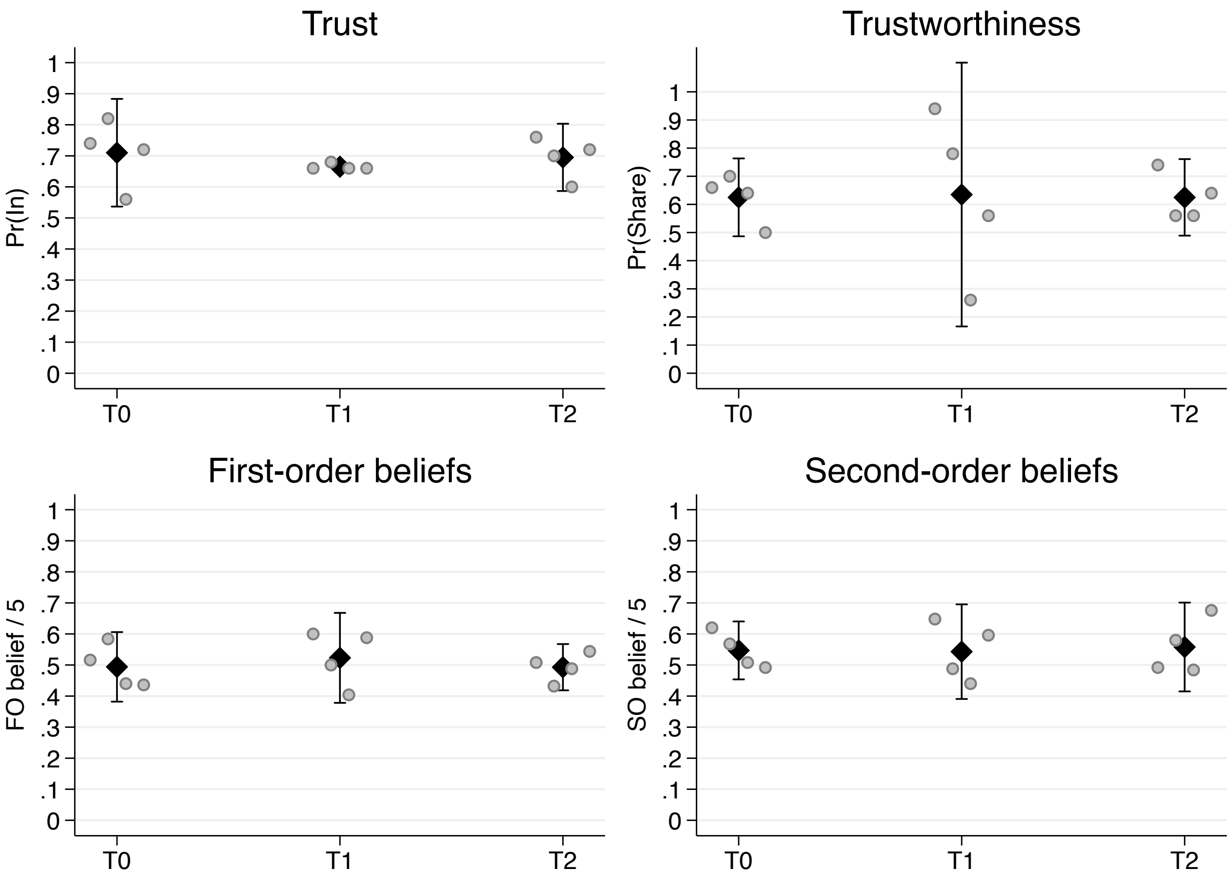

Figure A2: Session-level evidence for rich-receiver matches: Trust, trustworthiness, and first- and second-order beliefs across treatments. Grey circles denote session averages; black diamonds denote treatment means; whiskers indicate 95% confidence intervals across sessions. Outcomes are averaged across the 10 rounds of the first block. Treatments are ordered T0 = High Inequality, T1 = Low Inequality, and T2 = Rank Reversal.

Figure A2 shows that, in period-1 matches involving "Rich" receivers, session means for trust, trustworthiness, and both first- and second-order beliefs remain broadly stable across treatments. In particular, the figure does not reveal a clear descriptive pattern comparable to that observed in poor-receiver matches. Taken together, the session-level evidence is consistent with the interpretation that the main treatment differences are concentrated in poor-receiver matches.

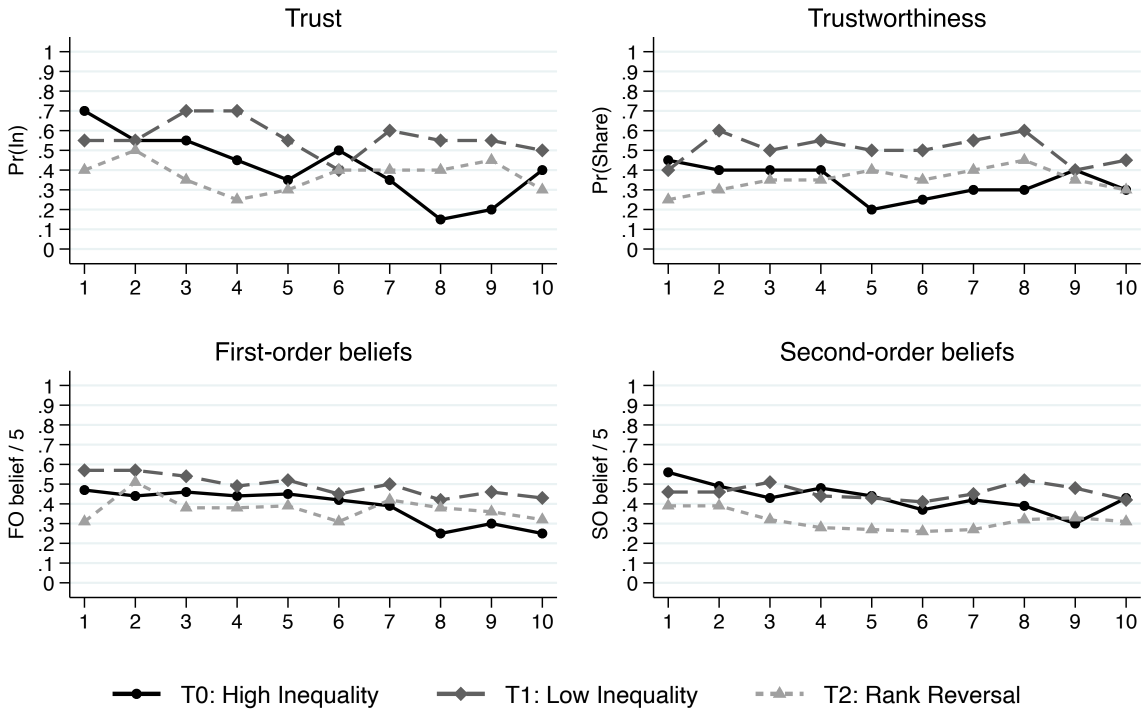

Figure A3: Round-by-round evidence for poor matches: Trust, trustworthiness, and first- and second-order beliefs across treatments over the 10 rounds of the first block. Each point denotes the treatment-specific average in a given round, with lines connecting round averages over time. Treatments are ordered T0 = High Inequality, T1 = Low Inequality, and T2 = Rank Reversal.

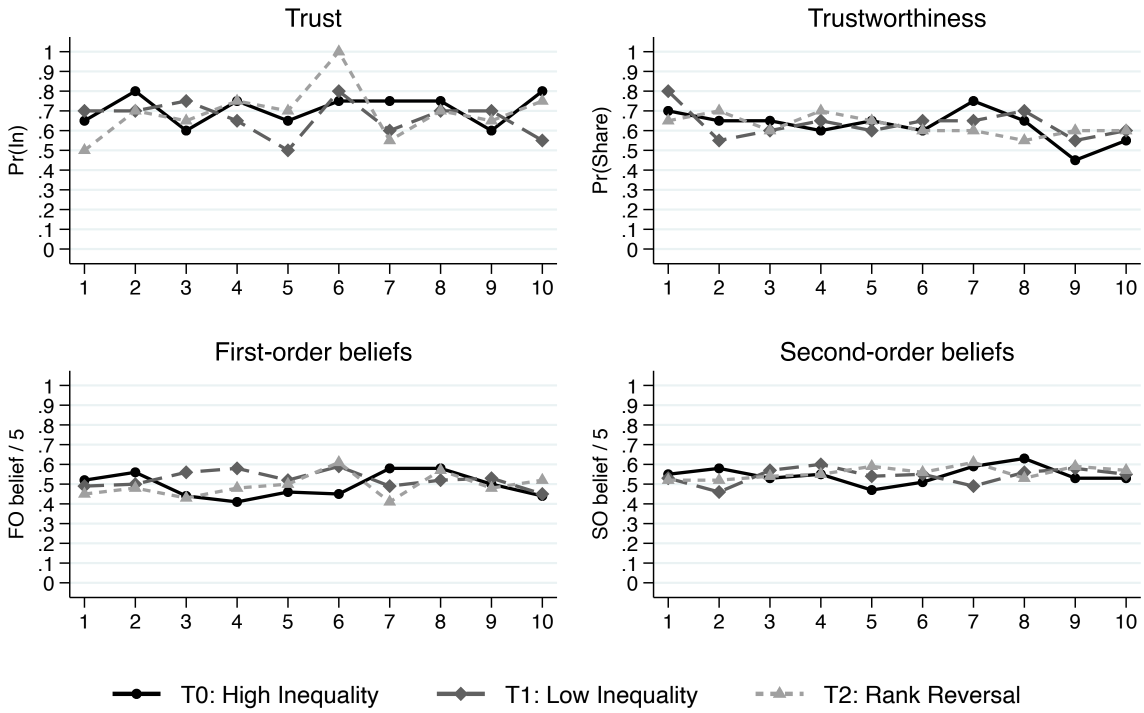

Figure A4: Round-by-round evidence for rich matches: Trust, trustworthiness, and first- and second-order beliefs across treatments over the 10 rounds of the first block. Each point denotes the treatment-specific average in a given round, with lines connecting round averages over time. Treatments are ordered T0 = High Inequality, T1 = Low Inequality, and T2 = Rank Reversal.

The round-by-round evidence in Figures A3 and A4 shows that the average treatment differences documented in the main text are not attributable to idiosyncratic variation in a small number of rounds. In both poor and rich matches, the treatment profiles evolve in a broadly stable manner over time, reinforcing the view that the reported averages capture persistent treatment effects rather than round-specific noise.

A.3 Econometric Analysis

In this Section, we report the tables with all the estimation results and average marginal treatment effects.

Table A2: Senders' trust: coefficients and average marginal effects

	T1–T0	T2–T0	T1–T2
A. Probit coefficients			
Low Inequality	-0.182		-0.063
	(0.160)		(0.130)
Rank Reversal		-0.086	
		(0.186)	
Poor receiver	-0.981***	-0.933***	-0.942***
	(0.200)	(0.191)	(0.273)
Low Inequality × Poor receiver	0.651***		0.616**
	(0.240)		(0.300)
Rank Reversal × Poor receiver		-0.024	
		(0.329)	
Constant	-0.270	-0.073	-0.945
	(1.433)	(0.716)	(1.331)
$\ln(\sigma_u^2)$	-0.400	-0.931**	-1.013**
	(0.290)	(0.417)	(0.415)
B. Marginal effects			
Poor receiver match	0.141**	-0.035	0.184***
	(0.067)	(0.075)	(0.068)
Rich receiver match	-0.049	-0.025	-0.019
	(0.043)	(0.054)	(0.039)
Observations	800	800	800

Notes: Panel A reports random-effects Probit estimates for the sender's decision to choose *In*; Panel B reports the corresponding average marginal effects by partner type. Columns (1)–(3) compare T1 vs. T0, T2 vs. T0, and T1 vs. T2, respectively. In Panel A, the omitted categories are the first treatment in each comparison and matches with a rich receiver. All specifications include age, gender, and round fixed effects (not reported). Standard errors clustered at the session level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Senders' first-order beliefs: coefficients and treatment effects

	T1–T0	T2–T0	T1–T2
A. Linear-model coefficients			
Low Inequality	0.023		0.022
	(0.048)		(0.042)
Rank Reversal		-0.002	
		(0.038)	
Poor receiver	-0.107***	-0.107***	-0.117***
	(0.014)	(0.014)	(0.018)
Low Inequality × Poor receiver	0.079**		0.089**
	(0.040)		(0.041)
Rank Reversal × Poor receiver		-0.010	
		(0.023)	
Constant	0.563***	0.339**	0.466***
	(0.136)	(0.148)	(0.162)
B. Treatment effects			
Poor receiver match	0.102***	-0.012	0.111***
	(0.036)	(0.032)	(0.032)
Rich receiver match	0.023	-0.002	0.022
	(0.048)	(0.038)	(0.042)
Observations	800	800	800

Notes: Panel A reports linear random-effects estimates for senders' first-order beliefs on the relevant pairwise sample: column (1) T1 vs. T0, column (2) T2 vs. T0, and column (3) T1 vs. T2. The omitted categories are the first treatment in each comparison and matches with a rich receiver. Panel B reports the implied treatment effects for poor and rich receiver matches. All specifications include age, gender, and round fixed effects. Standard errors clustered at the session level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Receivers' sharing: coefficients and average marginal effects

	Poor receivers			Rich receivers		
	T1–T0	T2–T0	T1–T2	T1–T0	T2–T0	T1–T2
A. Probit coefficients						
Low Inequality	0.889**		0.972***	0.121		0.119
	(0.411)		(0.291)	(0.835)		(1.643)
Rank Reversal		-0.595			0.011	
		(0.649)			(0.368)	
Constant	-8.113***	-0.493	-4.213	4.497	2.393	6.518
	(2.680)	(2.505)	(4.148)	(3.052)	(3.003)	(9.428)
$\ln(\sigma_u^2)$	0.382	0.717**	0.715**	1.384***	1.267***	2.282
	(0.273)	(0.327)	(0.283)	(0.437)	(0.357)	(1.664)
B. Marginal effects						
Treatment effect	0.197**	-0.121	0.210***	0.019	0.002	0.013
	(0.089)	(0.122)	(0.048)	(0.134)	(0.064)	(0.180)
Observations	400	400	400	400	400	400

Notes: Panel A reports random-effects Probit estimates for the receiver's decision to choose *Share*; Panel B reports the corresponding average marginal effects. Columns (1) and (4) compare T1 vs. T0, columns (2) and (5) T2 vs. T0, and columns (3) and (6) T1 vs. T2. Poor and rich receivers are estimated separately, with the first treatment in each comparison as the omitted category. All specifications include age, gender, and round fixed effects. Standard errors clustered at the session level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Receivers' second-order beliefs: linear-model coefficients

	Poor receivers			Rich receivers		
	T1–T0	T2–T0	T1–T2	T1–T0	T2–T0	T1–T2
Low Inequality	0.045		0.140***	-0.023		-0.007
	(0.071)		(0.049)	(0.061)		(0.069)
Rank Reversal		-0.159**			-0.015	
		(0.078)			(0.047)	
Constant	0.014	0.735***	-0.087	0.709**	0.805**	0.815**
	(0.345)	(0.131)	(0.338)	(0.337)	(0.314)	(0.326)
Observations	400	400	400	400	400	400

Notes: Each column reports linear random-effects estimates for receivers' second-order beliefs. Columns (1) and (4) compare T1 vs. T0, columns (2) and (5) T2 vs. T0, and columns (3) and (6) T1 vs. T2. Poor and rich receivers are estimated separately, with the first treatment in each comparison omitted. All specifications include age, gender, and round fixed effects (not reported). Standard errors clustered at the session level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.4 Robustness Analysis

Receivers (fully saturated model). As a robustness check, we estimate a pooled receiver specification that allows both the treatment effect and all nuisance parameters to vary by receiver type. Specifically, we estimate the following fully saturated random-effects Probit model:

$$\begin{aligned} Share_{ir} = \alpha + \beta_1 Poor_i + \beta_2 Treat_i \\ + \beta_3\Big(Poor_i \times Treat_i\Big) + \boldsymbol{\gamma}'\mathbf{Z}_{ir} + \boldsymbol{\delta}'\Big(Poor_i \times \mathbf{Z}_{ir}\Big) + u_i + \varepsilon_{ir}, \end{aligned} \tag{9}$$

The vector $\mathbf{Z}_{ir}$ includes age, gender, and round fixed effects; u_i denotes an individual-specific random effect and ε_{ir} an idiosyncratic error term. By interacting $Poor_i$ with $\mathbf{Z}_{ir}$, the specification allows all nuisance parameters, as well as the treatment effect, to vary by receiver type. It thus provides a pooled counterpart to the split-sample receiver regressions in the main text, while relaxing the restriction that these coefficients be identical for poor and rich receivers. Table A6 reports the estimated coefficients (Panel A) and the corresponding average marginal effects for poor and rich receivers (Panel B).

Table A6: Receivers' sharing: coefficients and average marginal effects (fully saturated model)

	T1–T0	T2–T0	T1–T2
A. Probit coefficients			
Low Inequality	0.104		0.127
	(0.739)		(0.865)
Rank Reversal		0.003	
		(0.345)	
Poor receiver	-13.092***	-2.760	-9.387*
	(3.997)	(3.460)	(5.697)
Low Inequality × Poor receiver	0.885		1.125
	(0.783)		(0.873)
Rank Reversal × Poor receiver		-0.676	
		(0.599)	
Constant	3.869	2.217	4.250
	(2.646)	(2.855)	(3.560)
$\ln(\sigma_u^2)$	0.905***	1.015***	1.391***
	(0.213)	(0.259)	(0.273)
B. Marginal effects			
Poor receiver match	0.187**	-0.123	0.212***
	(0.086)	(0.118)	(0.046)
Rich receiver match	0.019	0.000	0.020
	(0.139)	(0.118)	(0.137)
Observations	800	800	800

Notes: Panel A reports random-effects Probit estimates for receivers' choice of *Share*. Panel B reports the corresponding average marginal effects by receiver type. Columns (1)–(3) compare T1 vs. T0, T2 vs. T0, and T1 vs. T2, respectively. All specifications include age, gender, and round fixed effects (not reported). Standard errors clustered at the session level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The fully saturated pooled specification confirms the pattern identified in the main empirical analysis, with marginal effects estimates that are comparable in magnitude and statistical significance to those reported in the baseline specifications.

Robustness to past feedback. Since participants observe their counterpart's choice at the end of each round, behavior in later rounds may partly reflect learning from previous interactions. We therefore estimate the treatment effect conditional on feedback by re-estimating the baseline specifications for Senders and Receivers, separately for Poor and Rich matches, while controlling for the counterpart's choice in the previous round. The figures below report the corresponding average marginal treatment effects. The estimates reproduce the same pattern of treatment differences documented in the main analysis.

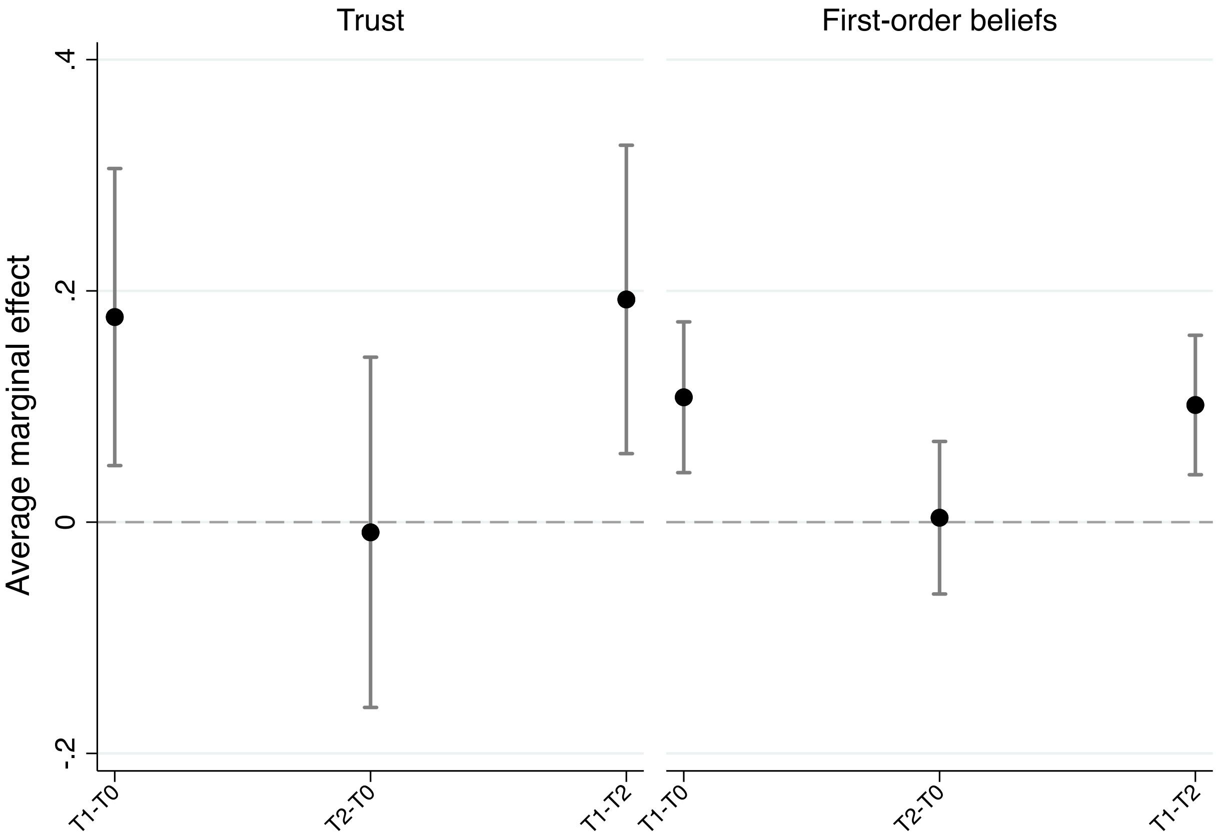

Figure A5: Marginal treatment effects for senders (poor-receiver matches). The left panel shows trust; the right panel shows first-order beliefs. Points denote estimates and whiskers 95% confidence intervals. The reported estimates control for the poor receivers choice at the previous round.

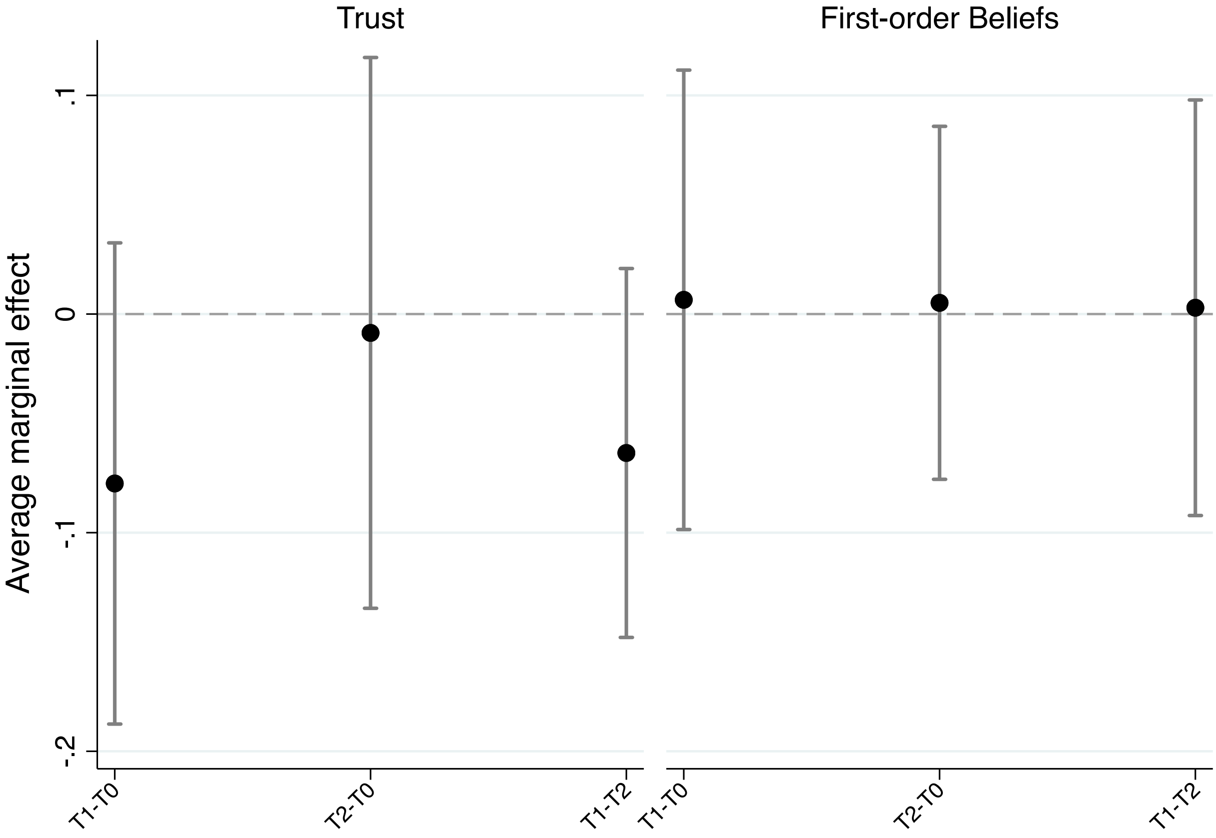

Figure A6: Marginal treatment effects for senders (rich-receiver matches). The left panel shows trust; the right panel shows first-order beliefs. Points denote estimates and whiskers 95% confidence intervals. The reported estimates control for the rich receivers choice at the previous round.

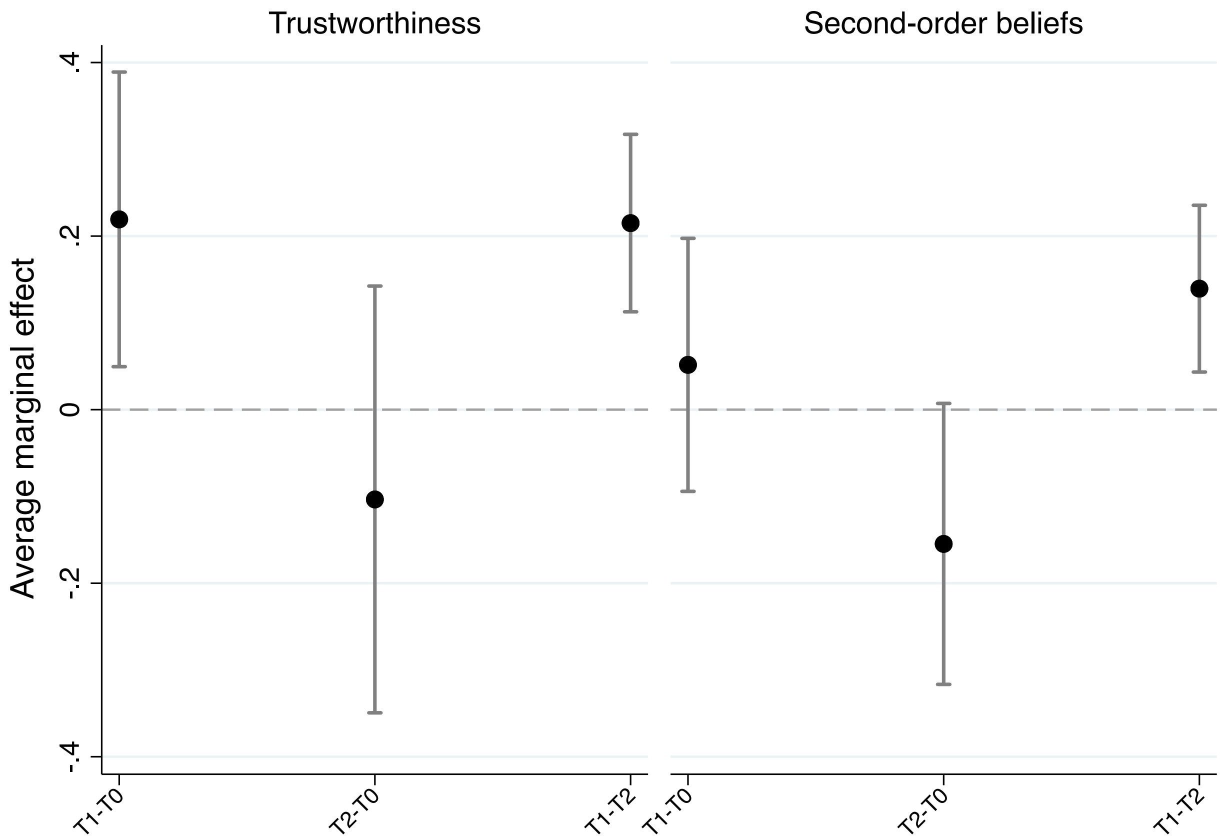

Figure A7: Average marginal treatment effects for poor receivers. The left panel shows sharing and the right panel second-order beliefs. Points denote estimates and whiskers 95% confidence intervals. The reported estimates control for sender' choice at the previous round.

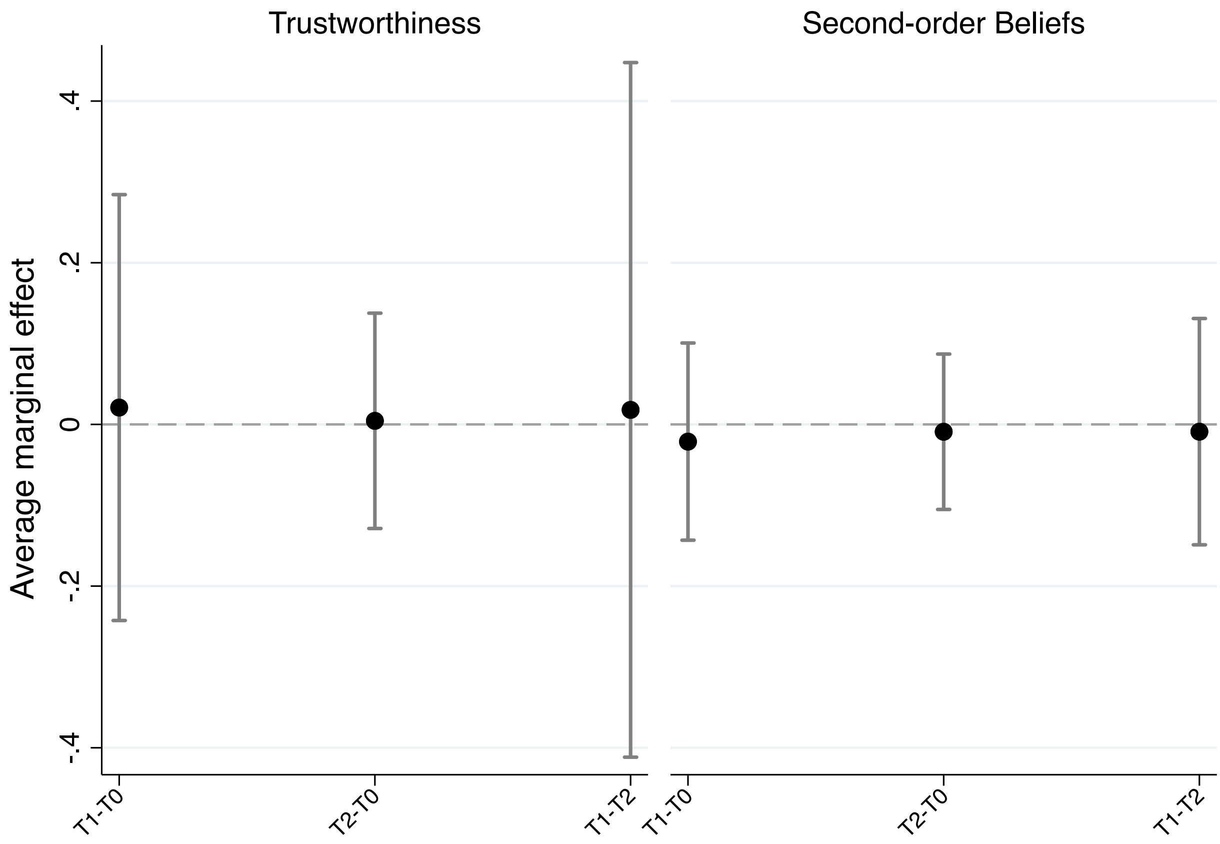

Figure A8: Average marginal treatment effects for rich receivers. The left panel shows sharing and the right panel second-order beliefs. Points denote estimates and whiskers 95% confidence intervals. The reported estimates control for sender' choice at the previous round.

Instructions – Roles and Endowments

Welcome. Thank you for participating in this experiment. Please do not talk to other participants during the session. If you have any questions, feel free to raise your hand and ask the lab assistants.

At the beginning of the experiment, you will be randomly assigned to **role A** or **role B**. If you are assigned to **role A**, you will be randomly paired with a participant in **role B**, and vice versa. You will remain in the same role for the entire session.

Each participant will be assigned to an endowment bracket and, according to that, they will receive an initial monetary endowment expressed in tokens (*1 token = 10 CZK*).

Participants in **role A** are in the Medium bracket and they will receive *6 tokens*.

Participants in **role B** will be evenly and randomly divided into two brackets.

Treatments T0 and T2:

- **Low bracket** : Endowment of 2 tokens
- **High bracket** : Endowment of 10 tokens

Treatments T1:

- **Low bracket** : Endowment of 4 tokens
- **High bracket** : Endowment of 8 tokens

The decision-making phase is divided into two blocks of **10 rounds** each. In the first block, you will be paired in each round with a different participant from the other role; within this block, you will never interact with the same person more than once. In the second block, you will again play one round with each of the same participants, but the order of the pairings will be different from the first block. At all times, your decisions remain completely anonymous: you will not know the identity of the other participants, and they will not know yours.

Participants in **role A** will be informed whether the **role B** participant they are paired with belongs to the low or high bracket.

Additional information for Treatment T2 (Mobility) only

Endowment change

Between the two blocks (i.e., after 10 rounds),each ***role B*** *participant in the low bracket has a 50% chance of being moved to the high bracket, and each* ***role B*** *participant in the high bracket has a 50% chance of being moved to the low bracket. The total number of participants in each bracket will remain the same across the two blocks.*

Instructions – Decision Task

During each round of the decision-making phase, the participant in **role A** chooses between two actions: **Stop** and **Continue**. Simultaneously, the participant in **role B** chooses between two actions: **Take** and **Share**.

Summary Table:		**Earnings Participant A**	**Earnings Participant B**	
			Low bracket	**High bracket**
A chooses **Stop**		6 tokens	2 tokens	10 tokens
A chooses **Continue**	B chooses **Take**	0 tokens	6 tokens	22 tokens
	B chooses **Share**	12 tokens	4 tokens	20 tokens

Figure A9: Payoffs in Treatments T0 and T2.

Summary Table:		**Earnings Participant A**	**Earnings Participant B**	
			Low bracket	**High bracket**
A chooses **Stop**		6 tokens	4 tokens	8 tokens
A chooses **Continue**	B chooses **Take**	0 tokens	10 tokens	18 tokens
	B chooses **Share**	12 tokens	8 tokens	16 tokens

Figure A10: Payoffs in Treatments T1.

Decisions are made independently and simultaneously. Neither participant knows the other's choice until the outcome is revealed.

If the participant in **role A** chooses **Stop**, both A and B will receive their initial monetary endowment.

If A chooses **Continue** and B chooses **Take**, A will receive 0 tokens, while B will receive:

- 6 tokens if they belong to the low bracket;
- 22 tokens if they belong to the high bracket.

} Treatments T0 and T2

- 10 tokens if they belong to the low bracket;
- 18 tokens if they belong to the high bracket.

} Treatments T1

If A chooses **Continue** and B chooses **Share**, A will obtain 12 tokens while B will receive:

- 4 tokens if they belong to the low bracket;
- 20 tokens if they belong to the high bracket.

} Treatments T0 and T2

- 8 tokens if they belong to the low bracket;
- 16 tokens if they belong to the high bracket.

} Treatments T1

This information is summarized in the table below:

Guesses and Estimates

Before making your choice, we will ask you to make an estimate if you are in **role A** or make a guess if you are in **role B**. In detail:

- **role A**: In each round, you will estimate how many participants in role B *within the same bracket (High or Low) as the B participant you are matched with in that round* will choose **Share** in that round. The estimate can be any number from **0** to **5**. Your estimate is evaluated against the actual number of role-B participants in that bracket who choose **Share** in that round. Each correct estimate earns **1 token**.
- **role B**: In each round, you will guess the estimate made by the role-A participant you are paired with in that round. The guess can be any number from **0** to **5**. Your guess is evaluated against the estimate actually made by your paired A in that round. Each correct guess earns **1 token**.

Payments

At the end of the experiment, the computer will randomly select one round within each block. You will be paid based on the choices made by you and the participant you were paired with during those two rounds of the decision-making phase.

Additionally, you will be paid for each correct estimate or guess made during all decision-making phases. Furthermore, you will receive a fixed participation fee for taking part in this session and for answering the final questionnaires.