Pre-Analysis Plan

1 Conceptual focus

Our main objects of interest in a Trust Mini Game are:

- **Trust**, measured by Senders' decisions to play *In*, and
- **Trustworthiness**, measured by Receivers' decisions to play *Share*.

The theoretical mechanism is:

1. Reducing inequality (via redistribution in T1 or mobility in T2) increases the **probability that Poor receivers choose Share**, driven by inequity aversion.
2. Anticipating this, Senders become more willing to choose *In*, especially when matched with Poor receivers.

We treat:

- **Primary outcomes:**
  - Senders' *In* decisions (trust)
  - Receivers' *Share* decisions (trustworthiness)

Primary tests focus on **block 1 (first 10 rounds)**, where the theoretical environment is cleanest (before rank reversal in T2 and before substantial learning).

2 Data structure, dynamics, and identification

Treatments T0 (baseline), T1 (redistribution), and T2 (mobility / rank reversal) are randomly assigned **between subjects** at the session level. Within each treatment:

- Each Sender and each Receiver participates in **10 rounds in block 1**, then 10 rounds in block 2.
- Matching between Senders and Receivers across rounds is random (absolute stranger matching).
- The **partner's type** (Poor vs Rich receiver) is random, and known, from the perspective of the Sender, conditional on treatment and round.

Dynamic information within block 1

Within block 1, players observe the outcome of each round, so their decisions can depend not only on the current treatment and partner type but also on past outcomes – e.g., whether a previous Poor receiver Shared or Took – and on their own earlier choices. As a result, behavior in block 1 exhibits serial correlation and learning dynamics.

However, because treatment status (T0, T1, T2) is randomly assigned and fixed for each individual, the distribution of histories within each treatment arm becomes part of the treatment environment. Our primary estimands are average treatment effects in block 1—conditional on the current partner type and round—while integrating over the history distribution generated by each treatment.

The panel structure allows us to include round fixed effects to capture common behavioral trends over time and to compute cluster-robust standard errors at the individual level. However, we do not use within-individual fixed effects or dynamic panel estimators in the primary analysis.

3 Block-1 econometric specification

All primary analyses use only **block 1 (rounds 1–10)**.

3.1 Notation

Let $r \in \{1, \ldots, 10\}$ index rounds in block 1, subject $\ell \in \{i, j\}$ with j index Senders, and i index Receivers. Treatments $T \in \{T0, T1, T2\}$ are assigned at the individual (session) level.

Senders.

- $In_{jr} = 1$ if Sender j chooses *In* in round r; 0 otherwise.
- $PoorPartner_{jr} = 1$ if the matched receiver in round r is Poor; 0 if Rich.

Receivers.

- $Share_{ir} = 1$ if Receiver i chooses *Share* in round r; 0 otherwise.
- $Poor_i = 1$ if Receiver i is Poor (in the current block/treatment); 0 if Rich.

For any *ordered pair* of treatments $(A, B) \in \{(T1, T0), (T2, T0), (T1, T2)\}$, define

$$D_\ell^{A,B} = \begin{cases} 1 & \text{if subject } \ell \text{ is in treatment } A, \\ 0 & \text{if subject } \ell \text{ is in treatment } B, \end{cases}$$

and restrict the sample to subjects with $T \in \{A, B\}$. We include a full set of round fixed effects λ_r (Senders) or η_r (Receivers). All regressions are estimated by OLS (Linear Probability Models) with standard errors clustered at the **session** level.

3.2 Pairwise models for Senders (trust)

For each treatment pair (A, B) we estimate, on Senders in A or B:

$$In_{jr} = \alpha_{A,B}^{In} + \beta_{1,A,B}^{In} PoorPartner_{jr} + \beta_{2,A,B}^{In} D_j^{A,B} + \beta_{3,A,B}^{In} (PoorPartner_{jr} \times D_j^{A,B}) + \lambda_{r,A,B}^{In} + \varepsilon_{jr,A,B}^{In}. \tag{1}$$

For this pair (A, B), the implied treatment effects on trust are:

$$\Delta^{R,In}_{A,B} := E[In \mid T = A,\ \text{Rich}] - E[In \mid T = B,\ \text{Rich}] = \beta^{In}_{2,A,B}, \tag{2}$$

$$\Delta^{P,In}_{A,B} := E[In \mid T = A,\ \text{Poor}] - E[In \mid T = B,\ \text{Poor}] = \beta^{In}_{2,A,B} + \beta^{In}_{3,A,B}. \tag{3}$$

Thus $\Delta^{P,In}_{A,B}$ is the effect of treatment A vs B on the probability of In when matched with a Poor receiver (our core estimand for trust), while $\Delta^{R,In}_{A,B}$ is the corresponding effect when matched with a Rich receiver (ancillary).

3.3 Pairwise models for Receivers (trustworthiness)

Similarly, for each treatment pair (A, B) we estimate, on Receivers in A or B:

$$Share_{ir} = \alpha^{Share}_{A,B} + \gamma^{Share}_{1,A,B}\, Poor_i + \gamma^{Share}_{2,A,B}\, D^{A,B}_i + \gamma^{Share}_{3,A,B}\, (Poor_i \times D^{A,B}_i) + \eta^{Share}_{r,A,B} + u^{Share}_{ir,A,B}. \tag{4}$$

For this pair (A, B), the implied treatment effects on sharing are:

$$\Delta^{R,Share}_{A,B} := E[Share \mid T = A,\ \text{Rich}] - E[Share \mid T = B,\ \text{Rich}] = \gamma^{Share}_{2,A,B}, \tag{5}$$

$$\Delta^{P,Share}_{A,B} := E[Share \mid T = A,\ \text{Poor}] - E[Share \mid T = B,\ \text{Poor}] = \gamma^{Share}_{2,A,B} + \gamma^{Share}_{3,A,B}. \tag{6}$$

Here $\Delta^{P,Share}_{A,B}$ is the effect of A vs B on the probability of *Share* for Poor receivers (core estimand for trustworthiness), and $\Delta^{R,Share}_{A,B}$ is the corresponding effect for Rich receivers (ancillary).

All these effects are *block-1* differences in the relevant probabilities, conditional on current partner type and round, and averaged over the distribution of histories induced by each treatment.

4 Primary hypotheses (block 1)

All primary hypotheses refer to **block 1 (rounds 1–10)** and focus on Poor partners/receivers. We use one-sided tests with $\alpha = 0.05$.

4.1 Trust (Senders' In) – core hypotheses (Poor partners)

- **H1 (Redistribution increases trust towards Poor, T1 vs T0)**

$$\text{Estimand: } \Delta^{P,In}_{T1,T0}, \quad H_1 : \Delta^{P,In}_{T1,T0} > 0.$$

- **H2 (Mobility is not worse than baseline for trust towards Poor, T2 vs T0)**

$$\text{Estimand: } \Delta^{P,In}_{T2,T0}, \quad H_2 : \Delta^{P,In}_{T2,T0} \geq 0,$$

implemented as $H_0 : \Delta^{P,In}_{T2,T0} \leq 0$ vs $H_1 : \Delta^{P,In}_{T2,T0} > 0$.

- **H3 (Redistribution $\geq$ mobility in trust towards Poor, T1 vs T2)**

$$\text{Estimand: } \Delta^{P,In}_{T1,T2}, \quad H_3 : \Delta^{P,In}_{T1,T2} \geq 0.$$

4.2 Trust towards Rich partners (ancillary)

For Rich partners we use two-sided exploratory tests:

- T1 vs T0: $\Delta^{R,In}_{T1,T0}$, test $H^{In}_{R,T1} : \Delta^{R,In}_{T1,T0} = 0$.
- T2 vs T0: $\Delta^{R,In}_{T2,T0}$, test $H^{In}_{R,T2} : \Delta^{R,In}_{T2,T0} = 0$.

4.3 Sharing (Receivers' Share) – core hypotheses (Poor receivers)

- **H4 (Redistribution increases sharing by Poor, T1 vs T0)**

$$\text{Estimand: } \Delta^{P,Share}_{T1,T0}, \quad H_4 : \Delta^{P,Share}_{T1,T0} > 0.$$

- **H5 (Mobility is not worse than baseline for sharing by Poor, T2 vs T0)**

$$\text{Estimand: } \Delta^{P,Share}_{T2,T0}, \quad H_5 : \Delta^{P,Share}_{T2,T0} \geq 0,$$

tested as $H_0 : \Delta^{P,Share}_{T2,T0} \leq 0$ vs $H_1 : \Delta^{P,Share}_{T2,T0} > 0$.

- **H6 (Redistribution $\geq$ mobility in sharing by Poor, T1 vs T2)**

$$\text{Estimand: } \Delta_{T1,T2}^{P,Share}, \quad H_6 : \Delta_{T1,T2}^{P,Share} \geq 0.$$

4.4 Sharing by Rich receivers (ancillary)

For Rich receivers we use two-sided exploratory tests:

- T1 vs T0: $\Delta_{T1,T0}^{R,Share}$, test $H_{R,T1}^{Share} : \Delta_{T1,T0}^{R,Share} = 0$.
- T2 vs T0: $\Delta_{T2,T0}^{R,Share}$, test $H_{R,T2}^{Share} : \Delta_{T2,T0}^{R,Share} = 0$.

4.5 Discussion

These hypotheses are derived through our game-theoretic analysis, grounded in the assumption that individuals care about both equality of opportunity and equality of outcome. Technically, we rely on Saito (2013)'s model assuming the δ parameter to close to 0.5. In the case of subjects care mostly about equality of outcomes ($\delta \approx 0$), then H2 and H5 are confirmed with $T2 = T0$. Instead, if subjects care mostly about equality of opportunity ($\delta \approx 1$), then H3 and H6 are confirmed with $T2 = T1$. Note that the purpose of this experiment is to examine the relationship between equality of opportunity and equality of outcomes. Consequently, the rejection of all hypotheses (H1–H6) does not, by itself, imply that the experiment has failed.

5 Beliefs

Since we measure both first- and second-order beliefs, we treat them as additional primary outcomes. For senders' first-order beliefs — i.e., their estimated probability that the receiver chooses Share — we assume that senders are rational and believe that receivers are both rational and inequity-averse. Under these assumptions, senders' beliefs should align with the senders' actual behavior. Accordingly, we estimate an analogous model for senders' behavior, using senders' first-order beliefs as the dependent variable.

For receivers' second-order beliefs — i.e., their beliefs about the first-order beliefs of the matched sender — we assume that rational receivers believe that a rational sender believes them to be rational and inequity-averse. Under these assumptions, receivers' second-order beliefs should mirror the pattern of senders' first-order beliefs. Therefore, we identify the treatment effect by estimating the same model as the one for senders' first-order beliefs, but with receivers' second-order beliefs as the dependent variable.

6 Additional analyses

Lagged-history (LDV) robustness. As a robustness check on our block-1 results, we will estimate lagged-dependent-variable (LDV) specifications that condition current behavior on very recent history. For Senders, we will regress the In decision in round r on treatment indicators, current partner type, and their interactions, while additionally controlling for the Sender's own previous In decision and for the partner's previous Share/Take choice (for rounds $r \geq 2$). An analogous LDV specification will be estimated for Receivers' Share decisions. These LDV models are not alternative primary estimators: they identify treatment effects *conditional* on last-round behavior and partner history, which differ from the *total* effects in our main block-1 models. We will therefore use them to assess how much of the overall treatment effect appears to operate through very recent experiences, rather than to redefine our primary estimands.

Exploratory analysis of block 2. In addition, we will conduct an exploratory analysis that replicates the main block-1 specifications using data from block 2. Specifically, we will re-estimate the block-1 regressions for trust (Senders' In) and trustworthiness (Receivers' Share) using only rounds in block 2 and the same treatment and partner-type indicators. This analysis is intended to shed light on the stability or evolution of treatment effects over time (e.g. whether redistribution and mobility have persistent, attenuating, or strengthening impacts on behavior), but will be clearly labelled as exploratory and will not form part of our primary hypothesis tests.

References

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