

1 Abstract

2 Conceptual Framework

Consider two risk-neutral players who interact for one period in a trust game. They have group identity $g \in \{S, N\}$ where $n = \text{Prob}(g = s)$. Player i , the trustor, receives some endowment x . They may send some amount $t \in [0, x]$ to player j , the trustee. After i 's decision, player j receives $3t$. They may then return some amount $r \in [0, 3t]$ to i .

In the original setup by Berg et al. (1995), t is a measure of trust and r is a measure of trustworthiness. The validity of this claim depends on the researcher's definition of trust. By the OECD definition (among others), I define trust as the belief that one has that *another party will act consistently with their expectations of positive behaviour*. Thus in my case I am interested in measuring $\hat{r}_j^i(t) \equiv \mathbb{E}^i(r_j|t)$: player i 's belief of the return from player j .

The action by player i does reveal a kernel of truth of $\hat{r}_j^i$ however, it does not cleanly identify this. Suppose that the researcher is interested in learning the difference in mean trust between S and N groups when the identity of j is unknown to i . Suppose additionally that each player has some other-regarding preferences e.g. altruism denoted by $z_{\{i,j\}}$. Thus player i will believe in a conditional expected return of $\hat{r}_j^i(t, \mathbb{E}^i(z_j))$ which is a function of the amount they send t and their expectation of the other-regarding preferences of the trustee. Thus, the trustor maximises their utility by sending t^* which solves the following problem.

$$t^* = \arg \max_{t \in [0, x]} U_i(z_i, \hat{r}_j^i(t, \mathbb{E}^i(z_j)))$$

Let the average other-regarding preferences for group g be z^g , then the difference in average amounts sent will reflect differences in trusting beliefs but also differences in average other-regarding preferences and expectations of such. Lets assume that t^* is weakly monotonically increasing in z_i . Then if $z^S > z^N$, the utility from sending some transfer t that an S trustor derives will exceed the utility an N trustor derives. Thus if this is the case, then there will exist some $\epsilon > 0$ such that $t_N^* = t_S^* - \epsilon$ where $\hat{r}^N(t, \mathbb{E}^N(z_j)) > \hat{r}^S(t, \mathbb{E}^S(z_j))$. In other words, it is possible that N trustors send less than S trustors while having higher beliefs of conditional returns from the trustee.

To combat this identification problem, I propose to use a 3rd party spectator k who is asked to state their beliefs about $\hat{r}_j^i(t, \mathbb{E}^i(z_j))$ and z_j by guessing the outcome of the above trust game and a dictator game (where there is no option for a return amount). Such a practise avoids the confounding of z_i in trust transfer decisions and allows a cleaner identification of the trust that one has in a trustee.

3 Experimental Set-Up

In this section, I will describe the baseline experimental procedure which consists of (i) a participant pre-screening stage, (ii) the 'game' and (iii) the spectator. I will describe each of these in order and then discuss the manipulations specific to the pilot in the following section.

3.1 Pre-screening

Participants in the experiment need to be categorised by their identity. This pre-screening will be present at the beginning of each session for both the 'game' and the spectator setup and thus warrants its own explanation. I identify an individual as a Southerner (Northerner) if they meet the following criteria:

- The individual was born in a Southern (Northern) region¹
- The individual currently lives in a Southern (Northern) region.
- The individual self-identifies with the Southern (Northern) identity.

The aim of this screening depends on the experiment session that is being run. In categorising individuals, I wish to identify if one has moved into or out of some region which would alter their exposure to other identities. Under the contact hypothesis, this may affect one's trusting beliefs and pro-social behaviour. For instance, given that there is substantial net-South migration in England (ONS, 2023), Southerners will naturally have a higher exposure to Northerners than the contrary: a point I may exploit in a follow-up experiment.

In the 'game', I need to recruit at least 10 Northerners and 10 Southerners for each role. The remaining participants may be of any English origin. Therefore, I adopt a recruitment quota policy in the pre-screening.

In the spectator setting, the aim is to exclude individuals who do not identify as a Northerner or a Southerner. This is since these are my two identities of interest in the experiment.

Participants who do not meet the recruitment criteria in any stage are screened out of the study and paid a small compensation (£0.10) for their time. Participants who pass are moved on to the next stage of the experiment.

3.2 The 'game'

The purpose of the game is not to measure or identify differences in trust beliefs or behaviour between identities but merely to provide real incentives for the spectator

¹Southern regions are London, South East, South West and East of England. Northern regions are North East, North West and the Yorkshire and Humber.

round. Participants are paid £1.00 as a base wage for participation and have the opportunity to earn up to £18.00 in bonus payment.

The game is split into two sub-experiments. Each experiment recruits 30 individuals. In the first sub-experiment, a recruitment quota system is used to ensure that there are at least 10 Northerners and 10 Southerners in the pool. All other participants may be of any English identity. The quota is implemented via the pre-screening feature. Participants are informed that they will play 2 games with another individual that is English. The players do not learn the identity of the person that they play these games with. In practice, I pair participants up across these two sub-experiments: I do not need these individuals to play simultaneously.

Players in the 'game' are told that in each game, they will be assigned a role of either person A or person B, with their partner playing the remaining role. They are informed that each token has a monetary value of £2.00 and that they will be paid their token balance in cash for one of the two games. They then proceed to play a trust game and a dictator game in a randomised order.

In the trust game, they are given the following instructions.

Person A starts with 3 tokens, and Person B starts with none. Person A can send some, all, or none of their tokens to Person B. Any tokens kept by Person A are theirs. For every token sent, Person B receives 3 tokens. For example, if Person A sends 2 tokens, Person B will get 6 tokens. Person B then decides how many tokens to return to Person A. Person A's earnings are the tokens they kept plus any returned to them. Person B's earnings are what they received minus what they returned.

In practice, players in the first sub-experiment are assigned the role of person B (the trustee) and players in the second sub-experiment are assigned the role of person A (the trustor).

The trustees are asked to state how many tokens they would like to return to person A for each possible amount that they could receive (0, 3, 6 or 9 tokens). This is a strategy method of elicitation. The trustees are told that their partner will make their decision also and this, together with their own decisions, will determine their payoff.

In the dictator game, players are given the following instructions.

Person A first receives 5 tokens while person B receives nothing. Person A may send some, none, or all of their tokens to person B. Person A's earnings from this round are the tokens they do not send. Person B's earnings equal the tokens they receive.

In practice, players in the first sub-experiment are assigned the role of person A (the dictator) and players in the second experiment are assigned to the role of person B (the receiver). The dictators are asked to decide how many tokens they wish to send. The receivers do not make a decision in this stage.

Before each game, participants are required to pass a comprehension test before they proceed.

3.3 Spectator

Spectators are recruited to guess the distribution of outcomes among 10 pairs who played the game. Spectators must also pass the pre-screening test. This way we can measure the trust of Northerners and Southerners separately. Spectators are paid £1.00 for their participation in the survey.

For each game, the spectators are given the same instructions as the players of the game and are then told that they must guess the frequency at which each possible decision was made among 10 trustees. In other words, the spectators are told that there are 10 pairs who did this experiment. They are told that the 10 trustees stated how many tokens they would like to return to the trustor conditional on them receiving each possible amount. They are then asked to state their beliefs about the distribution of decisions among these 10 trustees. Thus they are asked to elicit their beliefs of 3 distributions (one for each conditional transfer amount from the trustor) and are awarded £2.00 if they guess the exact frequency distribution for each question. After being given a chance to adjust their estimates, the spectators are then asked to guess the distribution of initial transfers in the trust game among 10 trustors where 5 were Northern and 5 were Southern.

The same procedure occurs for the dictator game where the spectators must state their beliefs about the distribution of transfers made by 10 dictators in the game. Spectators are awarded £2.00 if they guess the exact distribution.

As a final check, I ask spectators how much they agree with the following statement using a 7 point Likert scale. I then ask them to state their belief of how many of 100 individuals of a specific identity agree or strongly agree with this statement. The statement is '*Most people can be trusted*'. Spectators are awarded £0.50 for each guess that is within 3 of the correct answer.

4 Pilot

In the pilot stage of this experiment, I aim to answer 3 questions:

1. Do Northerners have lower trusting beliefs about a general individual than Southerners?
2. Do spectators have higher or lower trusting beliefs about their in-group compared to their out-group?
3. Are these effects mirrored in beliefs about altruism?
4. Do individuals misestimate the level of trust among their identity?

To answer these questions, I employ 3 treatments which vary the identity of the trustee that the spectator observes. To control for the size of potential identities, I inform all participants that everyone who is recruited to the study either identifies as Northern or Southern and that when the identity of a player is unknown, there is an equal chance that the player is a Northerner or a Southerner. Importantly, I am not assuming that being Northern or Southern is random and thus causal interpretations of trust differences can be attributed to this. I will however assess causality via information provision of the identity of the trustee.

To answer the first question, I compare the beliefs of Northern and Southern spectators regarding the decisions of trustees whose identity is unknown. I will recruit equally as many Northerners and Southerners through a quota system in the pre-screening exercise.

To answer the second question, I will compare the beliefs of spectators about the trustworthiness of their in-group vs. their out-group. I will examine for potential heterogeneity among spectator identities too. The third question speaks to the beliefs of spectators in the dictator game.

The final question speaks to an exploratory analysis which will be used to motivate a further study over whether differences in trusting beliefs are rational. In essence, assuming that trust and trustworthiness are positively correlated, I am testing for whether individuals correctly guess the share of individuals of some identity that have trusting beliefs using the same question as in the World Value Survey.

4.1 Econometric specification

I will employ three tiers of formal tests to determine the difference in trust across treatments.

1. First, I compare the believed distribution of trust held by Northerners and Southerners. For this, I use a chi-squared test.
2. Second, I use the distributions to approximate the spectator's conditional expectation of returns and then use linear estimation methods to compute the marginal effect of identity on the trust exhibited by the spectator for all three possible transfers.
3. Third, I approximate the unconditional expectation of returns using two approaches: first by uniformly weighting the the conditional expectations of returns, and second by using the spectators' weights of each outcome occurring. Once more I will use linear methods to compute the marginal effect of identity.

Using these three tiers, I can estimate the difference in trusting beliefs between different treatment arms while controlling for different covariates.

Say that at some belief elicitation, the question asks the spectator to guess the frequency of $n = 3t$ outcomes occurring. Then this question elicits a discrete distribution of beliefs of returns made by trustee $R_i^t = (\rho_i^{0|t}, \rho_i^{1|t}, \dots, \rho_i^{n|t})$ where $t \in \{1, 2, 3\}$

and $\sum_{k=0}^n \rho_i^{k|t} = 1$ such that $\rho_i^{k|t}$ represents the share of outcomes that i believes will be k among trustees. Then in this case, I can compute the conditional expectation of returns as

$$\hat{r}^i(t) \equiv \mathbb{E}^i(r|t) = \sum_{k=0}^{3t} \rho_i^{k|t} k$$

This is the first dependent variable used in the second tier of econometric tests. In the third tier, I need to compute two unconditional expectations. The first unconditional expectation is denoted by $\bar{r}^i$ since it utilises uniform probability weighting on each possibility of t

$$\bar{r}^i = \frac{1}{3}\hat{r}^i(1) + \frac{1}{3}\hat{r}^i(2) + \frac{1}{3}\hat{r}^i(3)$$

The second unconditional expectation utilises the spectator specific probability weighting of potential outcomes so is denoted by $\tilde{r}^i$. Here I introduce another vector of distribution beliefs: $\tau_i = (\tau_i^0, \tau_i^1, \tau_i^2, \tau_i^3)$ where τ_i^w measures i 's beliefs about the probability of observing a trustor transfer of w .

$$\tilde{r}^i = \tau_i^1 \hat{r}^i(1) + \tau_i^2 \hat{r}^i(2) + \tau_i^3 \hat{r}^i(3)$$

Let N_i be a binary variable for each spectator which ascertains whether they are a Northerner. Further define the binary variable I_i to indicate if the spectator guesses the trustworthiness of an in-group trustee and O_i indicates that this is an out-group. The baseline is then that the spectator does not know the identity of the trustee.

4.2 Controls

The models below include the control vector Z_i which may explain a large proportion of the variation I find. I now outline the variables that I collect and intend to use sequentially:

- **Demographics:** Sex, number of children, age.
- **Economics:** Education, income, profession.

4.3 Hypothesis 1: Northerners have lower levels of general trust than Southerners

To test whether Northerners have lower levels of trust than Southerners, I use data from the baseline treatment condition. I can first use a chi-squared test to test the differences between the belief distributions. Following this, I can adopt the following linear regression model where Y_i is the dependent variable of choice for the specific testing tier and ϵ_i is the stochastic error.

$$Y_i = \beta_0 + \beta_1 N_i + Z_i + \epsilon_i \quad Y_i \in \{\hat{r}^i(t), \bar{r}^i, \tilde{r}^i\}, t \in \{1, 2, 3\}$$

Thus to test this hypothesis, I test to reject the null hypothesis that $\beta_1 = 0$. I use the WVS to suggest that Northerners trust approximately 8% less than Southerners. Using this number, I simulate data to propose that a sample of 85 is sufficient to have 90% power to find this result in the first tier of tests for all 3 possible distributions. With higher samples, the effect on sufficient sample size depends on the assumption of the increase in variance that comes from a wider domain of options. A sample size of 100 should suffice as long as the variance within each individual is not too large.

4.4 Hypothesis 2: Northerners and Southerners exhibit in-group bias

To test whether such differences are driven by in-group or out-group trust levels, I compute the following model using data from the treatments I and O .

$$Y_i = \beta_0 + \beta_1 N_i + \beta_2 O_i + \beta_3 N_i \times O_i + Z_i + \varepsilon_i$$

Here I have a prior that while Northerners and Southerners trust their in-group equally. This trust is more than an out-group individual. To test for in-group bias, I test whether $\beta_2 = 0$. If Northerners trust their out-group less than Southerners do then it may be that this is reflected in a negative β_1 and/or β_3 is positive.

In estimating this difference-in-difference effect hence, I need a larger sample. Using conservative estimates of effect sizes from Dharami et al. (2024) in the Hindu group, I half their effect size and find a 50% DiD effect in their primed treatment category for a baseline 18% in-group/out-group difference. Using a 9% in-group/out-group difference, I am powered at 90% to identify the 50% DiD effect with 150 observations per treatment cell.

4.5 Hypothesis 3: Demographic and economic controls are not able to close the trust gap

This hypothesis is tested by including both demographic and economic controls into my vector Z_i sequentially and testing whether the coefficients of interest above are affected by such.

4.6 Budget

The following outlines the budget for the pilot study which outlines a total budget of just below 35 600 NOK for the pilot in Prolific.

Description	Prescreen	Trustors	Trustees	Spectators
Base Payment	0.1	1	1	1
Fee	0.033	0.33	0.33	0.33
Bonus	0	5	5	0.2
Bonus Fee	0	1	1	0.04
Net Charges	0.133	7.33	7.33	1.57
Sample size	600	33	33	900
Total Net Charges	79.8	241.9	241.89	1413
Total Gross Charges	99.75	302.4	302.3625	1766.25
Exchange Rate	14.2	14.2	14.2	14.2
Total Gross Charges (NOK)	1416.45	4293.55	4293.55	25080.75

References

- Berg, J., Dickhaut, J. & McCabe, K. (1995), ‘Trust, reciprocity, and social history’, *Games and economic behavior* **10**(1), 122–142.
- Dhami, S., Wei, M. & Mamidi, P. (2024), ‘Religious identity, trust, reciprocity, and prosociality: Theory and evidence’, *Journal of Development Economics* **166**, 103192.