SPITE, FEAR AND INTERGROUP CONFLICT
Evidence from Muslims and Christians in Nigeria

Pre-Analysis Plan

Miguel Ortiz

Abstract

Despite the high societal costs generated by conflict between social identity groups, we still have little clarity of what the micro-foundations of these conflicts are. Understanding and disentangling the deep drivers of conflict is important because they determine which policies are effective. In this paper, I first ask to what extent is intergroup conflict driven by spite vs fear. To answer this, I use a lab-in-the-field to estimates preferences and beliefs, and understand to what extent each one prevents cooperation between groups. Then, I ask how popular local policies tackle these two channels to increase cooperation. To answer this, I evaluate the effects of radio drama with a message to increase intergroup cooperation, and estimate how the treatment changes the parameters I estimate in the lab, and other real behavior outcomes.

1 Introduction

Despite the high societal costs generated by conflict between social identity groups in the developing world, we still have little clarity of what the microfoundations of these conflicts are. Expressions of conflict like group segregation or communities arming themselves could have multiple drivers: they could be driven by the spiteful preference to harm the other, or by the fear of being harmed and need of protection. Understanding and disentangling the deep drivers is important because they determine what the best policy against conflict is. In the case of segregation, policies of integration could generate violence if segregation is driven by spite, or they could create trust if driven by fear. Disarming only one group could lead to peace, if the weapon acquisition was driven by fear, or it could lead to violence if it was driven by spite. Despite its importance for policy, drivers are hard to operationalize in a simple theory and disentangle empirically. This is what I do in this paper, through a framework of spite vs fear.

I pose two main questions. (1) To what extent is intergroup conflict driven by spite vs fear? Where spite is a preference for harming the outgroup, and fear is a belief about the outgroup's spite towards the ingroup. Here, I further ask two sub-questions: (1.1) If conflict is driven by fear, are beliefs accurate or misperceptions; (1.2) If driven by spite, is reciprocity shaping these preferences. After analyzing the drivers, I connect them to policy by asking: (2) How are popular local policies effective (if at all)? I focus on radio dramas to promote intergroup cooperation, a policy used all around sub-Saharan Africa, and analyze which drivers is it tackling to promote cooperation.

This project takes place in Jos, Nigeria, a state capital situated in the region of the country where the Muslim North and Christian South meet. Historically, the city had been populated with Christians and Muslims living in harmony. With democratization in the 90s, competition for power between groups increased tensions, and in 2001 a spontaneous outbreak of violence perpetrated by regular citizens occurred, leaving almost 1000 deaths. Similar crises happened in 2004, 2008 and 2011. These broke ties and trust between religious groups and set in motion a process of segregation in all dimensions. Nowadays there is virtually no contact between groups, and high levels of mistrust and animosity. Religion is the key political cleavage, and politicians fuel negative narratives on the outgroup for political gain.

To answer the first question, I will run a lab-in-the-field to estimate the parameters of a model, and determine to what extent each one drives non-cooperation. For the second question, I will run an RCT of a radio drama promoting cooperation and estimate its effects on the parameters of the model. I aim around 800 people participating in this project, but this number may vary a little depending on how slow or quickly interviews go (given that the contract is for amount of days, not of interviews).

2 Overview of experimental design

	C	NC
C	10 , 10	0 , 9
NC	9 , 0	5 , 5

I first model intergroup conflict as a coordination game, like the one above. In it, cooperation between members of different groups is an equilibrium, and the highest payoff for each. However, players may prefer to not cooperate if they feel enough spite for the outgroup: if a player prefers to sacrifice 1 payoff unit to take away 9 from the outgroup player, NC would be her dominant strategy. But lack of cooperation may stem also from fear. For a player that is not spiteful, but believes the outgroup player to be spiteful and therefore non-cooperative, her unique best response will be to not cooperate. To what extent is non-cooperation driven by preferences vs beliefs is the empirical question I tackle here.

I measure cooperation by having Christians and Muslims play coordination games between them. To understand what drives non-cooperation, I elicit participants' social preferences and their beliefs about others' social preferences. To estimate social preferences, I use a BDM mechanism to price out the social preference parameter of each participant: I elicit their willingness to pay to decrease or increase in a fix amount the payoff of their match from the outgroup. To elicit beliefs about others social preferences, I ask participants, in an incentivized manner, to guess the outgroup's willingness to pay to decrease or increase in a fix amount the payoff of someone from the ingroup. I use this information to modify the empirical distribution I estimate of the outgroup's social preferences, and get a subjective distribution for each participant. Using these elements (and a measure of risk aversion) I calibrate the model and test how well it predicts the decision to cooperate in the game. Then, I perform counterfactual analysis to understand what percentage of non-cooperation was driven by spite and fear. In addition, I can replace people's subjective distribution with the actual distribution of the outgroup's social preferences to see how much non-cooperation is driven by inaccurate beliefs. Lastly, I will test if preferences are endogenous to beliefs or not, meaning that preferences are reciprocal and open the change if beliefs are corrected.

To answer the second questions, I will run an RCT where I randomly give people access to a radio drama that promotes intergroup cooperation, and see how this affects social preferences, beliefs and cooperation. This policy is important to study because of its popularity around Africa and its alleged benefits: (i) fiction allows to touch upon sensitive topics of conflict; (ii) stories increase attention and retention of the message; (iii) it can be implemented where intergroup contact policies cannot. I partner with the radio production company hired by Search for Common Ground to do their shows, and we develop a new show that speaks directly to the mechanisms of interest: the story is about two communities that, driven by spite and fear, end up in an inefficient equilibrium. The show is recorded in Hausa and it has 24 episodes, each 10-15 minutes long. In order to target the treatment, the show will not be broadcasted, instead participants will receive the episodes periodically through WhatsApp, using one-

use links to avoid spillovers. They will listen to an episode four days a week, during six weeks, and take weekly incentivized quizzes on the content to check that they are listening. The control group will listen to a radio drama with a message about health. At the end of the interventions, participants return to the lab, so that I can measure the effects on cooperation and the different parameters (plus some real-life outcomes, see below).

The main concern might be demand effects in the lab. To address this, I take various strategies. In the lab I never explicitly name religion, but implicitly suggest religious membership of the match with a mechanism that proved effective in the pilot. (ii) Participation is anonymous, no names are collected, and receivers are informed about an extra payment affected by many things, so that they can't infer their match's decisions. (iii) I include a module on social desirability developed by psychologist and used by Dhar et al (2022) that will allow me to control for this. (iv) From the perspective of participants, the radio show and the lab experiment will be unconnected projects that are being done by the same surveying company. (v) I will include multiple real-life outcomes to complement my results: I partner with an organization working on interfaith peace, and ask participants if they want to donate to it, register in one of their workshops, and I will later observe who actually goes to it.

3 The empirical model

Consider a society where there are two groups, A and B. In it i is a member of A, and j a member of B. i has the following utility function:

$$u_i(x_i, x_j) = x_i + \beta_i x_j$$

Where x_k is the payoff of agent $k \in \{i, j\}$, and $\beta_i \in [-1, 1]$ is an individual parameter of social preferences. Individual i is altruistic towards j if $\beta_i > 0$, or spiteful if $\beta_i < 0$.

Member from both groups interact between them in a coordination game like the one below.

	C	N
C	1.00 , 1.00	0.50 , 0.90
N	0.90 , 0.50	0.75 , 0.75

Without social preferences, $\beta_i, \beta_j = 0$, there are two equilibria: (C, C) and (N, N), with (C, C) being the Pareto efficient one (and socially efficient, by any definition). However, with social preferences, the equilibria may change. If i is spiteful enough ($\beta_i < \bar{\beta}$), then N is a dominant strategy. On the other hand, if j is spiteful enough ($\beta_j < \bar{\beta}$), then N is i's best response, *regardless* β_i. So beliefs on β_j matter.

Now let's look at a more general social preferences model. Where there are different social preferences for the ingroup and the ougroup: β_{iA}, β_{iB}. And open social preferences following (Levine 1999), with β_{iG}, for $G \in \{A, B\}$, being the following:

$$\beta_{iG} = \frac{\alpha_{iG} + \lambda_G \tilde{\beta}_{iG}}{1 + \lambda_G}$$

Where still $\beta_{iG} \in [-1, 1]$. $\alpha_{iG} \in [-1, 1]$ is the base social preferences. $\tilde{\beta}_{iG}$ is equal to i's belief on $\mathrm{E}[\beta_j | j \in G]$. $\lambda_G \in [0, 1]$ is a reciprocity parameter that adjust the base social preferences, which means that people have higher social preferences with those who they believe to higher social preferences for them. And now the utility function looks in the following way:

$$u_i(x_i, x_j) = x_i + \left(\frac{\alpha_{iA} + \lambda_A \tilde{\beta}_{iA}}{1 + \lambda_A} \right) x_j \cdot \mathbf{1}(j \in A) + \left(\frac{\alpha_{iB} + \lambda_B \tilde{\beta}_{iB}}{1 + \lambda_B} \right) x_j \cdot \mathbf{1}(j \in B)$$

And the expected utility function by which a person decides its strategy in the game is the following:

$$\begin{aligned} W(s_i) = &\left\{ \tilde{\mathrm{P}}_{iA}(s_j{=}\mathrm{C}) \cdot u_i(x_i, x_j | s_i, \mathrm{C})^{\gamma_i} \; + \; \tilde{\mathrm{P}}_{iA}(s_j{=}\mathrm{N}) \cdot u_i(x_i, x_j | s_i, \mathrm{N})^{\gamma_i} \right\} \cdot \mathbf{1}(j \in A) \\ &+ \left\{ \tilde{\mathrm{P}}_{iB}(s_j{=}\mathrm{C}) \cdot u_i(x_i, x_j | s_i, \mathrm{C})^{\gamma_i} \; + \; \tilde{\mathrm{P}}_{iB}(s_j{=}\mathrm{N}) \cdot u_i(x_i, x_j | s_i, \mathrm{N})^{\gamma_i} \right\} \cdot \mathbf{1}(j \in B) \; + \; \varepsilon_{is} \end{aligned}$$

Where $\tilde{\mathrm{P}}_{iG}(s_j = \mathrm{N})$ is i's subjective probability of $s_j = \mathrm{N}$, when $j \in G$. γ is a parameter of risk aversion. And ε_{is} is an error term with extreme value distribution. This is the utility function to estimate. The objective is to recover the following parameters: α_{iG}, $\tilde{\beta}_{iG}$, λ_G, $\tilde{\mathrm{P}}_{iG}(s_j{=}\mathrm{N})$, and γ_i.

4 Estimating the parameters

One first (restrictive) approach to estimate the parameter is the following. First, I get the compound social preferences, β_{iG}, with the revealed willingness to pay to give or take a fix amount of money from their match from the ingroup and outgroup. Specifically, every money allocation decision reveals a willingness to pay or not to pay that bounds the parameter in the following way. Let Option 1 offer payments (x_i^1, x_j^1) and Option 2 offer payments (x_i^2, x_j^2). It will always be the case that $x_i^1 > x_i^2$, so that $x_i^1 - x_i^2$ is the price to pay to either help or hurt their match. Define $\bar{\beta} = \frac{x_i^1 - x_i^2}{x_j^2 < x_j^1}$ If the decision presented is altruistic, then $x_j^2 < x_j^1$, and picking Option 2 would mean that $\beta_{iG} > \bar{\beta}$, and picking Option 1 would mean that $\beta_{iG} < \bar{\beta}$. If the decision presented is spiteful, then $x_j^2 < x_j^1$, and picking Option 2 would mean that $\beta_{iG} < \bar{\beta}$, and picking Option 1 would mean that $\beta_{iG} > \bar{\beta}$. In the end, I bound the parameter in one of the following intervals: $\beta_{iG} \in \{(-1, -0.9), (-0.9, -0.8), ..., (0.8, 0.9), (0.9, 1)\}$. To have a value for β_{iG} to use in the model, I will simply assign the parameter to be the arithmetic mean of the interval it is in, i.e., if $\beta_{iG} \in (0.3, 0.4)$ then $\beta_{iG} = 0.35$.

I use the same logic to estimate $\tilde{\beta}_{iG}$ for each individual. If the participant guesses that, when playing with them and offered an altruistic decision, her match picked Option 2, then $\tilde{\beta}_{iG} > \bar{\beta}$, (and $\tilde{\beta}_{iG} < \bar{\beta}$ otherwise). And if the participant guesses that, when playing with them and offered a spiteful decision, her match picked Option 2, then $\tilde{\beta}_{iG} < \bar{\beta}$, (and $\tilde{\beta}_{iG} > \bar{\beta}$ otherwise).

To estimate $\tilde{\mathrm{P}}_{iG}(s_j = \mathrm{N})$, recall that there exits a $\bar{\beta}$ such that $\beta_j > \bar{\beta}$ means that N is a dominant strategy. This means that $\tilde{\mathrm{P}}_{iG}(s_j = \mathrm{N}) = \tilde{\mathrm{P}}_{iG}(\beta_j > \bar{\beta})$ (this equality holds under the assumption

that players don't have second order beliefs in the game, which was validated in the pilot). The first and preferred approach will be to elicit this probability directly, by asking the participant to guess how many people out of a group of 10 members of the outgroup (or ingroup) picked Option 2 in the money allocation decision that tested the bound that corresponded to the $\bar{\beta}$ of the game. This question, unfortunately, wasn't tested in the pilot of the project, so in case something unexpected goes wrong, I propose an alternative approach to calculate this probability. In this one, I will start with the empirical distribution of social preferences of each group that I get with people's responses. Then, for each person, I set $\tilde{\beta}_{iG}$ to be the new mean of the distribution, shifting the whole distribution and keeping it's shape. With this new distribution unique to each person, I can calculate numerically $\tilde{\mathrm{P}}_{iG}(\beta_j > \bar{\beta})$. One extra thing that could be done is to use the question of "how many people..." as a measure of dispersion of the distribution, in order to not only change the mean but also the variance of the empirical distribution.

With these three parameters, the one that are at the individual level, I can then proceed to structurally estimate the rest of the parameters of the expected utility function. Two additional parameters will be estimated, ψ and ϕ (see below), and one will be adjusted (or redefined, if you will), β_{iG}.

First, the parameter of risk aversion is defined in the following way. A base level of risk aversion,γ, which is then adjusted by the answers on the questions of risk aversion in the survey that placed people in one of five groups, such that $\mathrm{RA}_i \in \{1, ..., 5\}$.

$$\gamma_i = \gamma + \phi \mathrm{RA}_i$$

The original β_{iG} will now be adjusted to get a new β_{iG}, using the old estimated parameter, now called β_{iG}^{old}, in the following way:

$$\beta_{iG} = \beta_{iG}^{\text{old}} + \psi \mathrm{SDS}_i$$

$$\beta_{iG}^{\text{old}} = \frac{\alpha_{iG} + \lambda_G \tilde{\beta}_{iG}}{1 + \lambda_G}$$

The reason to do this is to adjust the social preference parameter by experimenter demand effects. The original social preferences parameter is estimated using the money allocation decisions. However, the game decision, despite being taken with the same social preference parameter, it is subject to less experimenter demand effects. The reason for this is that in the game, the undesirable action (not cooperating) could be taken for a good reason (fear), so there is plausible deniability. Instead, in the money allocation decisions, the undesirable action, being spiteful, has no plausible deniability. To address this, I adjust the social preference parameter by the social desirability. To do this I ask a survey module form psicology that calculates a social desirability score (SDS), which measure the tendency of people to give social desirable answers to others. Each person receives a SDS and I use this information to adjust the social preference parameter to a more accurate estimate that controls for this bias.

The utility of i for the payoff pair (x_i, x_j) is:

$$u_i(x_i, x_j) = x_i + \beta_{iA} \cdot x_j \cdot \mathbf{1}(j \in A) + \beta_{iB} \cdot x_j \cdot \mathbf{1}(j \in B)$$

The expected utility of i for picking strategy s_i in the game is:

$$W(s_i) = \left\{\tilde{\mathrm{P}}_{iA}(s_j{=}\mathrm{C}) \cdot u_i(x_i, x_j|s_i, \mathrm{C})^{\gamma_i} \; + \; \tilde{\mathrm{P}}_{iA}(s_j{=}\mathrm{N}) \cdot u_i(x_i, x_j|s_i, \mathrm{N})^{\gamma_i}\right\} \cdot \mathbf{1}(j \in A)$$
$$+ \left\{\tilde{\mathrm{P}}_{iB}(s_j{=}\mathrm{C}) \cdot u_i(x_i, x_j|s_i, \mathrm{C})^{\gamma_i} \; + \; \tilde{\mathrm{P}}_{iB}(s_j{=}\mathrm{N}) \cdot u_i(x_i, x_j|s_i, \mathrm{N})^{\gamma_i}\right\} \cdot \mathbf{1}(j \in B) \; + \; \varepsilon_{is}$$

To estimate the parameters I build the following maximum likelihood function. Here, W^s_{ig} is the expected utility of picking strategy s_i for person i in the game g. And $d_{ig}{=}1$ if i decides to not cooperate.

$$\Lambda_{ig} = \frac{\exp(W^N_{ig} - W^C_{ig})}{1 + \exp(W^N_{ig} - W^C_{ig})}$$

$$L(\,\cdot\,) = \prod_{i=1}^{I}\prod_{g=1}^{G} \Lambda_{ig}^{d_{ig}} \, (1 - \Lambda_{ig})^{1-d_{ig}}$$

Some other things to note. λ_G can also be estimated using only the money allocation decision, and this may be preferable, as way to separately test if is different from zero. Also, note because given that I already have calculated β^{old}_{iG} without using the λ_G, I don't really need λ_G to estimate the other parameters.

5 Conterfactual analysis

Once the model is estimated, the following tests and counterfactuals can be done to answer some of the questions established initially.

To what extent is non-cooperation driven by spite vs fear? We can answer this question by looking at what percentage of $s_i = N$ change when $\tilde{\mathrm{P}}_{iG}(s_j{=}\mathrm{N}) = 0$. This will give us the percentage of non-cooperation that could be purely driven by spite. To calculate the percentage that could be purely driven by fear, we can set all $\beta_i = 0$. If the two percentage add up to something greater than 100, then the extra will be the percentage of non-cooperation that is driven by both spite and fear. This number will also speak of how tightly correlated spite and fear are, which will also be tested with a regression.

Is non-cooperation out of fear driven by misperceptions? We can answer this question by looking at what percentage of $s_i = N$ change when $\tilde{\mathrm{P}}_{iG}(s_j{=}\mathrm{N}) = \mathrm{P}_G(s_j{=}\mathrm{N})$. To define $\mathrm{P}_G(s_j{=}\mathrm{N})$, the real probability of non-cooperation, recall that if $\beta_j < \bar{\beta}$, N is a dominant strategy. So, $\mathrm{P}_{iG}(s_j{=}\mathrm{N}) = \mathrm{P}_{iG}(\beta_j{<}\bar{\beta})$. The assumption for this equality to hold is that higher order beliefs don't matter, which was validated in the pilot. We get $\mathrm{P}_{iG}(\beta_j{<}\bar{\beta})$ from the empirical distribution of social preferences that is estimated in the experiment.

Is non-cooperation out of spite determined by reciprocity? We can answer this question by testing if λ_G is statistically different from 0. If reciprocity matters, I will look at what happens when beliefs about others' social preferences are accurate, that is, if people were accurately reciprocal. To look at accurate beliefs, we use the empirical distribution of social preferences calculated with the experiment, and select the mean. I can also look at what happens if there is no reciprocity, which means what happens when $\lambda = 0$.

6 Regression analysis

The estimating equation to analyze the effects of the radio drama will be the following:

$$y_{it} = \alpha + \beta\big(T_i \times \textit{Post}_t\big) + \psi_i + \phi_t + \varepsilon_{it}$$

Controling for experimenter demand effects with the SDS, the estimation equation looks as follows

$$y_{it} = \alpha + \beta\big(T_i \times \textit{Post}_t \times \textit{SDS}_i\big) + \psi_i + \phi_t + \varepsilon_{it}$$

The outcomes to analyze the effects of the radio show will be: social preferences towards the outgroup, beliefs about social preferences the outgroup has towards the ingroup, cooperation as measured by the coordination game, support for policies of integration, donations to an NGO working on conflict, registration to one of their workshops, and participation in one of their workshops. I will also look at effects on these outcome variables but with the ingroup. Also, the data could be all pulled together and add an dummy for outgroup ($Outgroup_o = 1$) and its respective interactions, in the following way:

$$y_{ito} = \alpha + \beta\big(T_i \times \textit{Post}_t \times \textit{Ougroup}_o\big) + \delta \cdot \textit{Outgroup}_o + \psi_i + \phi_t + \varepsilon_{it}$$

Another specification will include fixed effect by neighborhood of the city.

To check for the correlation between preferences and beliefs, I will run the following regression:

$$\textit{Beliefs}_{it} = \alpha + \beta \textit{SocPref}_{it} + \psi_i + \phi_t + \varepsilon_{it}$$

Another element to check is if answers in the lab correlate with policy attitudes and real life behavior. I'll check if having more fear or spite towards the outgroup correlates with less donations to NGO working on integration, less registration on one of their workshops and less attendance to their workshops. I ask questions about policy where I ask people if they agree that the following reasons is a downside of a particular policy that aims to promote integration. For each policy proposed, people are asked about two possible downsides they might agree with, in which one related to fear and the other to spite. The objective is to see if people that show more spite in the lab agree more with the spiteful downside of the policy, and people that show more fear in the lab agree more with the fearful downside of the policy. The outcome variable could be the reason of each policy, or an index of spiteful reasons and an index of fearful reasons. The regression is the following:

$$y_i = \alpha + \beta \textit{SocPref}_i + \gamma \textit{Beliefs}_i + \varepsilon_i$$

An alternative specification will classify people as spiteful or fearful if they don't cooperate in the game out of spite or out of fear. Using these two dummy variables, the following regression can be run:

$$y_i = \alpha + \beta \textit{Spiteful}_i + \gamma \textit{Fearful}_i + \varepsilon_i$$

7 Heterogeneity

The main heterogeneity that will be analyzed will be by religion. The fieldwork and focus group from the exploratory stage led me to state the following hypothesis: Christians will be more spiteful to Muslims than Muslims to Christians. Christians will be believe Muslims to be more spiteful towards them than the level of spite Muslims believe Christians to have towards them. Christians will cooperate less with Muslims in the coordination game than Muslim with Christians. Because of this, Christians are also expected to show stronger effects from the radio show. Although this will require a deeper explanation, my fieldwork leads me to believe that this imbalance is mainly explain by three factors. The first one is that historically the inter-faith crisis have been seen as started by Muslims. Second, Christians tend to relate Muslims to violence because the most violent part of the country is the North, the Muslim part, and because the biggest terrorist groups in the country are Muslims, like Boko Haram. Third, Christians in this context believe themselves to be the rightful owners of that land because they were the first indigenous of it, so they see the Muslims as invaders because they arrive many years later (despite being around for more than 100 years) and this has created resentment in a context of scarce resource of land (where Christians need it for agriculture and Muslims for their pastoral activities).

8 Data cleaning

Bad quality responses can affect the estimation, so I will clean for them in the following ways.

All participants will answer at the beginning of the money allocation decisions these four questions with each match they have: (A0) is the question that offers the possibility of being altruistic for free, that is, increasing in 500 naira the payoff of the match without decreasing its own; (A50) is the question that offers the possibility of being altruistic by paying 50 naira, that is, increasing in 500 naira the payoff of the match by decreasing in 50 naira its own; (S0) is the question that offers the possibility of being spiteful for free, that is, decreasing in 500 naira the payoff of the match without decreasing its own; (S50) is the question that offers the possibility of being spiteful by paying 50 naira, that is, decreasing in 500 naira the payoff of the match by decreasing in 50 naira its own. In each question they can pick option 1 or option 2, and option is always about reducing your payoff to be altruistic/spiteful with your match. There are 16 possible combinations of answers that I classify in the following way:

.

Rational selfish

A0: 1 / A50: 1 / S0: 1 / S50: 1

.

Rational altruistic

A0: 2 / A50: 1 / S0: 1 / S50: 1

A0: 2 / A50: 2 / S0: 1 / S50: 1

.

Rational spiteful

A0: 1 / A50: 1 / S0: 2 / S50: 1

A0: 1 / A50: 1 / S0: 2 / S50: 2

.

Leans altruistic

A0: 2 / A50: 2 / S0: 1 / S50: 2

A0: 2 / A50: 2 / S0: 2 / S50: 1

A0: 1 / A50: 2 / S0: 1 / S50: 1

A0: 1 / A50: 2 / S0: 2 / S50: 1

.

Leans spiteful

A0: 2 / A50: 1 / S0: 2 / S50: 2

A0: 1 / A50: 2 / S0: 2 / S50: 2

A0: 1 / A50: 1 / S0: 1 / S50: 2

A0: 2 / A50: 1 / S0: 1 / S50: 2

.

No coherence

A0: 2 / A50: 2 / S0: 2 / S50: 2

A0: 2 / A50: 1 / S0: 2 / S50: 1

A0: 1 / A50: 2 / S0: 1 / S50: 2

.

The first version of cleaning will be to drop individuals which answers combination is classified as "No coherence". The second version of cleaning, that will be used in case there are still too many bad quality answers affecting the estimation, will be to drop individuals with answers combination classified as "Leaning altruistic" or "Leaning spiteful".

Another check of quality responses focus on the answers of the game. The sign of bad quality answer is people that cooperate when the price of cooperating is high, but then don't cooperate when the price of cooperating is low. This observations can be dropped in the cleaning process.

To do further checks on the quality of responses I include some questions that check that preferences and beliefs are constant throughout the survey. For preferences I ask people if they would be willing to give up 100 to decrease in 500 the payoff of their match. They have already revealed this preferences in the money allocation decisions, so they should be coherent. I also ask if they would be willing to give up 300 to decrease in 500 the payoff of their match. For beliefs I ask people how many (out of 10) people of the group in question do they think were willing to give up 100 to decrease in 500 the their payoffs. I already have the median of the distribution of beliefs distribution, so to be coherent the median should indicate if they will pick a number above 5 or below 6. I also ask the case of give up 300 for beliefs.

Answers that are not coherent in preferences or beliefs throughout the survey can be dropped.

Related to the past quality check, one last one will be to see the number of people that participants guess that are willing to sacrifice 100 to decrease in 500 their payoff is greater or equal than the number of people that they guess that are willing to sacrifice 300 to decrease in 500 their payoffs.