1 Experimental Design

The main objective of the experiment is to measure the intensity of outcome-based and responsibility-based preferences for the charity's well-being. By relying on the sharp revealed preference test that I propose in the theoretical framework, I can disentangle responsibility-based from outcome-based preferences. Moreover, based on Proposition **??**, varying the unit reparation cost allows me to measure the intensity of linear outcome-based and responsibility-based preferences. This section discusses in detail first the design elements that map the comparative statics exercise from the sharp revealed preference test proposed in Proposition **??** into an experimental setting. After that, it discusses the design elements that are needed to measure the relative strength of linear responsibility-based and outcome-based preferences. Finally, it discusses the implementation of the experiment.

1.1 First Stage: Generation of Negative Externalities

In order to ensure that participants faced a passive other as opposed to another participant who could respond to their decisions, I opted for a charity in the role of the passive other. Therefore, participants were informed that their decisions had consequences for themselves as well as for the donation made to another charity. To maximise the probability that the participant cares about the donation made the charity, they were given a charity selection task before receiving any instructions about the experiment. In this charity selection task, participant could select the recipient of a donation among a list of three charities. Because participants were Spanish residents, I focused on charities that are operating in Spain, and hence the participants are likely to be familiar with them. The donation made to these charities finances the provision of certain charity items. Table **??** presents the list of charities, the number of charity items financed by the donation, and the country in which the charity is operative.

Table 1: List of charitable organizations that could receive a donation of €7 in a charity selection task. The selected charity acted in the role of the passive other throughout the experiment. In each session, a subject was selected at random for whom the donation was carried out.

Charity	**Social Issue**	**Charity Items Financed by € 7 Donation**	**Geographical Area**
Plant-for-the-Planet	Global Warming	Planting one tree	Peruvian Amazonas
Fundación Educo	Child Poverty	Distribution of two meals	Spain
Doctors without Borders	Inadequate Medical Care Access	Distribution of one vaccination kit (=14 vaccines) against Meningitis	Central Africa

The selected charity remains in the role of the passive other in all decision situations a participant faces in this experiment. In the first stage of the experiment, a participant faces 7 decision situations. In each of the 7 decision situations, the selected charity starts with a donation it is assigned to receive. Eventually, the participant will observe that part of the donation has been reduced. While she knows the amount of the donation that has been cancelled, she does not know its origin except that it is unrelated to the decision she makes in the first stage of the experiment. The participant is offered a monetary benefit. In case she

accepts it, as an additional consequence, the remainder of the donation is cancelled, and thus the selected charity receives no donation. If the participant rejects the monetary benefit, the part of the donation that is not already cancelled remains intact. The donation that remains intact is carried out by the experimenter if the selected decision situation is selected for payment.[1] In order to make the negative externality salient to participants, it is not framed as a cancellation of the donation but in terms of the corresponding loss in charity items.[2] Figure **??** shows a screenshot of a decision screen a participant observed in the first part of the experiment. In order to incentivize participants to earn the monetary benefit, they were informed that they could spend the monetary earnings to restore the total number of items that have been lost.[3]

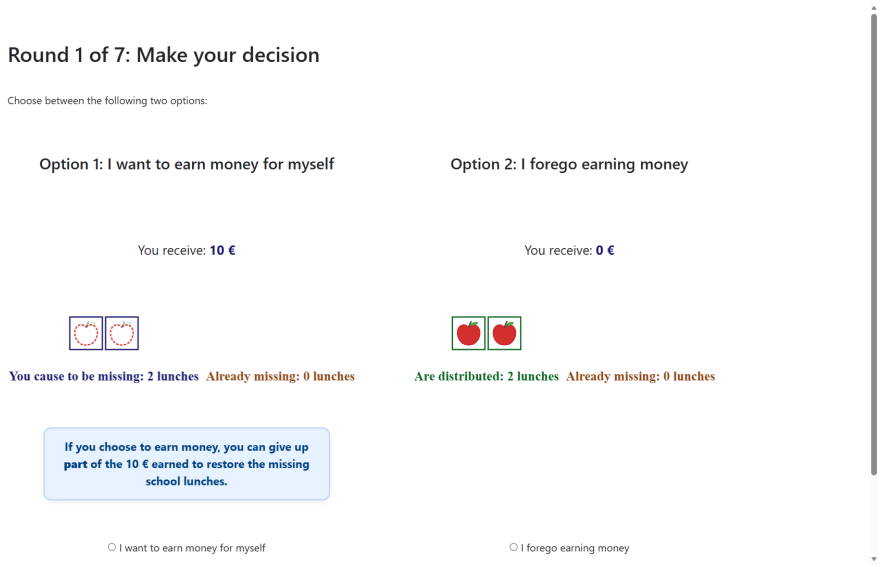


Figure 1: Decision Screen in the Externality Creation Stage

In the theoretical framework, I measure the causal history of the negative externality by the decision maker's relative share on the exact same negative externality that has been caused in the first stage. To isolate the causal history of the negative externality from the final outcomes of reparation choices in the second stage, the participant's relative share on the cancelled donation needs to vary in the experiment. This requires the participant to accept the monetary benefit in the first-stage decision situation. Then, the exact same donation cancellation can be generated in different ways. I design three pairs of decision situations in the first stage in which I fix the monetary benefit as well as the total amount of the cancelled donation but vary the participant's relative share. In two pairs the donation cancellation has a single source. In decision situations **`FullParticipant7`** and **`FullBackground7`** the same cancelled donation of €7 is either fully participant-caused or fully exogenous. In the latter decision situation, earning a monetary benefit did not cause any additional cancellation in the donation. In **`FullParticipant14`** and **`FullBackground`** the cancelled donation of €14 is either fully participant-caused or exogenous. In one pair of decision situations, the total amount of the cancelled donation is €21 and of mixed source. In **`ParticipantLow`**, the share of participant-caused

[1] See the logistics of the experiment for more details on the payment.

[2] Experimental designs represent a negative externality in different ways. For example, in the experimental design of Bartling et al. (bartling$_m$$a$$r$$k$$e$$t$$s_2$015)$t$$h$$e$$n$$e$$g$$a$$t$$i$$v$$e$$e$$x$$t$$e$$r$$n$$a$$l$$i$$t$$y$$c$$o$$r$$r$$e$$s$$p$$o$$n$$d$$s$$t$$h$$e$$r$$e$$d$$u$$c$$t$$i$$o$$n$$i$$n$$p$$a$$y$$o$$f$$f$$o$$f$$a$$p$$a$$s$$s$$i$$v$$e$$p$$a$$r$$t$$i$$c$$i$$p$$a$$n$$t.$$S$$o$$m$$e$$e$$x$$p$$e$$r$$i$$m$$e$$n$$t$$a$$l$$d$$e$$s$$i$$g$$n$$s$$r$$e$$p$$r$$e$$s$$e$$n$

[3] Participants did not know how much they had to pay to restore the lost items, and thus could make any intertemporal decisions.

harm is 1/3 and in `ParticipantHigh` it increases to 2/3.