

# Charitable Giving and Self-Signaling

Carlos Cueva\* and Roberta Dessì†

December 7, 2010

## Abstract

We provide an experimental test of the role of self-signaling in decisions to donate to charity. Our data strongly supports the theoretical prediction of a non-monotonic, hill-shaped relationship between self-confidence and prosocial behavior. Making self-image concerns more salient can more than double donations by individuals with medium self-confidence.

## 1 Introduction

Ample empirical evidence shows that people, on average, tend to behave (more) “prosocially” after observing others behaving (more) prosocially. For example, in laboratory and field experiments alike, subjects tend to donate more to charity when they observe that others have been generous than when they observe that others have donated very little<sup>1</sup>. In the absence of any information about others’ behavior, individuals also give more when they are first asked to *guess* the behavior of previous experimental participants,

---

\*University of Cambridge.

†Toulouse School of Economics (IDEI and GREMAQ).

<sup>1</sup>Several field experiments have shown that people tend to donate more when they are informed that others are also being generous. For example, Bruno Frey and Stephan Meier (2004) found that students were more likely to donate to a charitable University fund if they thought a higher proportion of students had donated in the past. Similar evidence has been found in field studies of voluntary contributions to a national park in Costa Rica (Alpizar, Carlsson, and Johansson-Stenman, 2008), public radio stations in the U.S. (Croson and Shang, 2008), maintenance of ski tracks in Sweden (Heldt, 2005) and an art gallery in New Zealand (Martin and Randal, 2008). The same tendency has also been directly observed in laboratory experiments using public goods games (Fischbacher, Gächter, and Fehr, 2001; Bardsley and Sausgruber, 2005) and dictator games (Krupka and Weber, 2009).

relative to a control condition where they just make their choices without any guesses about others being elicited<sup>2</sup>. The differences are economically important: eliciting guesses can increase the proportion of subjects behaving prosocially by as much as 38%, while observing others' prosocial behavior can increase it by as much as 47%<sup>3</sup>. A variety of explanations have been put forward for these observations, including reciprocity, conformity, and the effect of focusing attention on social norms.

In economics, the main theories that help to shed light on these findings are *signaling* theories. Hermalin (1998) developed a model in which leaders endowed with some private information on the returns to effort for a communal project may credibly signal this information, through "*sacrifice*" or *example*, thereby influencing the behavior of followers. We will refer to this as the "leader information" signaling theory, to emphasize the fact that the nature of the information held by the leader and conveyed by his action to the followers is the key to the theory's behavioral predictions. In the context of experiments with sequential decisions to contribute to charity, it has been argued that first-movers' decisions may signal private information about the quality of the charities concerned and the social value of their activities.

More recently, Bénabou and Tirole (2010) have developed a model of *self-signaling*, based on the idea that people have imperfect self-knowledge and value a "positive" self-image (e.g. prosocial, fair, honest), for hedonic or instrumental reasons. In the model, individuals with imperfect memory, who receive private, "soft" information about their "identity" (e.g. through introspection), may want to signal to their future selves by choosing certain observable, identity-relevant actions. The game here is an intra-personal one, between the current self and the future self, motivated by the fact that the signal received by the current self may be forgotten, or manipulated in a self-serving way: the future self will therefore update its beliefs based on "hard" information. This is what may give a signaling value to the decision to undertake costly, identity-relevant *observable* actions - for example, giving money or time to charity, or other kinds of prosocial behavior.

In this setting, observing others' behavior before choosing one's own action can influence decisions in two important ways. First, if signals are *correlated*, knowledge of others' actions will affect the future self's inference process: this can give rise to strategic complementarities and "conformity"<sup>4</sup>. Second, even when signals are independent, observing others'

---

<sup>2</sup>Krupka and Weber (2009).

<sup>3</sup>These figures represent the biggest differences reported by Krupka and Weber (2009).

<sup>4</sup>See Bernheim (1994) for a signaling theory of conformity where individuals care about their status, which depends on the inferences *other people* will make about their "type", based on their observable behavior.

identity-relevant actions may have an impact if it makes identity concerns more *salient*.

We refer to the Bénabou and Tirole model as the “identity” signaling theory. Clearly the “identity” theory and the “leader information” theory can both account for a positive correlation between prosocial behavior of first-movers and prosocial behavior of second-movers. We therefore focus primarily on a different implication of the identity theory. This will enable us to distinguish between the two theories and investigate an important prediction of the Bénabou and Tirole model.

The prediction may be stated succinctly as follows (a more detailed account is provided in section 2). Under plausible assumptions, identity investments will exhibit a non-monotonic, *hill-shaped* relationship with *self-confidence*. Intuitively, someone who is very confident about his identity has little need for costly self-signaling, and will not make identity-affirming investments irrespective of whether his private signal is “good” or “bad”. For lower levels of initial self-confidence, the good type (who receives the good private signal) will have an incentive to invest to distinguish himself from the bad type and raise his *ex-post* self-confidence; i.e. raise his posterior belief on the probability of being a good type. For the bad type, on the other hand, the return from investing will increase with initial self-confidence: when the prior belief on being the good type is very low, the gain from pooling with the good type is small; when the prior is relatively high, the gain from pooling is greater.

If we could elicit individuals’ prior beliefs and then observe their behavior in an appropriately designed experiment, we could test this prediction directly. However, it is difficult to elicit such beliefs at the beginning of an experiment without priming subjects and potentially influencing their behavior<sup>5</sup>. We therefore proceed as follows. In section 2, we identify the key implications of the identity theory for the relationship between observable behavior and *posterior* beliefs. This allows us to develop an experimental design in which we manipulate the salience of identity concerns so as to increase it for the treatment condition relative to a control group, without any potentially confounding effect due to *ex-ante* belief elicitation. Beliefs are elicited *ex post*.

The basic design of the experiment is quite simple. In each session, subjects are given some information about the aims and work of three very well-known charities (Red Cross, Save the Children and Amnesty Interna-

---

<sup>5</sup>There is a large empirical literature documenting the importance of “priming” effects: for some recent examples see Benjamin et al. (2010), LeBoeuf and Shafir (2010), Mazar et al. (2008).

tional). A subject is then randomly and anonymously selected to be the leader (“decision-maker”): she is given an endowment of £10, and has to decide whether to keep it all, or to give £4 to a charity of her choice among the three just described. The leader enters her choice on her computer, while the other subjects enter their answer to an unrelated question. The leader’s chosen donation amount (not the charity), i.e. 4 or 0, then appears, anonymously, on every subject’s computer screen. This sequence of moves is explained to all subjects at the beginning of the experiment. Following disclosure of the leader’s decision, subjects learn that in the second and last part of the experiment each individual will receive an endowment of £10, and will decide freely, and anonymously, how much of the endowment to give to a charity of their choice (one of the same three).

We vary this basic design in the treatment condition so as to make identity concerns more salient. Specifically, we now require the leader to stand up after making his decision, holding a card with his chosen donation amount written on it and clearly visible to the other session participants. The experimenters exit the laboratory while this is happening and an assistant not otherwise involved in the experiment just checks that indeed someone stands up and holds the right decision card, then sits down again. This modified sequence of moves is explained to all participants at the beginning of the session. In the second part of the experiment, it is made clear to all participants that decisions will be made privately. Thus manipulating the leader’s visibility should only affect the followers’ decisions, holding constant the leader’s chosen amount, by making *identity more salient*: seeing someone else stand up and “show his colors” focuses attention on the implications of one’s own choices for “the kind of person I am”.

Since the theory’s focus is on self-signaling, we adopt a double-blind protocol for the followers; this also minimizes any possible experimenter demand effects.

At the end of the experiment, we elicit subjects’ posterior beliefs. We do this by asking them to complete the Multidimensional Personality Questionnaire (MPQ). The questionnaire contains a large number of questions, intended to evaluate 11 primary trait dimensions. One of these, Social Potency, provides a good measure of self-confidence in the sense of this paper; indeed, Burks, Carpenter, Goette and Rustichini (2010) have found that subjects with a high score on the Social Potency scale make more confident judgements about their performance in an IQ test, holding actual ability constant. In our setting, which involves decisions about donations to charity rather than performance on an IQ test, we conjecture that a high score on the Social potency scale will reflect greater confidence in one’s values (belief on being a good type). To check whether this is a reasonable interpretation,

we also have a group of subjects who participated two weeks earlier in a completely unrelated experiment on cognition and social networks<sup>6</sup>: these subjects completed the MPQ at the end of the earlier experiment and are not asked to do so again in our experiment. We conjecture that the Social Potency score for these subjects will reflect confidence in their social skills, their ability to perceive social networks accurately and become key players within them. Since this is largely orthogonal to confidence in moral identity, if our conjecture is correct we should not find the same relationship between behavior and beliefs, as captured by the Social Potency score, for the group that took the MPQ earlier as for our main group.

Indeed, for the "visible leader" treatment we find no significant relationship between followers' donation decisions and self-confidence as measured by the Social Potency score obtained at the end of the previous experiment on social networks and cognition. In contrast, we find a significant, non-monotonic relationship between the amount donated to charity and self-confidence as measured by the ex-post Social potency score. This relationship is consistent with the prediction from a pooling equilibrium in the Bénabou and Tirole model, as discussed in section 2. Moreover, the relationship is not present in the "anonymous leader" control, where identity concerns are less salient. The difference in amounts donated between the treatment and the control gives a quantitative measure of the importance of signaling in this context. This difference is highly significant, and economically important: the average amount donated is more than doubled in the presence of strong identity concerns.

The remainder of the paper is organized as follows. Section 2 relates our work to the existing literature in economics, and in particular to the identity theory. Section 3 describes our experimental design and procedures. Section 4 presents our main results on the relationship between donations to charity and self-confidence. Some additional results are contained in section 5. Section 6 concludes.

## 2 Existing literature and implications

We focus here on the identity theory developed by Bénabou and Tirole (2010) to derive the implications that motivate our experimental design.

---

<sup>6</sup>See Dessí, Gallo and Goyal (2010).

## 2.1 The model

In the baseline version of their model, an individual has imperfect self-knowledge to begin with: for simplicity, he may be a “good” type,  $v_H$ , or a “bad” type,  $v_L$ , and he has a prior belief  $\rho$  which represents the probability of being the good type. The individual’s type  $v$  reflects his values (identity): for example, how prosocial (generous, altruistic, fair, etc.) he is. The prior belief  $\rho$  captures his initial self-confidence.

The individual then receives a private signal concerning his type; for simplicity, this signal is assumed to be fully informative ( $v$ ). However, the signal is “soft” information - for example, “an instinctive feeling of empathy, a temptation to cheat or a conscious self-assessment”. The individual has a momentary glimpse into his true nature, but his subsequent recall of this insight will be very imperfect and self-serving<sup>7</sup>. This is modeled by assuming that in the next period, with a given probability (denoted by  $1 - \lambda$ ), the individual will no longer recall (reliably) the signal, and will update his belief on his type using only “hard” information.

Hard information is provided by an observable action (or its absence): just after receiving his private signal, the individual has the option to undertake an identity-affirming action, whose cost  $c_0$  is (weakly) lower for the good type (e.g. buying “green” products, donating to charity). Denote by  $a_0 \in \{1, 0\}$  this investment decision. In the next period, with probability  $1 - \lambda$ , the individual no longer recalls reliably his private signal; he therefore updates his belief on his type based on the observation of  $a_0$ . This yields the following updated beliefs  $\rho^*(a_0)$ :

$$\rho^*(1) = \frac{\rho x_H}{\rho x_H + (1 - \rho)x_L}; \quad \rho^*(0) = \frac{\rho(1 - x_H)}{\rho(1 - x_H) + (1 - \rho)(1 - x_L)} \quad (1)$$

where  $x_H$  and  $x_L$  are the investment probabilities for the good type and the bad type, respectively. We will refer to the individual choosing the action  $a_0$  as the current self, and the same individual updating his belief the next period as the future self.

Bénabou and Tirole develop different versions of the demand side of the model, based on different possible reasons for people to wish to have high self-confidence *ex post* (i.e. high posterior beliefs  $\rho^*$ )<sup>8</sup>. For our purposes, what matters is their characterization of the equilibrium of the signaling game

---

<sup>7</sup>See Bénabou and Tirole (2010) for a discussion of the evidence in support of this assumption.

<sup>8</sup>The reasons they consider are: the hedonic benefits of high self-esteem; anticipatory utility; the instrumental benefits in dealing with self-control problems.

between the current self and the future self, which does not depend on the particular reason for desiring high self-confidence *ex post*. The equilibrium is described in their Proposition 1, which states:

**Proposition 1** (Bénabou and Tirole (2010)) *There exists a unique (monotonic, undominated) equilibrium, characterized by thresholds  $\rho_l$  and  $\rho_h$  with  $0 < \rho_l \leq \rho_h \leq 1$  and investment probabilities  $x_H(\rho)$  and  $x_L(\rho)$  such that:*

- (1)  $x_H(\rho) = 1$  for  $\rho < \rho_h$  and  $x_H(\rho) = 0$  for  $\rho > \rho_h$ ;
- (2)  $x_L(\rho)$  is non-decreasing on  $[0, \rho_l]$ , equal to 1 on  $[\rho_l, \rho_h)$  when  $\rho_l < \rho_h$  and equal to 0 on  $[\rho_h, 1]$ .

The intuition for this result may be summarized as follows. When initial self-confidence is sufficiently high ( $\rho > \rho_h$ ), neither type needs to invest in identity, since in the absence of investments *ex-post* confidence will still be high ( $\rho^* = \rho$ ), and could not have been increased much. When initial self-confidence falls below the critical threshold  $\rho_h$ , it becomes worthwhile for the good type to invest to try to separate from the bad type, and thereby raise his *ex-post* confidence  $\rho^*$ . What happens in equilibrium depends on the behavior of the bad type. If his cost  $c_0^L$  is sufficiently high, he will never invest (separation:  $x_H(\rho) = 1, x_L(\rho) = 0$ ). For lower values of  $c_0^L$ , he will invest with strictly positive probability. The probability is increasing in  $\rho$ , since the net gain from pooling with the good type increases with  $\rho$  (from (1)).

In the model, the individual is assumed to hold an accurate prior belief  $\rho$ . It is straightforward to verify from Proposition 1 that in this case average investment in identity, defined as  $x = \rho x_H + (1 - \rho)x_L$ , exhibits a non-monotonic, hill-shaped relationship with initial self-confidence,  $\rho$ .

## 2.2 Implications for behavior and ex-post self-confidence

We now turn to the relationship between observed behavior,  $a_0$ , and ex-post self-confidence,  $\rho^*$ . There are three cases to consider.

*Case 1* (separation). This is the simplest case: in equilibrium, the bad type never invests, while the good type invests if, and only if,  $\rho < \rho_h$ . Thus for  $\rho > \rho_h$ , there will be no investment, and both types will have posterior belief  $\rho^* = \rho$ . For  $\rho < \rho_h$ , only the good type invests: his posterior belief will be  $\rho^* = 1$ , while the bad type will have posterior belief  $\rho^* = 0$ . Ex post, investment will be associated with very high self-confidence ( $\rho^* = 1$ ). Note that in this case the relationship between investment and posterior beliefs is very different from the hill-shaped relationship between investment and prior beliefs discussed above.

*Case 2* (full pooling over some range). This is the case where, for  $\rho < \rho_h$ , the bad type invests with increasing probability up to the threshold  $\rho_l$ , and

then pools fully with the good type up to the higher threshold  $\rho_h$ . The good type always invests in the range  $\rho < \rho_h$ . Neither type invests for  $\rho > \rho_h$ . Posterior beliefs are the same for both types in the range  $\rho > \rho_h$ , and are given by  $\rho^* = \rho$ . The same is true in the range  $\rho_l \leq \rho < \rho_h$ , since there is full pooling. In the range  $\rho < \rho_l$ , since the bad type randomizes while the good type always invests,  $\rho^*(0) = 0$  and  $\rho^*(1) > \rho$ . Moreover, it can be verified that full pooling over some range occurs when  $\rho^*(1) \leq \rho_h$ .<sup>9</sup>

Ex post, investment will therefore exhibit a hill-shaped relationship with self-confidence, with the peak in the range  $\rho_l \leq \rho^* < \rho_h$ .

*Case 3* (partial pooling). In this case, the bad type invests with increasing probability over the range  $\rho < \rho_h$ , but the probability remains always strictly lower than one. The good type invests with probability one over the same range. Neither type invests for  $\rho > \rho_h$ . As in the other cases, posterior beliefs are given by  $\rho^* = \rho$  when  $\rho > \rho_h$ . In the range  $\rho < \rho_h$ , we have  $\rho^*(0) = 0$  and  $\rho^*(1) > \rho$ . Moreover, it can be verified that partial pooling occurs when  $1 > \rho^*(1) > \rho_h$ , where  $\rho^*(1)$  is a constant.<sup>10</sup>

Ex post, investment will be associated with high self-confidence ( $\rho^*(1) > \rho_h$ ).

### 2.3 Implications: summary

We can summarize the predictions that will motivate our experiment as follows:

(i) in the case of separation and in the case of partial pooling, identity investments will be significantly higher among individuals who will be very self-confident ex post.

(ii) in the case of full pooling over some range, identity investments will exhibit a non-monotonic, hill-shaped relationship with ex-post self-confidence.

## 3 Experimental Design and Procedures

The study was conducted with 135 students at the University of Cambridge. Participants were unaware of the nature of the study before entering the lab. Subjects were only identified through personal IDs which were assigned randomly as they entered the lab. Once in the lab, they were informed that the experiment studied decisions to donate to charity. They were assured that none of the experimenters would be able to link their identities with their decisions and that payments at the end of the experiment would be

<sup>9</sup>For details, see the proof of Proposition 1 in Bénabou and Tirole (2010).

<sup>10</sup>For details, see the proof of Proposition 1 in Bénabou and Tirole (2010).



made so as to preserve confidentiality<sup>11</sup>. They were presented with a choice of three charities and a short summary of their aims and operations. The charities used were *The British Red Cross*, *Amnesty International* and *Save the Children*. We chose three international well known charities to maximise the likelihood that subjects would be familiar with them.

The experiment was divided into four parts: *leader announcement*, *follower donations*, *follower guesses* and *personality questionnaire*. We used a between-subjects  $2 \times 2$  design, with the two factors being *leader visibility* and *number of leaders*. Some sessions had two leaders instead of one. This allowed us to check whether the degree of social influence was affected when the number of leaders increased from one to two. It also allowed us to evaluate the effect of conflicting announcements sent by the leaders. All corresponding procedures were explained before subjects made their decisions.

*Leader announcement*: subjects were informed that one or two subjects in the room would be randomly selected as *leaders* (to avoid priming, we referred to these subjects as “decision-makers”). The leaders would be endowed with £10 and would have to decide to either *donate £4 to a charity and keep £6*, or *donate £0 and keep £10*. The selected leaders would simply be prompted to enter their decision privately into their computer, while everyone else would be prompted to make an unrelated decision. This was done to ensure the anonymity of the leaders. Once all subjects had completed their task, the choices of the leaders would be announced via subjects’ PC screens. The initial instructions for the *visible leaders* sessions contained an additional part, specifying that the leaders would have to stand and hold up a sign with the amount donated written on it for all participants to see. We explained that the experimenters would leave the lab while this took place: specifically, once the leaders’ decisions appeared (anonymously) on the PC screens, the experimenters would walk out of the room and an assistant, otherwise unrelated to the experiment, would walk in bringing cards corresponding to the two possible decisions. Each leader would then stand, choose a card and hold it up so that the other session participants could read the amount he or she was donating to charity. They would then return the card to the assistant and go back to sit in front of their PC screen. At this point the assistant would go out, the experimenters would go back to the lab, and the experiment would continue.

*Follower donations*: in the next part of the experiment all subjects (including the leaders) received a (new) £10 endowment which they could freely and privately allocate between themselves and a charity of their choice<sup>12</sup> (to

---

<sup>11</sup>This was achieved by preparing sealed envelopes containing the appropriate cash payments, with the ID code written on the envelope.

<sup>12</sup>Participants could choose one of the three charities used in the first part of the exper-

the nearest pound). Subjects entered their choices individually and confidentially into their PCs. Note that subjects did not know about this stage in the first part of the experiment.

*Follower guesses:* after all decisions were made, subjects were asked to guess the donation of the leader and the average donation of the other followers. Subjects with at least one correct guess were rewarded with an additional £2. Again, subjects did not know about this stage in the previous part of the experiment.

*Personality questionnaire:* as the final stage of the experiment, subjects were asked to complete a brief (155-item) form of the Multidimensional Personality Questionnaire (Patrick, Curtin, and Tellegen, 2002). The MPQ is a standard personality profile test (Patrick, Curtin, and Tellegen, 2002; Tellegen, 1988; Tellegen and Waller, 1994). It consists of a series of multiple choice questions concerning 11 primary trait scales. The primary traits measured by the MPQ (and by its brief form) are: Social Potency, Wellbeing, Achievement, Social Closeness, Stress Reaction, Alienation, Aggression, Control, Harm Avoidance, Traditionalism and Absorption. Questions were ordered randomly and subjects had to answer all 155 questions, which were presented one at a time on their PC screens.

The Social Potency scale is particularly relevant to our study, as discussed in the Introduction. A high score on this scale indicates that the subject considers himself to be persuasive, that he likes to be a leader and that he feels others value his judgements. These characteristics tend to be highly correlated with self-confidence. For subjects taking the MPQ at the end of our experiment, confidence about their values and moral judgements will have been the salient dimension of self-confidence. Their score on the Social Potency scale should therefore be a good measure of the posterior belief  $\rho^*$  in the model of section 2. We also had a control group of subjects who had taken the MPQ a couple of weeks earlier, at the end of an unrelated experiment on social networks. For these subjects, confidence in their social and networking skills will have been the salient dimension of self-confidence when completing the MPQ. Since this is largely uncorrelated with confidence in moral identity, we believe that these subjects' score on the Social Potency scale will be largely irrelevant in explaining behavior in our experiment.<sup>13</sup>

Some of the other traits measured by the MPQ could also be relevant to decisions to donate to charity. In what follows, we will check whether this is the case, notably for *Social Closeness*, *Stress Reaction*, *Alienation*

---

iment. Note that the leader(s)' chosen charity, if any, was never revealed.

<sup>13</sup>Moreover, we cannot control for any events that may have led these subjects to update their beliefs during the time between the two experiments.

**Table 1:** Followers after different leader announcements; numbers in brackets are sessions.

Leader announcement	One leader		Two leaders		
	£4	£0	both £4	one £4, one £0	Both £0
Anonymous leader	24 (2)	0	7 (1)	27 (3)	0
Public leader	23 (2)	0	14 (2)	9 (1)	10 (1)
Totals	47 (4)	0	21 (3)	36 (4)	10 (1)

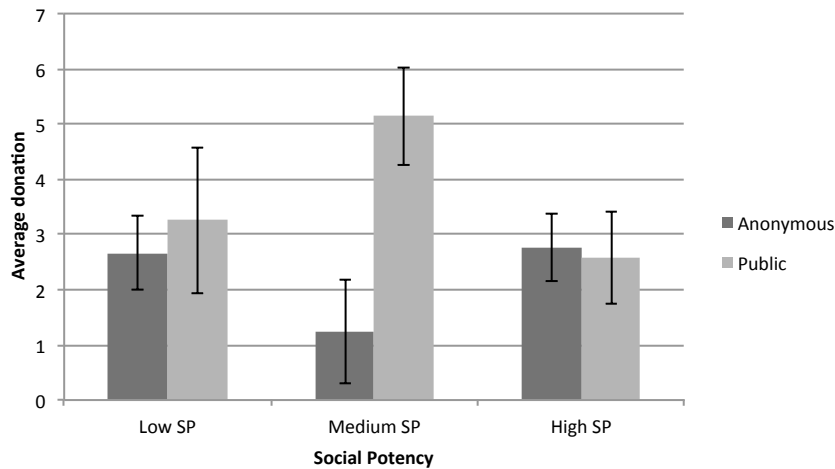
and *Traditionalism*. Subjects with a high score in Social Closeness describe themselves as sociable, warm and affectionate, valuing close relationships, and welcoming support from others. High Stress Reaction indicates that a subject is easily upset, nervous, and prone to feel guilty and vulnerable. Finally, subjects with a high score in Traditionalism tend to advocate high moral standards, condemn selfishness, endorse religion and strict child rearing, oppose permissiveness and value propriety.

## 4 Results: donations and self-confidence

A total of 12 sessions were conducted with group sizes ranging from 9 to 14 subjects. In half of the sessions the identity of the leader(s) was completely private and anonymous (henceforth “anonymous” sessions) and in the other half it was visible to the other participants (henceforth “public” sessions). Four sessions had a single leader, whereas the remaining eight had two leaders. Approximately half of our participants were women (70 out of 135), and about one third of our subjects were studying for an economics or business-related degree.

We summarize the distribution of followers across the different leader announcements in Table 1.

The Table reveals that all leaders in the single-leader sessions (both public and anonymous) chose to donate the same amount, £4, to charity. Leaders’ decisions are more varied, on the other hand, in the sessions with two leaders: some sessions had two generous leaders, some one, and one session had both leaders choosing not to give anything to charity. We therefore begin by analysing the single-leader sessions, since these enable us to investigate the relationship between follower donations and self-confidence while holding constant the number of leaders (one) and the amount donated by the leader (£4). We will then go on to our main analysis, using the whole sample; this will also enable us to investigate the impact of having two leaders, with different leader donations.



**Figure 1:** Donations of followers in single leader sessions: averages in anonymous and public treatments for subjects with low, medium and high scores in the social potency trait (SP). Error bars are standard errors of the mean.

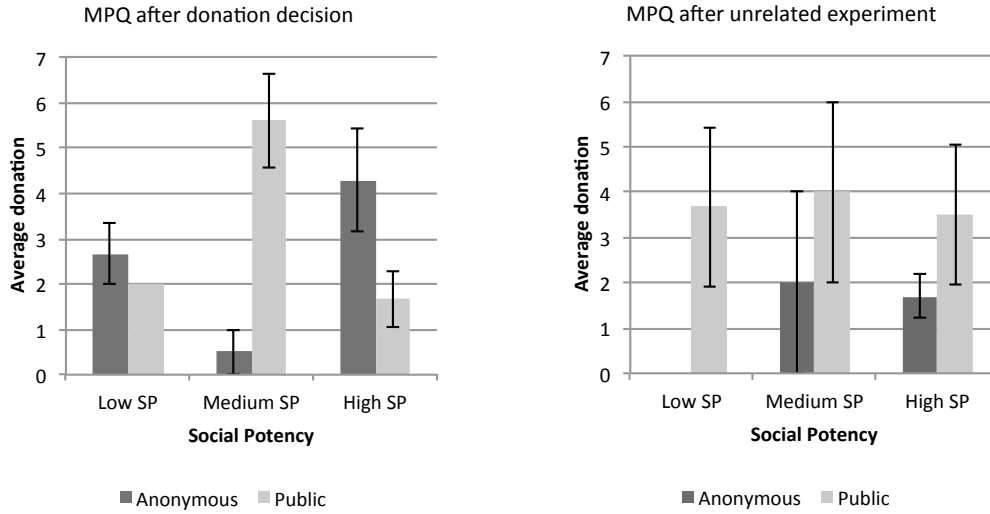
#### 4.1 Single-leader sessions

We had 4 single-leader sessions, 2 public and 2 anonymous. As just noted, the leader’s decision was the same in all four sessions (give £4 to charity). Figure 1 shows average follower donations for public and anonymous sessions. Subjects are divided into three groups depending on their score on the Social Potency scale of the MPQ: low, medium and high.<sup>14</sup>

Donations for the public treatment should reflect individuals’ identity investments (self-signaling), if any, since watching the leader stand up and “show his colors” should make moral identity salient for the followers. Indeed, Figure 1 shows that average follower donations for this treatment exhibit a clear *hill-shaped* relationship with self-confidence, as measured by the score on the Social Potency scale. The amount donated by the medium-potency group is significantly higher than the amount donated by the other two groups ( $p = 0.0300$ ).<sup>15</sup>

<sup>14</sup>The identity theory discussed in section 2 gives no guidance as to the precise values of the relevant thresholds. We therefore chose plausible values by inspection of the scatter plot for the whole sample, and use these values throughout the paper, including here where we are analysing the sub-sample of single-leader sessions.

<sup>15</sup>The difference between the medium-potency group and the high-potency group is significant, while that between the low-potency group and the medium-potency group is not, possibly owing to the small number of observations in the low-potency group.



**Figure 2:** Donations of followers in single leader sessions. *Left panel:* Social Potency scores from MPQ administered after donation decision. *Right panel:* Social Potency scores from MPQ administered after unrelated experiment. Error bars are standard errors of the mean.

This is not the case for the anonymous treatment: here identity was not made salient, and the amount donated does not differ significantly between the three potency groups. The difference between the amount donated in the public and in the anonymous treatment can be viewed as a measure of the importance of signaling motives. This difference is not significant for the low- and for the high-potency groups, but it is significant for the medium-potency group ( $p = 0.0287$ ). Signaling motives more than quadruple the amount donated by this group.

The results summarized by Figure 1 do not differentiate between subjects who took the MPQ at the end of our experiment and the control group who had taken the MPQ at the end of a previous unrelated experiment on social networks. We have argued earlier in the paper that the social potency score should be a good measure of ex-post self-confidence ( $\rho^*$ ) for subjects who took the MPQ at the end of our experiment, while it should be largely irrelevant for subjects who took the MPQ following the earlier experiment. Figure 2 gives the same information as Figure 1 separately for the two sets of subjects.

The left panel of Figure 2 shows average follower donations for subjects who took the MPQ at the end of our experiment. It exhibits a clear hill-shaped relationship between the average amount donated and the social potency score for the public treatment, with a significantly higher amount being

donated by the median-potency group ( $p = 0.0078$ ). The pattern is reversed for the anonymous treatment. This can be compared with the right panel of Figure 2, showing average follower donations for subjects who had taken the MPQ at the end of the earlier experiment. Here there is no significant difference between the three social potency groups, neither in the public nor in the anonymous treatment. While the data for the single-leader sessions should be interpreted with caution because of the small sample sizes, it is clearly consistent with our predictions. We now turn to the main analysis.

## 4.2 Whole sample

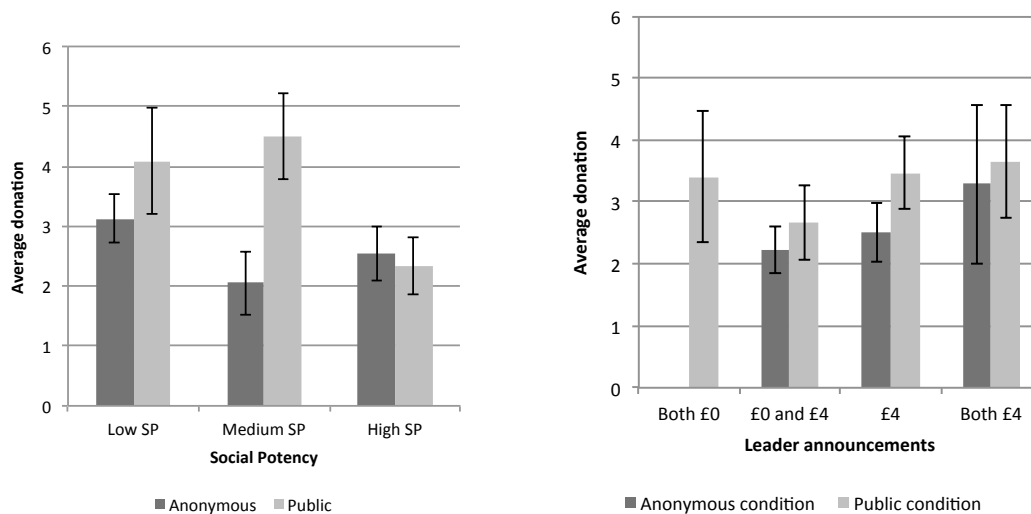
We now examine the relationship between follower donations and self-confidence for the whole sample. We therefore need to control for the possible effect of different leader announcements. Table 2 reports the results from (robust) Tobit regressions for the amount donated in the public and in the anonymous treatment. We estimate Tobit regressions because the amount donated is censored at zero and ten (the size of the endowment). Among our explanatory variables we include three dummies to control for the possible effects of different choices by leaders; the residual category is one leader giving £4. To study the relationship with self-confidence, we include a dummy for medium social potency, and one for low social potency.

The results for the whole sample confirm those found for the single-leader

**Table 2:** Follower donations and self-confidence.

Robust Tobit regression		
Dependent variable: Follower donation		
	Public condition	Anonymous condition
Medium Social Potency	2.899** (0.014)	-0.408 (0.667)
Low Social Potency	2.480* (0.053)	1.227 (0.129)
Both £4	-0.095 (0.945)	1.135 (0.499)
£4 and £0	-1.180 (0.299)	-0.324 (0.717)
Both £0	-0.092 (0.948)	-
$N$	56	58
Log pseudolikelihood	-123.79665	-121.44094

\* significant at 0.1 level  
\*\* significant at 0.05 level  
(p-values in brackets)



**Figure 3:** *Left panel:* average donation in the whole sample for subjects with low, medium and high scores in the Social Potency scale in Public and Anonymous conditions. *Right panel:* average donation after different leader announcements in Public and Anonymous conditions. Error bars are standard errors of the mean.

sessions. In the public treatment, where identity concerns are made salient, the amount donated to charity exhibits a hill-shaped relationship with self-confidence: the coefficient on the dummy for the medium potency group is positive and significant ( $p = 0.014$ ). Moreover, it is larger in magnitude than the coefficient for the low potency group, which is also positive and significant ( $p = 0.053$ ). This is not the case for the anonymous treatment, where identity is far less salient, and the coefficients for the same variables are insignificant.<sup>16</sup>

Table 2 also shows that follower donations are *not* significantly affected by the amount(s) donated by the leader(s): what seems to matter is not the leader(s)' *example* (his/their choice), but rather the fact that seeing someone stand up and show his decision openly in front of his peers makes all partic-

<sup>16</sup>We also ran the regressions in Table 2 separately for subjects who took the MPQ at the end of the experiment and those who took it following an unrelated experiment two weeks earlier. The results for ex-post self-confidence are very similar to those for the full sample: in the public condition, the dummy for median potency is positive, highly significant ( $p = 0.005$ ) and larger than the one for low potency, which is also positive ( $p = 0.063$ ). In the anonymous condition neither is significant. For subjects who took the MPQ two weeks earlier, social potency is never significant except for a positive coefficient on low social potency in the anonymous condition ( $p = 0.077$ ).

ipants think carefully about the signal that their choice will send about the kind of person they are.

Figure 3 illustrates our *key results*: the left panel shows the relationship between follower donations and the social potency score. As expected, the hill-shaped relationship is present for the public treatment and not for the anonymous treatment. It is consistent with the prediction for a pooling equilibrium (Case 2) obtained in section 2. For the median potency group, the difference between the amount donated in the public and the anonymous treatment is highly significant ( $p = 0.0059$ ) and large: the amount is more than double in the public treatment.

The right panel of Figure 3 shows the relationship between follower donations and leader donations. The pattern is broadly increasing, except for the one session with two leaders where they both chose to give nothing to charity. However, none of the differences are statistically significant.

### 4.3 Discussion

At this point it is worth considering possible alternative explanations for our results.

It could be conjectured that participants who chose to donate very little to charity tried to rationalize their behavior ex post as due to an aversion to conformity (hence unwillingness to follow the leader’s example), and that this was reflected in high scores on the Social Potency scale (“I am not selfish, I am a leader”). This could generate a negative correlation between donations and the ex-post potency score for those sessions where the leader(s) chose to donate to charity.

However, we also find the negative correlation for sessions where one or both leaders gave nothing to charity: in this case, aversion to conformity should either have no effect or increase the propensity to donate. It could not be a valid argument to rationalize low donations ex post. Yet when we ran a regression like the one in Table 2 for the public condition restricting the sample to subjects who had completed the MPQ at the end of the experiment and had participated in sessions with one or two “selfish” leaders, we found that the dummy for median potency had a large positive coefficient ( $p = 0.067$ ) while the low potency dummy had a smaller positive and insignificant coefficient.

Moreover, if the negative correlation overall had been driven by ex-post rationalization of “selfish” behavior as “aversion to conformity”, we would expect to find it also in the anonymous condition, to some degree. A Tobit regression for this condition, restricted to subjects who had completed the MPQ after the experiment and had participated in sessions with “generous”



leaders (i.e. where we might expect to find the negative correlation), yielded negative estimated coefficients for the medium and low potency dummies.

These findings cast doubt on the conjectured alternative explanation.

A different objection to our interpretation of the results might be that subjects will not have had enough time to "forget" their internal states of mind at the beginning of the experiment by the time they complete the MPQ at the end. Thus in terms of the model of section 2, questionnaire answers should reflect  $v$ , not  $\rho^*$ . However, "forgetfulness" is only one possible reason for the self to update beliefs on the basis of observable actions ( $\rho^*$ ) rather than internal states of mind ( $v$ ). There are other possible reasons, consistent with the model of section 2, that do not require memory to decay over time: for example, the presence of self-serving biases in the interpretation as well as recall of private information, documented in the social psychology literature<sup>17</sup>. Indeed, the experimental evidence from social psychology suggests that individuals quickly update beliefs about the self on the basis of their observable past behavior<sup>18</sup>. Our results are, in this respect, in line with previous experimental findings.

## 5 Additional results: personality traits, and expectations

In this section we report some additional findings. First, we investigate whether personality traits such as social closeness, traditionalism, stress reaction or alienation, measured by the MPQ, influence decisions to donate to charity. Second, we examine the data on our experimental subjects' expectations.

### 5.1 Personality traits

As noted in section 3, the MPQ asks questions concerning 11 different primary trait scales. In section 4 we have focused exclusively on the Social Potency scale, as a measure of self-confidence. We now briefly investigate the possible role of some of the other traits.

Table 3 reports the results from (robust) Tobit regressions for the amount donated to charity in the public and in the private treatment, similar to those in Table 2 above. The difference is that this time we also include among

---

<sup>17</sup>See Dunning (2001).

<sup>18</sup>See Bem (1967) for a discussion of the evidence. Bem's argument is that individuals have only limited access to their internal states of mind, and rely a great deal on observation of their past behavior to update their beliefs about the self.

**Table 3:** Follower donations and psychological traits.

Robust Tobit regression		
Dependent variable: Follower donation		
	Public condition	Anonymous condition
Medium Social Potency	3.165** (0.014)	0.316 (0.787)
Low Social Potency	2.892* (0.087)	1.016 (0.292)
Both £4	-0.222 (0.875)	1.242 (0.498)
£4 and £0	-1.133 (0.373)	-0.473 (0.630)
Both £0	0.223 (0.878)	-
Closeness	-0.022 (0.987)	0.852 (0.295)
Stress reaction	0.727 (0.565)	0.523 (0.606)
Alienation	-0.985 (0.469)	-1.808 (0.132)
Traditionalism	0.905 (0.293)	0.787 (0.394)
<i>N</i>	56	58
Log pseudolikelihood	-123.16711	-118.46376

\* significant at 0.1 level

\*\* significant at 0.05 level

(p-values in brackets)

our explanatory variables the scores on the following trait scales: Social Closeness, Stress Reaction, Alienation, Traditionalism<sup>19</sup>.

Our results show that none of the personality traits other than Social Potency have a significant impact on the amount donated to charity.

## 5.2 Expectations

After making their donation decisions, subjects in our experiment were asked to guess the leader(s)' second donation decision (followers only), and the average amount donated by followers. They were given a reward of £2 for each correct guess. Table 4 reports the results from a (robust) Tobit regression of the guessed average amount donated by followers. Two features of the results stand out: first, subjects who donate more also expect others, on average, to donate more. Second, subjects expect others to be influenced by

<sup>19</sup>Since we have no reason to expect non-monotonic effects for these variables, we simply use the scores.

**Table 4:** Followers’ expectations and leader announcements. Robust Tobit regression.

Dependent variable: Followers’ expectations about average donation of other followers

	Coef.	p-value
Own donation	0.402***	0.000
Both £4	0.792***	0.004
£4 and £0	-0.102	0.659
Both £0	-1.155***	0.004
Public	0.130	0.479
Constant	1.909***	0.000
<i>N</i>	114	
Log pseudolikelihood	-164.98391	

the leader(s)’ example, giving more after observing two “generous” leaders instead of one, and less after observing two “selfish” leaders. These effects are significant, while the corresponding effects on actual behavior, examined in section 4, were not: in a sense, our participants appear to expect greater “conformity” from others than they exhibit themselves.

## 6 Conclusions

A growing body of empirical evidence points to the importance of social signaling as a driver of prosocial behavior (e.g. Dana, Cain, and Dawes, 2006; Andreoni and Petrie, 2004; Rege and Telle, 2004; Soetevent, 2005; Ariely, Bracha, and Meier, 2009; Andreoni and Bernheim, 2009.) In this paper, we have focused instead on the role of self-signaling. This yielded testable predictions concerning the relationship between self-confidence and prosocial behavior.

Our experimental findings lend support to these predictions, and suggest that self-signaling can play an economically important role.

## References

- ALPIZAR, F., F. CARLSSON, AND O. JOHANSSON-STENMAN (2008): “Anonymity, reciprocity, and conformity: Evidence from voluntary contributions to a national park in Costa Rica,” *Journal of Public Economics*, 92(5-6), 1047 – 1060.
- ANDREONI, J. (1990): “Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving?,” *Economic Journal*, 100(401), 464–77.

- ANDREONI, J., AND B. D. BERNHEIM (2009): “Social Image and the 5050 Norm: A Theoretical and Experimental Analysis of Audience Effects,” *Econometrica*, 77(5), 11 – 28.
- ANDREONI, J., AND R. PETRIE (2004): “Public goods experiments without confidentiality: a glimpse into fund-raising,” *Journal of Public Economics*, 88(7-8), 1605 – 1623.
- ARIELY, D., A. BRACHA, AND S. MEIER (2009): “Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially,” *American Economic Review*, 99(1), 544–55.
- BARDSLEY, N., AND R. SAUSGRUBER (2005): “Conformity and reciprocity in public good provision,” *Journal of Economic Psychology*, 26(5), 664 – 681.
- BEM, D. J. (1967): “Self-Perception: An Alternative Interpretation of Cognitive Dissonance Phenomena,” *Psychological Review*, 74(3), 183–200.
- BÉNABOU, R., AND J. TIROLE (2010): “Identity, Morals and Taboos: Beliefs as Assets,” *Quarterly Journal of Economics*, (forthcoming).
- BENJAMIN, D., J. CHOI, AND J. STRICKLAND (2010): “Social Identity and Preferences,” *American Economic Review*, (forthcoming).
- BERNHEIM, B. D. (1994): “A Theory of Conformity,” *Journal of Political Economy*, 102(5), 841–77.
- BURKS, S. V., J. P. CARPENTER, L. GOETTE, AND A. RUSTICHINI (2010): “Overconfidence is a Social Signaling Bias,” *IZA Discussion Paper*, 4840.
- CROSON, R., AND J. SHANG (2008): “The impact of downward social information on contribution decisions,” *Experimental Economics*, 11(3), 221–233.
- DANA, J., D. M. CAIN, AND R. M. DAWES (2006): “What you don’t know won’t hurt me: Costly (but quiet) exit in dictator games,” *Organizational Behavior and Human Decision Processes*, 100(2), 193 – 201.
- DESSÍ, R., E. GALLO, AND S. GOYAL (2010): “Network Cognition and Behavior,” *Paper presented at the Workshop on Information and Behavior in Networks, Nuffield College, Oxford*.

- DUNNING, D. (2001): “On the motives underlying social cognition,” in A. Tesser & N. Schwarz (eds.) *Blackwell Handbook of Social Psychology: Intraindividual Processes*.
- FISCHBACHER, U., S. Gächter, AND E. FEHR (2001): “Are people conditionally cooperative? Evidence from a public goods experiment,” *Economics Letters*, 71(3), 397 – 404.
- FREY, B. S., AND S. MEIER (2004): “Social Comparisons and Pro-Social Behavior: Testing ”Conditional Cooperation” in a Field Experiment,” *The American Economic Review*, 94(5), pp. 1717–1722.
- GAECHTER, S., D. NOSENZO, E. RENNER, AND M. SEFTON (2009): “Who Makes A Good Leader? Cooperativeness, Optimism And Leading-By-Example,” Discussion Papers 2009-19, The Centre for Decision Research and Experimental Economics, School of Economics, University of Nottingham.
- GINO, F., S. AYAL, AND D. ARIELY (2009): “Contagion and Differentiation in Unethical Behavior,” *Psychological Science*, 20(3), 393–398.
- HELDT, T. (2005): “Conditional cooperation in the field : Cross-country skiers’ behavior in Sweden,” *Working Paper. Dalarna University College Electronic Archive*.
- HERMALIN, B. E. (1998): “Toward an Economic Theory of Leadership: Leading by Example,” *American Economic Review*, 88(5), 1188–1206.
- KRUPKA, E., AND R. A. WEBER (2009): “The focusing and informational effects of norms on pro-social behavior,” *Journal of Economic Psychology*, 30(3), 307 – 320.
- LEBOEUF, R. A., E. SHAFIR, AND J. B. BAYUK (2010): “The conflicting choices of alternating selves,” *Organizational Behavior and Human Decision Processes*, 111(1), 48–61.
- MARTIN, R., AND J. RANDAL (2008): “How is donation behaviour affected by the donations of others?,” *Journal of Economic Behavior & Organization*, 67(1), 228–238.
- MAZAR, N., O. AMIR, AND D. ARIELY (2008): “The Dishonesty of Honest People: A Theory of Self-Concept Maintenance,” *Journal of Marketing Research*, Vol. 45, No. 6, pp. 633-644, 2008.

- MAZAR, N., AND D. ARIELY (2006): “Dishonesty in Everyday Life and its Policy Implications,” *Journal of Public Policy and Marketing*, Vol. 25, No. 1, pp. 117-126, 2006.
- MESSICK, D. M. (1999): “Alternative logics for decision making in social settings,” *Journal of Economic Behavior & Organization*, 39(1), 11 – 28.
- PATRICK, C. J., J. J. CURTIN, AND A. TELLEGEN (2002): “Development and validation of a brief form of the Multidimensional Personality Questionnaire,” *Psychological Assessment*, 14(2), 150 – 163.
- POTTERS, J., M. SEFTON, AND L. VESTERLUND (2007): “Leading-by-example and signaling in voluntary contribution games: an experimental study,” *Economic Theory*, 33, 169–182, 10.1007/s00199-006-0186-3.
- REGE, M., AND K. TELLE (2004): “The impact of social approval and framing on cooperation in public good situations,” *Journal of Public Economics*, 88(7-8), 1625–1644.
- SOETEVEENT, A. R. (2005): “Anonymity in giving in a natural context—a field experiment in 30 churches,” *Journal of Public Economics*, 89(11-12), 2301–2323.
- TADELIS, S. (2007): “The Power of Shame and the Rationality of Trust,” *SSRN eLibrary*.
- TELLEGEN, A. (1988): “The analysis of consistency in personality assessment,” *Journal of Personality*, 56, 621 – 663.
- TELLEGEN, A., AND N. G. WALLER (1994): “Exploring personality through test construction: Development of the Multi-dimensional Personality Questionnaire (MPQ),” *Personality measures: Development evaluation*.

## 7 Appendix

### 7.1 Instructions

*General rules:*

During this experiment, all your answers and decisions will be recorded in a completely anonymous manner, and the resulting data will be identified only by the **ID number** you picked randomly before starting the experiment.

Please type in your ID number, exactly as it is written on the ticket you picked before the start of the experiment.

If you have a question, please raise your hand.

If there are no questions, we can move on to the specific instructions.

-NEXT SCREENSHOT-

*Specific instructions (stage 1):*

This experiment studies decisions to donate to charity. There are two parts to the experiment. In both parts, *participants who are given an endowment have to decide how much to keep for themselves and how much to give to charity*. We will provide specific instructions for each part before the start.

**In the first part, one participant will be picked randomly to make a decision: we refer to him or her as "the decision-maker".**

The decision-maker will receive an endowment of **£10**. He or she will have to decide whether to **keep all the endowment or to donate £4 to a charity**. The charity can be chosen out of three possibilities: the Red Cross, Save the Children, and Amnesty International.

We will now explain the procedure in detail.

-NEXT SCREENSHOT-

The following information will appear on participant's screens.

All participants, including the decision-maker, will see some information describing the three charities. Then the decision-maker will find out that he or she has been randomly picked to make the decision, while the other participants find out that they have not.

On the following screen, the decision-maker will be asked to specify his or her allocation choice. The other participants will be asked a completely unrelated question (so that the decision-maker cannot be readily identified as the only participant answering a question).

Finally, **all participants will see on their screen whether the decision-maker chose to give £4 to charity or not.**

ANONYMOUS TREATMENT ONLY: Note that the identity of the decision-maker will never become known to other participants, nor to the experimenters who will identify participants only by their ID codes. Moreover, payments at the end of the experiment will be made so as to ensure that the experimenters cannot deduce any participants decisions from his or her earnings.

PUBLIC TREATMENT ONLY: At this point, the **decision-maker** will be asked to **stand up and hold up** either a **card** saying “*Zero to charity*” or a card saying “*Four pounds to charity*”, corresponding to his or her choice. We will walk out of the laboratory while this happens, and another person who is not one of the experimenters and will have nothing further to do with the experiment will come in briefly to bring the two cards, leaving as soon as the decision-maker holds up the relevant card.

The decision-maker will then sit down again, and the experiment will continue.

This is to ensure that the identity of the decision-maker will never become known to the experimenters. Note that payments at the end of the experiment will be made so as to ensure that the experimenters cannot deduce any participants decisions from his or her earnings.

-NEXT SCREENSHOT-

Here is some information about the three charities:

*Amnesty International*

We are a campaigning organisation; it’s what we do. Our purpose is to protect people wherever justice, fairness, freedom and truth are denied. We work on lots of issues. Right now our priority campaigns are Stop Violence Against Women and Terrorism, Security and Human Rights.

*British Red Cross (Haiti Earthquake Appeal)*

Drawing on resources around the world the Red Cross is carrying out a huge emergency response operation. This includes the provision of vital supplies (such as kitchen kits, shelter kits, personal hygiene kits, blankets and containers for storing drinking water) as well as providing medical aid and water for survivors.

*Save the Children*

We’re working flat out to ensure children get proper healthcare, food, education and protection. We’re saving lives in emergencies, campaigning for children’s rights, and improving their futures through long-term development work.

-NEXT SCREENSHOT-

*Leader’s screenshot:* Your ID code was randomly selected to be the decision-maker’s.

We are giving you an endowment of **£10**. Please specify your decision:



I keep all the endowment for myself  
Or  
I keep 6 for myself and give 4 to charity.

I would like my donation (if any) to go to:

Amnesty International  
Or  
Red Cross  
Or  
Save the Children

*Follower's screenshot:* Your ID code was not selected to be the decision-maker's.

Please answer the following question: Which of these charities are you most familiar with?

Amnesty International  
Or  
Red Cross  
Or  
Save the Children

-NEXT SCREENSHOT-

**PUBLIC TREATMENT ONLY:** Please now wait for the decision-maker to stand up and show the card corresponding to his or her decision. Do not click on "Next" until the decisionmaker sits down again.

*Specific instructions (stage 2):*

In this part of the experiment, **all participants** are given an endowment of **£10**.

They all **independently and privately** choose how much to keep for themselves and how much to donate to a charity. Any integer amount can be donated. The recipient charity can be chosen from the same list of three used in the previous part: Save the Children, Red Cross and Amnesty International.

-NEXT SCREENSHOT-

We are giving you an endowment of **£10**. Please specify your decision:

I donate the following amount to charity:  
I would like my donation (if any) to go to:

-NEXT SCREENSHOT-

The decision-making part of the experiment is now complete. We would like you to make two guesses.

If at least one of your answers is correct, you will receive **£2** in addition to your other earnings.

How much do you think was donated this time by the decision-maker from the first part of the experiment?

How much do you think was donated on average by each participant this time, excluding yourself and the decision-maker?

-NEXT SCREENSHOT-

The experiment has now ended. Thank you for your participation. We now ask you to complete the following questionnaire.

As soon as you finish the questionnaire you can leave the laboratory.

Before the questionnaire starts an experimenter will come to check if you brought an ID number given to you in a previous experiment. This is done so that participants who already took part in a previous study don't have to fill out the same questionnaire twice.