SIEPR Discussion Paper No. 12-022

Incentives for Prosocial Behavior: The Role of Reputations

By

Christine Exley

Stanford Institute for Economic Policy Research
Stanford University
Stanford, CA 94305
(650) 725-1874

Incentives for Prosocial Behavior: The Role of Reputations

Christine Exley[a]

[a]*Stanford University, Department of Economics, 579 Serra Mall, Stanford, CA 94305-6072, USA. clexley@stanford.edu, 1-757-748-3515*

Abstract

In public settings, the impact of monetary incentives on prosocial behavior is empirically mixed. Existing theory explains these findings by noting that incentives can introduce public signals that may or may not crowd out motivation to volunteer. The strength of these public signals are normally unobserved by the researcher, so it remains unclear as to when significant crowding out is likely to occur and render incentives ineffective. I overcome this ambiguity by examining individuals for whom the signal strength is likely zero - those with strong public reputations. In a laboratory experiment, I show that the crowd out in response to public incentives is much less likely among those with public reputations as opposed to private reputations, particularly for women.

JEL: D64, J16, Z13

Keywords: altruism; monetary incentives; image motivation; gender

1. Introduction

With over a quarter of Americans volunteering annually at an estimated market value of 173 billion dollars, understanding how to encourage volunteers to provide more help may yield significant benefits to crucial societal services.[1] While offering monetary incentives may extrinsically motivate volunteers, they may also backfire by crowding out intrinsic motivation.[2] Bénabou and Tirole (2006) closely investigate this crowd-out by separately considering intrinsic and extrinsic motivations from image motivation - how much individuals value their public reputation. In particular, in their model of prosocial behavior, public incentives may crowd out image motivation since publicly accepting an incentive introduces a greedy signal that hinders one's ability to appear prosocial.[3] Ariely et al. (2009) correspondingly find that incentives effectively encourage private volunteer behavior, but fail to encourage public volunteer behavior where crowding out of image motivation is a concern. While some studies similarly provide evidence of this image motivation crowd out, others find that incentives increase public volunteer behavior. That is, it remains theoretically and empirically unclear as to when significant crowding out is likely to occur and render incentives ineffective.[4]

This paper proposes a mechanism where the observability of reputations impacts the likelihood of crowding out of image motivation. In particular, among individuals with strong public reputations, the strength of image signals is likely zero, as no single piece of evidence will be informative either for or against their well-established type. Hence, their image motivation crowd-out due to the greedy signal of publicly accepting an incentive is unlikely.

This paper extends Bénabou and Tirole (2006)'s model to account for the role of reputations and tests three key image-related effects in a laboratory experiment. First, to determine if public incentives generate a greedy signal and hence crowd out image motivation to volunteer, this study tests *The Negative Image Effect - public incentive offers, relative to private incentive offers, discourage public volunteer behavior.* Note that by only manipulating the observability of the incentive offer (not the observability of the volunteer behavior or the level of the incentive), the Negative Image Effect excludes mechanisms other than the greedy signal that may cause crowding out to occur.[5] Second, to examine if public reputations decrease the strength of the prosocial signal and hence crowd out image motivation to volunteer, this study tests *The Reputations Effect - public reputations, relative to private reputations, discourage public volunteer behavior.*[6] Third, to investigate if public reputations decrease the strength of the greedy signal and hence prevent public incentives from crowding out image motivation, this study tests *The Interactions Effect - among individuals with public*

[1] See Corporation for National and Community Service (2011) and Independent Sector (2010).

[2] This crowding out of intrinsic motivation was first mentioned in Titmuss (1970), modeled in Bénabou and Tirole (2003), and argued in many empirical studies, such as Frey and Oberholzer-Gee (1997), Gneezy and Rustichini (2000) and Frey and Jergen (2001). Some later studies provide evidence against this crowding out, such as Goette and Stutzer (2008) and Ashaaf et al. (2012). Also, see Gneezy et al. (2011) for a survey of incentives' varying effects across a variety of contexts.

[3] While image motivations can be considered extrinsic motivations since it involves motivations that rely on external conditions, this paper will always distinguish the two by only referring to non-image external conditions, such as incentives, as extrinsic motivations.

[4] Gneezy and Rustichini (2000) find a negative effect for small incentives offered to school children eliciting donations but no effect for larger incentives; Mellström and Johannesson (2008) find a negative effect of incentivizing the completion of a health examination to become blood donors among females and a null effect among males; Ariely et al. (2009) find a null effect of volunteer incentives in the laboratory; Carpenter and Myers (2010) find a positive effect of small stipends among volunteer firefighters, unless they have vanity license plates; Lacetera et al. (2011) and Lacetera et al. (2012) find positive effects of a variety of incentives in large scale natural field experiments of American Red Cross blood drives.

[5] For instance, the impact of incentives on public versus private volunteer behavior may also result from diminishing returns to incentives as discussed in Section 5. Also, note that Ariely et al. (2009) point out that a situation with a private incentive offer would eliminate the greedy signal, which is an idea that corresponds with the this study's approach to test the image crowding out hypothesis.

[6] Note that manipulating the observability of one's current reputation differs from the well-established result that an individual positively changes his behavior if he knows his future reputation will be public (Harbaugh, 1998b,a; Bénabou and Tirole, 2006; Andreoni and Bernheim, 2009; Ariely et al., 2009; Lacetera and Macis, 2010b).

reputations, the observability of the incentive offer does not impact public volunteer behavior.

The above effects imply that public incentives are more likely to crowd out image motivation of individuals with private reputations than public reputations because the strength of the greedy signal is near zero for the latter group. Interestingly, this stands as a promising explanation for the mixed empirical findings in the literature: studies that find a net positive effect of incentivizing public volunteer behavior often involve established volunteers with likely public reputations (Carpenter and Myers, 2010; Lacetera et al., 2011, 2012), while studies which conclude that incentives fail to encourage more public volunteer behavior often involve unestablished or new potential volunteers with likely private reputations (Gneezy and Rustichini, 2000; Mellström and Johannesson, 2008; Ariely et al., 2009).[7]

Outside of the theory, but likely of interest to the large but mixed gender literature on prosocial preferences, this paper examines if the above image-related effects are stronger for females. Gender differences in these image-related effects may be particularly relevant to Mellström and Johannesson (2008) and Lacetera and Macis (2010a); they find that females but not males are discouraged when monetary incentives are introduced from completing a health examination to become a blood donor and from donating blood, respectively. Gender differences may also add to the existing evidence that females' actions are more easily influenced, especially in prosocial and experimental contexts.[8]

The experiment involves a simple, incentivized, and publicly observable volunteer task, called the clicking task, that is very similar to Ariely et al. (2009). The clicking task is simple since it merely requires participants to push a button (i.e., "click") on an electronic tally counter for eight minutes. For every five clicks a participant completes, the experimenter donates one cent to the American Red Cross (ARC), so the more a subject clicks, the more she is effectively volunteering for the ARC. The clicking task is incentivized at the same rate for which she earns money for the ARC: for every five clicks a participant completes, the experimenter adds one cent to her study compensation. The clicking task is publicly observable since a participants' level of volunteering, i.e., number of clicks, is publicly revealed to other participants, who observe but do not participate in the clicking task.

To experimentally test the three image effects and potential gender differences, a two-by-two design is implemented by varying the observability (public or private) of participants' incentive offers and reputations. Consistent with females being more image-motivated or susceptible to influences in the lab than males, I find strong support for all three testable predictions among females but not males: (1) The Negative Image Effect - public incentives discourage clicking by about nine percent, (2) The Reputations Effect - public reputations discourage clicking also by about nine percent, and (3) The Interactions Effect - among those with public reputations, there is no significant difference in the impact of public and private incentives on clicking.[9] That is, the observability of reputations impacts when public incentives are likely to crowd out image motivation and hence when public incentives are likely to be effective.

This paper proceeds as follows: Section 2 derives the testable predictions, Section 3 details the study design, Section 4 presents the results, and Section 5 concludes.

[7] Note that Carpenter and Myers (2010) find a null effect of incentives for people with vanity license plates. Also, encouragingly, Lacetera et al. (2011) find that monetary incentives at a *particular* blood drive led to, in absolute value, significantly more donations those with likely public reputations from previously donating at that *particular* drive relative to those with likely private reputations since they have not previously donated at that drive.

[8] Andreoni and Vesterlund (2001) details a careful experiment that demonstrates how gender results for altruistic behavior is price sensitive. In their survey paper, Croson and Gneezy (2009) argue that the large literature on gender differences in social preferences, in particular with respect to altruism and generosity, may be explained by females being "more sensitive to subtle cues than" men. DellaVigna et al. (2013) similarly find in a field experiment that females are more likely to be on the margin of giving and volunteering.

[9] Note that these results only apply to females who are above average volunteers, because of heterogeneous treatment effects among types of volunteers, which accounts for 76% of all females.

2. Theoretical Framework

I derive testable predictions from an extension of the model of prosocial behavior in Bénabou and Tirole (2006), henceforth BT. Agents in BT's model receive utility from participating in a prosocial activity, or volunteering, at a level of a that is incentivized at a rate of y according to their intrinsic, extrinsic, and image motivations as well as their private cost of volunteering. An agent's intrinsic motivation, whether purely altruistic or warm-glow related, involves her "prosocial" preference for how much she values volunteering, v_a. An agent's extrinsic motivation results from the incentive rate y and her "greedy" preference for how much she values money, v_y, and together with her intrinsic motivation, leads to a utility of $(v_a + yv_y)\,a$. The most crucial component of BT's model to this study involves an agent's image motivation or payoff, $I(a, y)$, that she receives from volunteering given an incentive.[10] This image payoff is the combination of how prosocial or greedy others believe the agent is based on her incentivized volunteer behavior, $\mathbb{E}(v_a|a, y)$ and $\mathbb{E}(v_y|a, y)$, weighted by her preference to appear prosocial and not greedy to others, μ_a and μ_y. Combining these motivations with the private cost of volunteering, $C(a)$, yields the following utility for an agent:

$$U(a, y) = (v_a + yv_y)\,a + I(a, y) - C(a) \tag{1}$$

where

$$I(a, y) = \mu_a \mathbb{E}(v_a|a, y) - \mu_y \mathbb{E}(v_y|a, y), \tag{2}$$

To solve the model, there are three helpful assumptions that I use from BT. First, I assume a specific convex cost function $C(a) = \frac{ka^2}{2}$ for some fixed parameter k. Second, I assume that all agents have the same image preferences, $\mu = (\bar{\mu}_a, \bar{\mu}_y)$. Third, I assume that agents' types, (v_a, v_y), are normally distributed as follows:

$$\begin{pmatrix} v_a \\ v_y \end{pmatrix} \sim \mathbf{N}\left(\begin{pmatrix} \bar{v}_a \\ \bar{v}_y \end{pmatrix}, \begin{pmatrix} \sigma_a^2 & \sigma_{ay} \\ \sigma_{ay} & \sigma_y^2 \end{pmatrix} \right).$$

Then, by solving for the conditional expectations through a signal extraction problem and a resulting differential equation (see Appendix C for details), the first order condition yields:

$$a(y, k) = \frac{v_a + yv_y}{k} + \mu_a \frac{\sigma_a^2 + y\sigma_{ay}}{\sigma_a^2 + 2y\sigma_{ay} + y^2\sigma_y^2} - \mu_y \frac{y\sigma_{ay} + y^2\sigma_y^2}{y(\sigma_a^2 + 2y\sigma_{ay} + y^2\sigma_y^2)}.$$

By imposing the assumption that $\sigma_{ay} = 0$, the aggregate level of volunteer behavior is:

$$\bar{a}(y, k|A_{pub}O_{pub}R_{prvt}) = \frac{\bar{v}_a + y\bar{v}_y}{k} + \underbrace{\bar{\mu}_a \frac{\sigma_a^2}{\sigma_a^2 + y^2\sigma_y^2} - \bar{\mu}_y \frac{y\sigma_y^2}{\sigma_a^2 + y^2\sigma_y^2}}_{\theta(a,y)}, \tag{3}$$

where $\theta(a, y)$ denotes the image payoff that agents receive from others observing them volunteering at level a for a given incentive rate y.[11] Also, I define this benchmark result as under the conditions of public volunteer behavior (A_{pub}), public incentive offers (O_{pub}) since it is assumed that others know the agents' incentive rate, and private reputations (R_{prvt}) since it is assumed that others do not know the agents' types (v_a, v_y). In addition to BT's assumptions in deriving their results, I impose one additional and simplifying assumption about agents' reputations:

[10] Earlier papers also included components where people value others' perception of them, such as in Bernheim (1994).

[11] $\sigma_{ay} = 0$ implies independence between greedy and prosocial preferences in the population, and is a simplifying assumption that Bénabou and Tirole (2006) also use as their benchmark case.

Additional Assumption (Reputations Assumption)**.** *An agents' reputation perfectly identifies their type* (v_a, v_y).[12]

While the benchmark case involves a greedy signal associated with a public incentive offer, I slightly modify BT's model to consider a case with private incentive offers (O_{prvt}) that does not necessarily involve a greedy signal. In particular, if the incentive offer is private, then I assume y_b is the fixed incentive rate that others believe an agent receives, although y is the true incentive rate.[13] This assumption may only be reasonable for when others believe an agent did not receive an incentive ($y_b = 0$), which is the relevant assumption for this study as discussed later in the experimental design. Her utility under private incentive offers is then:

$$U(a, y) = (v_a + y v_y)\, a + \mu_a \mathbb{E}(v_a | a, y_b) - \mu_y \mathbb{E}(v_y | a, y_b) - C(a).$$

After solving in a similar manner to the benchmark case (see Appendix C), this yields the following aggregate level of volunteer behavior:

$$\bar{a}(y, k | A_{pub} O_{prvt} R_{prvt}) = \frac{\bar{v}_a + y\bar{v}_y}{k} + \underbrace{\bar{\mu}_a \frac{\sigma_a^2}{\sigma_a^2 + y_b^2 \sigma_{y_b^2}} - \bar{\mu}_y \frac{y_b \sigma_{y_b^2}}{\sigma_a^2 + y_b^2 \sigma_{y_b^2}}}_{\theta(a, y_b)}. \tag{4}$$

When $y_b \to 0$, note that a private incentive offer causes an agent to receive $\theta(a, y_b) = \bar{\mu}_a$, which is the largest potential image payoff. In contrast, a public incentive offer leads to a lower image payoff of $\theta(a, y) = \bar{\mu}_a \frac{\sigma_a^2}{\sigma_a^2 + y^2 \sigma_y^2} - \bar{\mu}_y \frac{y \sigma_y^2}{\sigma_a^2 + y^2 \sigma_y^2}$, since the agent now only receives a fraction of the potential prosocial image payoff and there is a negative greedy image payoff. More generally, this result is summarized in Prediction 1 below.

Prediction 1 (The Negative Image Effect)**.** *In the case of incentivized and public volunteer behavior, with private reputations, public incentive offers discourage volunteer behavior, relative to private incentive offers, as* $y_b \to 0$.[14]

In order to derive the second and third predictions, l consider BT's model under public reputations, instead of private reputations, so an agent's type, (v_a, v_y), is public. Then, $I(a, y)$ is fixed, so she receives no additional image payoff from volunteering, or in other words:

$$\bar{a}(y, k | A_{pub} O_{pub} R_{pub}) = \bar{a}(y, k | A_{pub} O_{prvt} R_{pub}) = \frac{\bar{v}_a + y\bar{v}_y}{k}. \tag{5}$$

Since $\theta(a, y) = \theta(a, y_b) = 0$ in Equation 5, two more testable predictions immediately follow.

Prediction 2 (The Reputations Effect)**.** *In the case of incentivized and public volunteer behavior, with private incentive offers, public reputations discourage volunteer behavior, relative to private reputations, as* $y_b \to 0$.[14]

Prediction 3 (The Interactions Effect)**.** *In the case of incentivized and public volunteer behavior, with public reputations, the observability of the incentive offer has no impact on volunteer behavior.*

The following section describes the experiment designed to test these three predictions by exogenously varying the observability of participants' incentive offers and reputations in a volunteer task.

[12] Note that an agent's reputation can be defined by a variety of characteristics as long as those characteristics perfectly identify their type.

[13] Introducing y_b to allow for private incentive offers is one way to separate the observability of y from a. In BT, there is an observability parameter x that is collapsed into the μ_a and μ_y terms. So, an approach that would be similar *in spirit* would be to have separate observability parameters x_a and x_y, although this would still require some assumption about what people believe y is when they do not observe it and then the corresponding theoretical predications would likely be similar.

[14] Predictions 1 and 2 hold more generally as long as $\theta(a, y_b) > \theta(a, y)$ or $\theta(a, y_b) > 0$, respectively, but $y_b \to 0$ is the relevant condition for this study, as explained later in the experimental design.

3. Experimental Design

This study involves a simple, incentivized, and publicly observable volunteer task, called the clicking task, that is very similar to Ariely et al. (2009).[15]

The clicking task is simple since it merely requires Clicker Participants (CPs) to push a button on, or "click," an electronic tally counter (shown in Appendix A.1) for eight minutes. For every five clicks a CP completes, the experimenter donates one cent to the American Red Cross (ARC), so the more a CP clicks, the more she is effectively volunteering for the ARC.

The clicking task is incentivized at the same rate for which she earns money for the ARC: for every five clicks a CP completes, the experimenter adds one cent to her study compensation.

The clicking task is publicly observable since a CP's level of volunteering, i.e., number of clicks, is publicly revealed to other participants called Panel Members (PMs). However, a CP's level of volunteering is not observable to other CPs, which helps to avoid competition between CPs from driving the results - a potential concern because of the well-established gender result that males are more eager to participate in and perform better in competitions (Niederle and Vesterlund, 2011). Also, exactly two PMs observe each CP, and each PM is incentivized to carefully record information about the CPs, including their number of clicks.[16]

Given the above properties of the clicking task, this study implements an across-subject two-by-two design that only varies according to whether (1) incentive offers to volunteer are public or private, and (2) volunteer reputations are public or private. Table 1 summarizes these treatments. To ensure participants understand the design, participants practice, receive instructions about, and are required to correctly complete an understanding quiz.

Table 1: Treatment Groups

	Public Reputations (R_{pub})	**Private Reputations (R_{prvt})**
Public Incentive Offers (I_{pub})	PMs informed of CPs' incentive offers and reputations (n=31)	PMs informed of CPs' incentive offers but not reputations (n=28)
Private Incentive Offers (I_{prvt})	PMs not informed of CPs' incentive offers but informed of CPs' reputations (n=38)	PMs not informed of CPs' incentive offers nor reputations (n=33)

In all treatment groups, CPs participate in the clicking task, which is a simple, incentivized and publicly observable volunteer task. The observability of the incentive offer varies across study sessions, while the observability of reputations varies within each study session. n displays the number of participants in each treatment group.

The observability of the incentive offers varies on the session level so that CPs within a study session either all have public or all have private incentive offers. CPs always know their incentive offers, but PMs are only aware of their incentive offers in the public condition. This difference in information is accomplished by having PMs leave the study room to complete an unrelated and non-incentivized task, such as a maze, to keep them occupied for several minutes. During this time, CPs are informed of their incentive offers to volunteer in both the public and private condition. When PMs return to the study room in the public condition, they are informed of these incentive offers. However, when PMs return to the study room in the

[15] In Ariely et al. (2009), participants click pairs of keys on a computer, and the experimenter donates to charity according to an increasing payment scheme in clicks. In incentivized treatment groups, participants receive an identical payment scheme to keep for themselves that is randomly implemented. In observable treatment groups, participants publicly announce how much money they earn for charity.

[16] Each PM is informed that he will receive an additional $5 in compensation if he correctly records a selected piece of information from one of his CPs. PMs are informed that this information may be a CP's number of clicks or some component of the CP's volunteer reputation if it is revealed.

private condition, they are not informed of these incentive offers nor even the possibility that there could be incentives to volunteer in this study.[17] The results from an incentivized understanding question show that this experimental protocol leads to, when incentive offers are private, PMs believing that the CPs receive no incentive and CPs also believing that PMs believe this.[18] In other words, this is consistent evidence with $y_b = 0$, which is the sufficient condition needed for the Negative Image Effect and the Reputations Effect to hold (and the Interactions Effect holds regardless).

While the observability of incentive offers is determined on the session level, volunteer reputations are randomly set as public or private for half of all CPs within each session. CPs will always observe their own reputation, but PMs only observe the CPs' reputations in the public condition. CPs self-report information relevant to their reputation at the beginning of the study, before knowing any details about the study except that this information may be revealed to the PMs with a 0.50 probability.[19] Subsequently, each CP is the labeled as an "above average volunteer" if they volunteered 23 or more hours in the past year (the national average among young adults) or a "below average volunteer" otherwise.[20] Interestingly, no known study examines the observability of such a reputation based on past volunteer behavior, perhaps because of the largely unavoidable endogeneity issues in the field (i.e., the more one volunteers, the more likely she is to possess a public reputation). Finally, note that if a CP's reputation correlates with her prosocial and greedy preferences, then this is consistent with a CP's reputation identifying her type (this study's reputations assumption), or effectively reducing the strength of the image signals.

After the clicking task is completed, participants are surprised with four activities, starting with two dictator games. First, in the "PMs' Allocations to CPs" dictator game, PMs are informed that they may receive an additional \$10 in compensation from which they must allocate \$2 or \$4 to one of their CPs. PMs then privately allocate amounts for each of their CPs, and one allocation is randomly implemented such that the anonymity of the PMs' allocations are ensured.[21] The results from this game will provide insight into how PMs perceive above versus below average public reputations. Second, in the "CPs' Donations to the ARC" activity, CPs are informed that they may receive an additional \$10 in compensation from which they must decide whether to allocate any dollar amount from \$0 - \$10 to the ARC. CPs then make their donation decisions, knowing that PMs will observe these decisions and one donation decision will be randomly implemented. The results from this game provide insight into whether above average volunteers are more altruistic or image motivated than below average volunteers. Third, in "Round 2," CPs click for one minute to only earn money for themselves (not the ARC as well) at the same incentive rate as in the clicking task. CPs' number of clicks are private in Round 2 and used as a control later in the analysis. Finally, all

[17] Note that at no point were PMs misinformed about incentives for CPs.

[18] CPs and PMs complete an incentivized understanding quiz of 3 and 4 questions, respectively. When PMs are asked "How much money will be given to the CP?" in a situation where the CP clicks 500 times, all PMs answer \$0 in the private incentive offer condition. When CPs are asked to indicate what they believe PMs answer to this question will be, the majority indicate \$0, with only 8.22% indicating \$1 (the true incentive offer) in the private incentive offer condition.

[19] In order to ease concerns about the self-reporting nature, CPs volunteer behavior is self-reported while PMs are not in the study room, and their volunteer behavior is presented to PMs without any mention of it being self-reported. Several studies also find that study participants do not lie much, and perhaps most encouragingly, Gneezy (2005) provides evidence that participants with a chance to lie believe that other participants will believe they are telling the truth.

[20] This cutoff of 23 hours was determined by calculating the average volunteer hours among young adults (16-24) from the Corporation for National & Community Service 2010 data about volunteering in America. If CPs reputations were public, then their stated number of volunteer hours and their willingness to participate in an additional survey were also public. Note that the latter had no impact on their classification as an above/below average volunteer. Participants were informed that this additional survey would take approximately 5-10 minutes, would be given on behalf of Stanford's Haas Center for Public Service, and could be completed immediately after their participation in this study or later via a link sent to them by email. This participation is used as a control later in this analysis.

[21] A CP may receive an allocation from either of her two PMs, so even if a CP receives an additional \$2 or \$4 as part of her compensation, she does not know from which PM's decision this resulted.

participants complete a short follow-up survey to gather demographic and other relevant information.

The subjects were 168 undergraduate students from Stanford University. For both the public and the private incentive offer conditions, there were five study sessions run between January and March 2012 in the Stanford Economics Research Laboratory.[22] Within each study session, random assignment to one role only resulted in 134 CPs and 34 PMs split over sessions such that 2 PMs always observed 5-10 CPs.[23] Also, among CPs within each study session, there were equal numbers of CPs (plus or minus one CP) randomly assigned to the public and private reputation conditions.[24] Four CPs are excluded in the analysis for not properly following instructions about how to use the electronic tally counter.[25] See Table 3.2 for descriptive data about the participants, and Tables Appendix B.1 and Appendix B.2 for more descriptive data broken down by the observability of participants' incentive offers and reputations.

Table 3.2: Average Characteristics of Participants

Above Average Volunteer	0.70
Willingness to Complete an Additional Survey	0.74
Male	0.52
Economics Major	0.12
Born in United States	0.78
Feel Favorably about ARC (as opposed to Neutral or Unfavorably)	0.77
Senior	0.10
Junior	0.12
Sophomore	0.19
Freshman	0.58
Volunteer Hours (in Past Year)	70.90
N	130

All of the above values indicate the fraction of participants with a given characteristic, except for the values associated with Volunteer Hours which indicate the average volunteer hours for participants.

Given this simple two-by-two design, the results relevant to the previously derived three predictions easily follow, as detailed in the following section.

4. Results

Table 4.1 displays the average level of clicks across the treatment groups.[26] While the differences in average clicks across these treatment groups are qualitatively consistent with the three theoretical predictions, these differences in clicking levels are not statistically significant. As detailed in Results 1-3, there is insufficient evidence for the Negative Image and Reputations Effects among the full sample, and while the Interactions Effect is confirmed, this is not very interesting given the lack of evidence for the first two effects.

Result 1 (The Negative Image Effect)**.** *Among the full sample, there is insufficient evidence to conclude that public incentive offers discourage volunteer behavior.*

[22] To limit possible time effects from running sessions on five separate days, I ran exactly one public and one private incentive offer condition each day but varied the order in which these conditions were run.

[23] To randomly assign roles, upon arrival to the study, each participant selects a marked poker chip from a bag that denotes if he/she was a PM or CP.

[24] Participants who sat at every other computer desk were in the public or private reputation condition.

[25] Participants were instructed to hold the electronic tally counter in only one self-chosen hand and to only use their thumb to push the button. Excluded participants did not do this.

[26] See Table Appendix B.3 for more descriptive statistics on clicks.

Result 2 (The Reputations Effect)**.** *Among the full sample, there is insufficient evidence to conclude that public reputations discourage volunteer behavior.*

Result 3 (The Interactions Effect)**.** *Among the full sample, the observability of the incentive offer does not significantly impact volunteer behavior.*

Table 4.1: Average Level of Clicks Across Treatment Groups Groups

	Public Reputations (R_{pub})	**Private Reputations (R_{prvt})**
Public Incentive Offers (I_{pub})	2159.94 (175.08)	2200.46 (195.54)
Private Incentive Offers (I_{prvt})	2181.71 (239.30)	2268.00 (286.16)

This data represents the full sample. Standard errors are in parentheses.

Although these raw results are inconclusive, there are a couple of reasons to explore heterogenous treatment effects. First, important differences that impact image motivation likely exist between above and below average volunteers, even though these reasons are not captured in this paper's theoretical predictions. It is plausible that participants with below average reputations are less image-motivated than participants with above average reputations. It is similarly likely that a public below average reputation is seen less favorably than a public above average reputation. Second, the treatment effects may hold more strongly among females if they are more image-concerned or susceptible to influences in the lab, both of which are common gender results noted in Section 1.

Via examining regressions of the specifications below, a deeper investigation of the data provides strong support for these heterogenous treatment effects by volunteer reputation and gender. While regressions results of Equation 6 will directly test all three predictions: $\beta_1 < 0$ (the Negative Image Effect), $\beta_2 < 0$ (the Reputations Effect), and $\beta_1 + \beta_3 = 0$ (the Interactions Effect), the regression results of Equation 7 will allow for heterogenous treatment effects between above and below average volunteers. Heterogeneous effects by gender will be examined via running regressions on samples restricted to only males or females.

$$clicks_i = \beta_0 + \beta_1 O_{pubi} + \beta_2 R_{pubi} + \beta_3 O_{pub} R_{pubi} + [Controls]_i + \epsilon_i \tag{6}$$

$$\begin{aligned} clicks_i = {} & \beta_0 + \beta_1 O_{pubi} + \beta_2 R_{pubi} + \beta_3 O_{pub} R_{pubi} \\ & + v^b \left(\beta_4 + \beta_5 O_{pubi} + \beta_6 R_{pubi} + \beta_7 O_{pub} R_{pubi}\right) + [Controls]_i + \epsilon_i \end{aligned} \tag{7}$$

where

$clicks_i$ = number of clicks participant i completes in the clicking task

O_{pubi} = 1 if participant i's incentive offer is public, 0 otherwise

R_{pubi} = 1 if participant i's volunteer reputation is public, 0 otherwise

$O_{pub}R_{pubi}$ = 1 if participant i's incentive offer and volunteer reputation are public, 0 otherwise

v_i^b = 1 if participant i has a below average volunteer reputation, 0 otherwise

Column 1 of Table 4.2 confirms what is already known from the raw data - there is inconclusive evidence for the treatment effects among the full sample, and this remains true even when allowing for heterogeneity

among above and below average volunteers as shown in Column 2. However, Columns 3 and 4 support the possibility that females are more susceptible to influences in the laboratory or more concerned about their image than males. While there is inconclusive evidence among males, there is strong evidence for these treatment effects among females.

The females' results support the expected heterogeneity between above and below average volunteers. Among females with below average reputations, the observability of the incentive offers or reputations do not significantly impact how much one clicks.[27] However, among females with above average volunteer reputations, there is significant evidence for all three treatment effects as detailed in Results 1F-3F.

Result 1F [The Negative Image Effect] Among females with above average reputations, public incentive offers discourage volunteering by 194 clicks, on average, or by 9.1% of the average number of clicks by females.

Result 2F [The Reputations Effect] Among females with above average reputations, public volunteer reputations discourage volunteering by 198 clicks, on average, or by 9.2% of the average clicks by females.

Result 3F [The Interactions Effect] Among females with above average reputations, the observability of the incentive offer has no significant impact on volunteer behavior. That is, I cannot reject the hypothesis that public reputations cancel out the effect of public incentive offers.[28] If anything, it appears that public reputations more than counteract the negative impact of public incentive offers.

It is worth noting that the dictator games following the volunteer clicking task lend support to (although do not causally prove) the reasons put forth for heterogenous treatment effects between females with above and below average reputations. First, results from the"CPs' Donations to the ARC" dictator game provides evidence that females with below average reputations are less image motivated than females with above average reputations, since the former publicly donate less to the American Red Cross (see Appendix A.2). Second, results from the "PMs' Allocations to the CPs" dictator game correspond to public below average reputations being perceived less favorably than public above average reputation, for PMs allocate relatively less to CPs with below average reputations (see Appendix A.3).

Finally, the results are robust to several specifications. First, the results are robust to running separate regressions according to participants' genders and volunteer reputations, as shown in Appendix B.4 and Appendix B.5. Second, the results are qualitatively and quantitatively similar without controls as shown in Appendix B.6. Third, the results remain similar even with the inclusion of session fixed effects as shown in Appendix B.7, although note that O_{pub} must be excluded form these regressions since the the observability of the incentive offer was only varied on the session level. Fourth, and perhaps most encouragingly as shown in Appendix B.8, the treatment effects, although smaller in magnitude, still hold with the inclusion of a control for an individual's clicking speed: $r_2clicks$ is a CP's number of clicks in Round 2 where she clicks for one minute and only earns money for herself. More descriptive statistics on $r_2clicks$, shown in Appendix B.3, also suggest that neither males nor females are reaching their maximum number of potential clicks in the clicking task.[29] Lastly, note that Appendix B.9 presents all of these different specifications for females with above average reputations.

[27] Using Equation 7, Negative Image Effect- $H_0 : \beta_1 + \beta_5 = 0$, p-value = 0.61; Reputations Effect- $H_0 : \beta_2 + \beta_6 = 0$, p-value = 0.49; Interactions Effect- $H_0 : \beta_1 + \beta_5 + \beta_3 + \beta_7 = 0$, p-value = 0.20.

[28] Using Equation 7, $H_0 : \beta_1 + \beta_3 = 0$, p-value = 0.24.

[29] If participants maintain the same rate of clicking in the one minute Round 2 as the eight minute clicking task, then $r_2clicks/clicks = 8$. However, for both females and males, $r_2clicks/clicks < 8$ at the 90th percentile. It is also unlikely that CPs max out on their clicks in Round 2 since the one minute time limit of this round significantly limits potential amount of money earned. This evidence suggests that CPs could have clicked more in the clicking task.

Table 4.2: The Effects on Clicks

	Full Sample	Full Sample	Males	Females
O_{pub}	-65.99	-56.55	48.83	-193.56**
	(58.54)	(73.64)	(119.86)	(93.36)
R_{pub}	-79.09	-76.57	8.92	-197.69**
	(55.31)	(67.37)	(102.30)	(87.83)
$O_{pub}R_{pub}$	45.95	71.35	-111.98	284.76**
	(81.32)	(97.14)	(158.63)	(118.88)
$v^b O_{pub}$		-20.17	-216.57	108.03
		(127.79)	(184.76)	(188.68)
$v^b R_{pub}$		2.65	-167.75	284.47*
		(118.47)	(184.48)	(152.92)
$v^b O_{pub} R_{pub}$		-139.26	188.56	-513.29
		(191.51)	(280.06)	(312.82)
v^b	-28.62	5.03	142.77	-187.35
	(53.31)	(85.65)	(125.95)	(112.80)
male	118.85***	126.62***		
	(41.77)	(42.63)		
constant	2198.42***	2195.26***	2221.65***	2310.11***
	(91.71)	(93.95)	(123.76)	(134.48)
Controls	yes	yes	yes	yes
N	130	130	67	63

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Results are from an OLS regression with a dependent variable equal to a CP's number of clicks. v^b indicates that someone is a below average volunteer. The following controls are included: indicator for willingness to complete an additional survey or not; number of stated volunteer hours in past year; indicator for economics major; indicator for the CP being born in the US; and, indicator for whether the CP feels favorably about ARC or not.

5. Conclusions and Discussion

There are several reasons as to why incentives for prosocial behavior may succeed or backfire and a correspondingly mixed literature on this topic. To provide needed insight into when incentives for prosocial behavior are likely effective or not, this paper focuses on how incentives interact with image motivation. In particular, this paper examines three theoretical predictions derived from an extension of the model of prosocial behavior in Bénabou and Tirole (2006): the Negative Image Effect- public incentives crowd out image motivation to volunteer due to the associated greedy signal; the Reputations Effect- public reputations crowd out image motivation to volunteer because they weaken the strength of the prosocial signal; and, the Interactions Effect- public incentive offers do not crowd out image motivation to volunteer among those with public reputations because those reputations weaken the strength of the greedy signal.

In experimentally testings these three predictions, this study finds heterogeneous effects by types of reputations and gender. In particular, these image-related effects do not hold among males, who are likely less image motivated or susceptible to influences in the lab, nor do they hold among individuals with below average volunteer reputations that are likely perceived unfavorably. However, this study finds strong support of the three predictions among females with above average reputations: public incentives and public reputations both independently decrease the level of volunteering by about nine-percent, but among those with public reputations, the observability of incentive offers has no significant impact. In other words, this paper provides support for an important theoretical implication that helps to unify the conflicting literature: public incentives to volunteer are likely more effective for those with public, rather than private, reputations.

Extending past these three results, I have several suggestions for future work. First, while publicly recognizing volunteers may be very effective at encouraging more volunteerng in the short run, this public recognition establishes volunteers' reputations and hence may be ineffective in the long run because of the Reputations Effect. Examining the robustness of the Reputations Effect in different prosocial behavior environments may therefore be very informative for policy. Second, diminishing returns to material and image incentives on volunteer behavior may explain why some studies find incentives to be ineffective at encouraging public volunteer behavior, and more importantly, that increasing a material incentive to outweigh any associated crowding out effects may be ineffective.[30]

[30]The effect of incentives on charitable giving has been extensively studied via estimating price elasticities of giving, with Andreoni (2006) providing a good overview of this literature, and Karlan and List (2007) detailing an insightful field experiment. However, there do not exist similar theoretical or empirical studies, to my knowledge, that examine how incentives interact with volunteer labor supply. Also, note that the results in this study cannot be explained by diminishing returns to incentives since the prosocial image and material incentives are constant across all treatment groups. However, this is a possible explanation for studies that examine the the effect of a greedy image without holding the other the prosocial image incentive constant.

Acknowledgements

I gratefully acknowledge funding for this study from the NSF (SES 1159032), Stanford's Institute for Research in the Social Sciences, and Stanford's Economics department. I would like to thank Ran Abramitzky, Doug Bernheim, John Beshears, Nick Bloom, Loy Campbell, Jeffrey Naecker, Muriel Niederle, Charlie Sprenger, and Stephen Terry for their helpful advice and support. I also am appreciative of feedback from participants in Stanford's Behavioral and Experimental Workshop, the 2012 North American Economic Science Association Conference in Tucson, and the 2012 Bay Area Behavioral and Experimental Economics Workshop.

References

Andreoni, J. (2006). Chapter 18 philanthropy. In S.-C. Kolm and J. M. Ythier (Eds.), *Applications*, Volume 2 of *Handbook on the Economics of Giving, Reciprocity and Altruism*, pp. 1201 – 1269. Elsevier.

Andreoni, J. and B. D. Bernheim (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica 77*(5), 1607–1636.

Andreoni, J. and L. Vesterlund (2001). Which is the fair sex? Gender differences in altruism. *Quarterly Journal of Economics*, 293–312.

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), 544–555.

Ashaaf, N., O. Bandiera, and K. Jack (2012, June). No margin, no mission? A field experiment on incentives for pro-social tasks. *Harvard Business School Working Paper 12-008*.

Bénabou, R. and J. Tirole (2003). Intrinsic and extrinsic motivation. *Review of Economic Studies 70*(3), 489–520.

Bénabou, R. and J. Tirole (2006). Incentives and prosocial behavior. *American Economic Review 96*(5), 1652–1678.

Bernheim, B. D. (1994). A theory of conformity. *Journal of Political Economy 102*(5), pp. 841–877.

Carpenter, J. and C. K. Myers (2010). Why volunteer? Evidence on the role of altruism, image, and incentives. *Journal of Public Economics 94*(11-12), 911 – 920.

Corporation for National and Community Service (2011). *Volunteering in America 2011 Research Highlights*, Washington DC. Corporation for National and Community Service.

Croson, R. and U. Gneezy (2009, June). Gender differences in preferences. *Journal of Economic Literature 47*(2), 448–474.

DellaVigna, S., J. A. List, U. Malmendier, and G. Rao (2013, February). The importance of being marginal: Gender differences in generosity. Technical report, NBER Working Paper No. 18748.

Frey, B. and R. Jergen (2001). Motivaiton of crowding theory. *Journal of Economic Surveys 15*(5), 589–611.

Frey, B. S. and F. Oberholzer-Gee (1997). The cost of price incentives: An empirical analysis of motivation crowding- out. *The American Economic Review 87*(4), pp. 746–755.

Gneezy, U. (2005). Deception: The role of consequences. *The American Economic Review 95*(1), pp. 384–394.

Gneezy, U., S. Meier, and P. Rey-Biel (2011, Fall). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives 25*(4), 191–210.

Gneezy, U. and A. Rustichini (2000). Pay enough or don't pay at all. *Quarterly Journal of Economics 115*(3), 791–810.

Goette, L. F. and A. Stutzer (2008). Blood Donations and Incentives: Evidence from a Field Experiment. *SSRN eLibrary*.

Harbaugh, W. T. (1998a). The prestige motive for making charitable transfers. *The American Economic Review 88*(2), pp. 277–282.

Harbaugh, W. T. (1998b). What do donations buy?: A model of philanthropy based on prestige and warm glow. *Journal of Public Economics 67*(2), 269 – 284.

Independent Sector (2010). *Value of Volunteer Time*. Independent Sector.

Karlan, D. and J. A. List (2007). Does price matter in charitable giving? Evidence from a large-scale natural field experiment. *The American Economic Review 97*(5), pp. 1774–1793.

Lacetera, N. and M. Macis (2010a). Do all material incentives for pro-social activities backfire? The response to cash and non-cash incentives for blood donations. *Journal of Economic Psychology 31*(4), 738 – 748.

Lacetera, N. and M. Macis (2010b). Social image concerns and prosocial behavior: Field evidence from a nonlinear incentive scheme. *Journal of Economic Behavior and Organization 76*(2), 225 – 237.

Lacetera, N., M. Macis, and R. Slonim (2011). Rewarding altruism? A natural field experiment. Working paper, National Bureau of Economic Research.

Lacetera, N., M. Macis, and R. Slonim (2012). Will there be blood? Incentives and displacement effects in pro-social behavior. *American Economic Journal: Economic Policy 4*(1), 186–223.

Mellström, C. and M. Johannesson (2008). Crowding out in blood donation: Was titmuss right? *Journal of the European Economic Association 6*(4), 845–863.

Niederle, M. and L. Vesterlund (2011). Gender and competition. *Annual Review of Economics*, 601–630.

Titmuss, R. M. (1970). *The Gift Relationship*. London: Allen and Unwin.

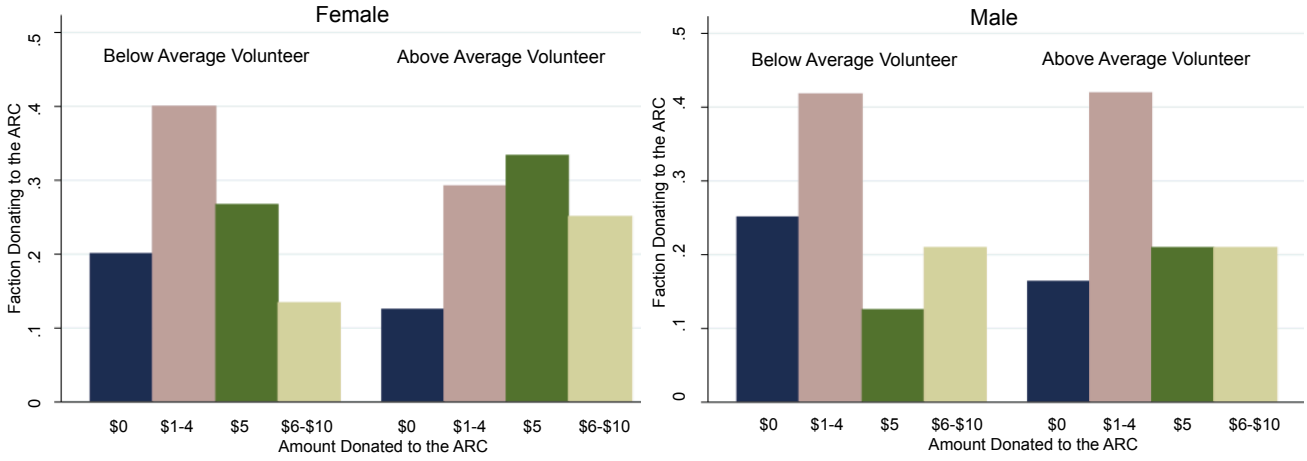
Appendices

Appendix A. Figures

Figure Appendix A.1: Electronic Tally Counter

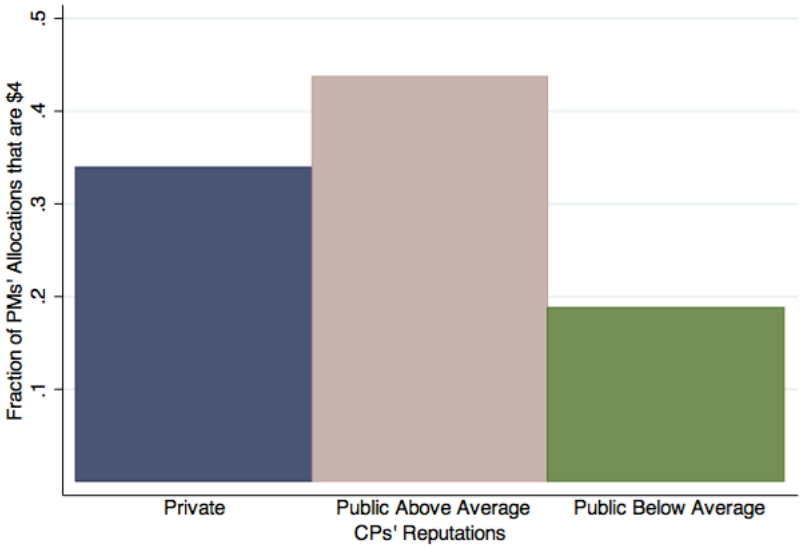
This Electronic Tally Counter, manufactured by Reddington, is both highly rated on Amazon.com and has a simple design suitable for this experiment. See http://www.amazon.com/Reddington-Electronic-Tally-Counter/dp/B0019VH9S0/ref=cm_cr_pr_product_top for more information.

Figure Appendix A.2: By Sex and Volunteer Reputation: CPs' Public Donations to ARC



This graph displays results from the "CPs' Donation to the ARC" dictator game, where CPs chose what dollar amount they would like to publicly donate to the ARC out of a potential $10 in additional compensation.

Figure Appendix A.3: PMs' Allocations to CPs



This graph displays results from the "PMs' Allocations to CPs" dictator game, where PMs decide how much to privately allocate, $2 or $4, out of $10, to each of their CPs. These results involve 248 allocation decisions from 31 PMs, since 3 PMs are excluded for not accurately recording information on their CPs.

Appendix B. Tables

Table Appendix B.1: By Observability of Incentive Offer: Characteristics of Participants

	Full Subsample			Subsample (1)	
	All	O_{pub}	O_{prvt}	O_{pub}	O_{prvt}
Above Average Volunteer (1)	0.70	0.75	0.66	1.00	1.00
Willingness to Complete Additional Survey	0.74	0.76	0.72	0.84	0.70
Male	0.52	0.53	0.51	0.43	0.51
Economics Major	0.12	0.10	0.13	0.05	0.06
Born in United States	0.78	0.85	0.73	0.89	0.70
Feel Favorably about ARC	0.77	0.81	0.73	0.84	0.72
Senior	0.10	0.14	0.07	0.14	0.09
Junior	0.12	0.12	0.11	0.14	0.09
Sophomore	0.19	0.15	0.23	0.05	0.19
Freshman	0.58	0.56	0.59	0.64	0.64
Volunteer Hours	70.90	81.49	62.10	106.00	91.57
N	130	59	71	44	47

All of the above values indicate the fraction of participants with a given characteristic, except for the values associated with Volunteer Hours which indicate the average volunteer hours for participants. For the full subsample, I cannot reject that joint hypothesis that the means of the above variables are different, or the individual hypotheses that the mean of any of the above variables are different, across the O_{pub} and O_{prvt} subsamples.

Table Appendix B.2: By Observability of Reputations: Characteristics of Participants

	Full Subsample			Subsample (1)	
	All	R_{pub}	R_{prvt}	R_{pub}	R_{prvt}
Above Average Volunteer (1)	0.70	0.75	0.64	1.00	1.00
Willingness to Complete Additional Survey	0.74	0.70	0.79	0.75	0.79
Male	0.52	0.49	0.54	0.46	0.49
Economics Major	0.12	0.12	0.11	0.06	0.05
Born in the United States	0.78	0.80	0.77	0.81	0.77
Feel Favorably about ARC	0.77	0.75	0.79	0.77	0.79
Senior	0.10	0.13	0.07	0.13	0.08
Junior	0.12	0.12	0.11	0.10	0.13
Sophomore	0.19	0.17	0.21	0.13	0.10
Freshman	0.58	0.55	0.61	0.60	0.69
Volunteer Hours	70.90	75.32	65.90	97.77	99.59
N	130	69	61	69	61

All of the above values indicate the fraction of participants with a given characteristic, except for the values associated with Volunteer Hours which indicate the average volunteer hours for participants. For the full subsample, I cannot reject that joint hypothesis that the means of the above variables are different, or the individual hypotheses that the mean of any of the above variables are different, across the O_{pub} and O_{prvt} subsamples.

Table Appendix B.3: Clicking Descriptive Stats

	Females					Males				
	mean	p50	p90	min	max	mean	p50	p90	min	max
$clicks$	2136.11	2155	2352	1545	2560	2264.85	2258	2635	1500	2800
$r_2clicks$	313.46	311	348	209	506	319.64	316	363	192	425
$clicks/r_2clicks$	6.89	6.92	7.61	4.36	9.10	7.11	7.08	7.63	5.36	9.45
N	63					67				

mean, p50, p90, min and max indicate the mean, 50th percentile, 90th percentile, minimum and maximum number of clicks. While clicks is the number of clicks in the eight minute clicking task, $r_2clicks$ in the number of clicks in the one minute Round 2. Recall that CPs clicked in Round 2 to earn money for themselves (but not the ARC), and their number of clicks was kept private.

Table Appendix B.4: By Sex and Volunteer Reputation: The Effects on Clicks

	Females		Males	
Vol. Reputation:	Above Avg	Below Avg	Above Avg	Below Avg
O_{pub}	-156.54*	-40.83	6.98	-90.57
	(83.63)	(206.99)	(115.15)	(132.46)
R_{pub}	-169.52**	24.67	-14.81	-107.40
	(79.49)	(146.37)	(101.52)	(145.10)
$O_{pub}R_{pub}$	225.32**	-180.17	-51.89	-85.23
	(109.27)	(343.26)	(153.51)	(205.20)
constant	2253.44***	2136.33***	2270.27***	2396.00***
	(62.02)	(103.50)	(74.72)	(93.66)
Controls	no	no	no	no
N	48	15	43	24

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Using the specification from Equation 6, results are from an OLS regression with a dependent variable equal to a CP's number of clicks.

Table Appendix B.5: By Sex and Volunteer Reputation: The Effects on Clicks

	Female		Male	
Vol. Reputation:	Above Avg	Below Avg	Above Avg	Below Avg
O_{pub}	-190.82**	-181.54	41.77	-69.28
	(87.75)	(283.24)	(122.02)	(126.73)
R_{pub}	-194.32**	-50.66	-3.01	-71.64
	(83.11)	(228.15)	(103.70)	(137.39)
$O_{pub}R_{pub}$	285.50**	553.52	-70.28	61.02
	(111.96)	(929.65)	(162.45)	(223.09)
constant	2288.16***	2203.69*	2172.75***	2252.86***
	(145.54)	(858.44)	(146.88)	(201.98)
Controls	yes	yes	yes	yes
N	48	15	43	24

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Using the specification from Equation 6, results are from an OLS regression with a dependent variable equal to a CP's number of clicks. The following controls are included: indicator for willingness to complete an additional survey or not; number of stated volunteer hours in past year; indicator for economics majors; indicator for the CP being born in the US; and, indicator for whether the CP feels favorably about ARC or not.

Table Appendix B.6: No Controls: The Effects on Clicks

	Full Sample	Full Sample	Males	Females
O_{pub}	-66.01	-72.76	6.98	-156.54*
	(57.42)	(71.75)	(115.15)	(90.51)
R_{pub}	-76.83	-87.09	-14.81	-169.52*
	(53.34)	(65.96)	(101.52)	(86.03)
$O_{pub}R_{pub}$	39.30	89.90	-51.89	225.32*
	(78.81)	(94.75)	(153.51)	(118.26)
$v^b O_{pub}$		23.64	-97.55	115.70
		(121.34)	(175.51)	(187.68)
$v^b R_{pub}$		41.07	-92.59	194.18
		(112.78)	(177.09)	(144.63)
$v^b O_{pub} R_{pub}$		-226.02	-33.34	-405.49
		(176.65)	(256.27)	(297.20)
v^b	4.13	14.98	125.73	-117.11
	(43.60)	(79.58)	(119.82)	(106.13)
male	125.72***	132.62***		
	(39.56)	(40.23)		
constant	2197.79***	2189.76***	2270.27***	2253.44***
	(46.63)	(54.63)	(74.72)	(67.12)
Controls	no	no	no	no
N	130	130	67	63

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Results are from an OLS regression with a dependent variable equal to a CP's number of clicks. v^b indicates that someone is a below average volunteer.

Table Appendix B.7: Controls & Session Fixed Effects: The Effects on Clicks

	Full Sample	Full Sample	Males	Females
R_{pub}	-89.13	-78.89	-30.03	-168.92*
	(55.86)	(65.01)	(103.08)	(88.51)
$O_{pub}R_{pub}$	63.95	85.04	-40.30	256.12**
	(82.60)	(88.55)	(152.88)	(115.23)
$v^b O_{pub}$				
$v^b R_{pub}$		-21.64	-94.20	146.18
		(106.96)	(167.63)	(161.78)
$v^b O_{pub} R_{pub}$		-123.55	90.83	-284.87
		(143.69)	(228.50)	(296.03)
v^b	-26.52	7.99	63.86	-133.15
	(53.98)	(69.25)	(107.79)	(101.80)
male	105.34**	113.18**		
	(43.60)	(44.57)		
constant	2167.34***	2171.37***	2170.91***	2167.74***
	(109.37)	(110.14)	(168.41)	(163.01)
Controls	yes	yes	yes	yes
Session FE	yes	yes	yes	yes
N	130	130	67	63

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Results are from an OLS regression with a dependent variable equal to a CP's number of clicks. v^b indicates that someone is a below average volunteer. The following controls are included: indicator for willingness to complete an additional survey or not; number of stated volunteer hours in past year; indicator for economics major; indicator for the CP being born in the US; and, indicator for whether the CP feels favorably about ARC or not.

Table Appendix B.8: Controls & Session Fixed & $r_2Clicks$: The Effects on Clicks

	Full Sample	Full Sample	Males	Females
R_{pub}	-65.43	-67.83	-19.20	-122.29*
	(43.09)	(50.09)	(73.09)	(68.08)
$O_{pub}R_{pub}$	26.61	48.32	-35.70	161.47*
	(63.72)	(68.34)	(108.38)	(89.56)
v^bO_{pub}				
v^bR_{pub}		15.97	-17.17	110.52
		(82.51)	(119.36)	(123.67)
$v^bO_{pub}R_{pub}$		-110.82	49.52	-282.59
		(110.70)	(162.10)	(226.00)
v^b	-9.27	4.94	49.43	-90.20
	(41.60)	(53.35)	(76.44)	(78.09)
male	102.46***	109.28***		
	(33.56)	(34.34)		
$r_2Clicks$	3.69***	3.68***	5.17***	2.83***
	(0.42)	(0.42)	(0.75)	(0.50)
constant	1084.17***	1094.37***	676.90***	1347.18***
	(149.19)	(150.09)	(247.44)	(191.51)
Controls	yes	yes	yes	yes
Session FE	yes	yes	yes	yes
N	130	130	67	63

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Results are from an OLS regression with a dependent variable equal to a CP's number of clicks. v^b indicates that someone is a below average volunteer. The following controls are included: indicator for willingness to complete an additional survey or not; number of stated volunteer hours in past year; indicator for economics major; indicator for the CP being born in the US; and, indicator for whether the CP feels favorably about ARC or not.

Table Appendix B.9: For Females: The Effects on Clicks

	1	2	3	4
O_{pub}	-156.54*	-193.56**		
	(90.51)	(93.36)		
R_{pub}	-169.52*	-197.69**	-168.92*	-122.29*
	(86.03)	(87.83)	(88.51)	(68.08)
$O_{pub}R_{pub}$	225.32*	284.76**	256.12**	161.47*
	(118.26)	(118.88)	(115.23)	(89.56)
$v^b O_{pub}$	115.70	108.03		
	(187.68)	(188.68)		
$v^b R_{pub}$	194.18	284.47*	146.18	110.52
	(144.63)	(152.92)	(161.78)	(123.67)
$v^b R_{pub} O_{pub}$	-405.49	-513.29	-284.87	-282.59
	(297.20)	(312.82)	(296.03)	(226.00)
v^b	-117.11	-187.35	-133.15	-90.20
	(106.13)	(112.80)	(101.80)	(78.09)
$r_2 Clicks$				2.83***
				(0.50)
constant	2253.44***	2310.11***	2167.74***	1347.18***
	(67.12)	(134.48)	(163.01)	(191.51)
Controls	no	yes	yes	yes
Session FE	no	no	yes	yes
N	63	63	63	63

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Using only female CPs, results are from an OLS regression with a dependent variable equal to a CP's number of clicks. v^b indicates that someone is a below average volunteer. In addition to any shown in the table, the following controls, when indicated, are included: indicator for willingness to complete an additional survey or not; number of stated volunteer hours in past year; indicator for economics majors; indicator for the CP being born in the US; and, indicator for whether the CP feels favorably about ARC or not.

Appendix C. Theory

I solve the prosocial behavior model in Bénabou and Tirole (2006) and a slightly modified version under four different sets of conditions, one for each of four treatment groups in this study. In Section Appendix C.1, I first solve the model of prosocial behavior in Bénabou and Tirole (2006), under their baseline conditions (although using the terminology of this paper) where an agent's prosocial activity is public (A_{pub}), incentive offer is public (O_{pub}), and reputation is private (R_{prvt}). In Section Appendix C.2, I secondly solve a slightly modified version of their model under the conditions where an agent's prosocial activity is public (A_{pub}), incentive offer is private (O_{prvt}), and reputation is private (R_{prvt}). In Section Appendix C.3, after introducing the concept of public reputations (R_{pub}), I then solve the model a third and a fourth time under the conditions $A_{pub}O_{pub}R_{pub}$ and $A_{pub}O_{prvt}R_{pub}$.

With the above models solved, I can then rather easily solve for the three testable predictions, as shown in Section Appendix C.4.

Appendix C.1. Solving Under Conditions: $A_{pub}O_{pub}R_{prvt}$

In Bénabou and Tirole (2006), henceforth BT, agents receive utility from volunteering, or participating in a prosocial activity, at a level of a that is incentivized at a rate of y, according to their intrinsic, extrinsic, and image motivations as well as their private cost of volunteering. An agent's intrinsic motivation involves her "prosocial" preference for how much she values volunteering, v_a. An an agent's extrinsic motivation results from the incentive rate y and her "greedy" preference for how much she values money, v_y. Her volunteer behavior is public (A_{pub}) and her incentive offer is public (O_{pub}), so her image motivation or payoff, $I(a, y)$, is a function of her chosen level of volunteer behavior a and incentive rate y. In particular, $I(a, y)$ is the combination of how prosocial or greedy others believe the agent is based off of her incentivized volunteer behavior, $\mathbb{E}(v_a|a, y)$ and $\mathbb{E}(v_y|a, y)$, weighted by her preference to appear prosocial and not greedy to others, μ_a and μ_y. Combining these three motivations with the private cost of volunteering, $C(a)$, yields the following utility for an agent:

$$U(a, y) = (v_a + yv_y)\, a + I(a, y) - C(a) \tag{C.1}$$

where

$$I(a, y) = \mu_a \mathbb{E}(v_a|a, y) - \mu_y \mathbb{E}(v_y|a, y) \tag{C.2}$$

To solve the model, there are three helpful assumptions. First, I assume a convex cost function $C(a) = \frac{ka^2}{2}$ for some fixed parameter k. Second, I assume that all agents have the same non-negative image preferences, $\mu = (\mu_a, \mu_y) = (\bar{\mu}_a, \bar{\mu}_y)$. Third, I assume that agents' types, (v_a, v_y) are normally distributed as follows:

$$\begin{pmatrix} v_a \\ v_y \end{pmatrix} \sim \mathbf{N}\left(\begin{pmatrix} \bar{v}_a \\ \bar{v}_y \end{pmatrix}, \begin{pmatrix} \sigma_a^2 & \sigma_{ay} \\ \sigma_{ay} & \sigma_y^2 \end{pmatrix} \right).$$

$$\bar{v}_y > 0$$

Now, I will solve for the optimal level of prosocial activity or volunteering, a, based on the preferences in Equation C.1, via the steps broken down by the following subsections.

Appendix C.1.1. Write the first order condition of Equation C.1

Differentiating Equation (C.1) with respect to a yields

$$\underbrace{C'(a)}_{=ka} = (v_a + yv_y) + \underbrace{\mu_a \frac{\partial \mathbb{E}(v_a|a,y)}{\partial a} - \mu_y \frac{\partial \mathbb{E}(v_y|a,y)}{\partial a}}_{i(a,y)} \tag{C.3}$$

Appendix C.1.2. Solve for $\mathbb{E}(v_a|a,y)$ in terms of the known values of a, y and parameters of the distribution of $\mathbf{v} = (v_a, v_y)$, and the unknown value of $i(a,y)$.

Equation (C.3) implies

$$v_a + yv_y = C'(a) - i(a,y),$$

so by subtracting $\bar{v}_a + y\bar{v}_y$ from both sides, we have

$$\underbrace{(v_a - \bar{v}_a)}_{\text{signal}} + \underbrace{y(v_y - \bar{v}_y)}_{\text{noise}} = \underbrace{C'(a) - i(a,y) - \bar{v}_a - y\bar{v}_y}_{\mathcal{Z}\ \text{(observed)}}$$

and from a standard signal extraction solution this implies that:

$$\begin{aligned}
\mathbb{E}((v_a - \bar{v}_a)|\mathcal{Z}) &= \left[\mathbb{E}((v_a - \bar{v}_a)\mathcal{Z})\right] \left[var(\mathcal{Z})\right]^{-1} \mathcal{Z} \\
\mathbb{E}((v_a - \bar{v}_a)|\mathcal{Z}) &= \frac{cov(v_a - \bar{v}_a, \mathcal{Z})}{var(\mathcal{Z})} \mathcal{Z}
\end{aligned} \tag{C.4}$$

where the $cov(v_a - \bar{v}_a, \mathcal{Z})$ and the $var(\mathcal{Z})$ are as follows.

$$\begin{aligned}
cov((v_a - \bar{v}_a), \mathcal{Z}) &= cov((v_a - \bar{v}_a), v_a - \bar{v}_a + y(v_y - \bar{v}_y)) \\
&= cov((v_a - \bar{v}_a), (v_a - \bar{v}_a)) + ycov((v_a - \bar{v}_a), (v_y - \bar{v}_y)) \\
&= var(v_a - \bar{v}_a) + y \left[\mathbb{E}\left((v_a - \bar{v}_a)(v_y - \bar{v}_y)\right) - \underbrace{\mathbb{E}(v_a - \bar{v}_a)\mathbb{E}(v_y - \bar{v}_y)}_{=0} \right] \\
&= var(v_a) + y\left[\mathbb{E}\left((v_a - \bar{v}_a)(v_y - \bar{v}_y)\right)\right] \\
cov((v_a - \bar{v}_a), \mathcal{Z}) &= \sigma_a^2 + y\sigma_{ay}
\end{aligned} \tag{C.5}$$

$$\begin{aligned}
var(\mathcal{Z}) &= var(v_a - \bar{v}_a + y(v_y - \bar{v}_y)) \\
&= var(v_a + yv_y) \\
&= var\left(\begin{bmatrix} 1 & y \end{bmatrix} \begin{bmatrix} v_a \\ v_y \end{bmatrix}\right) \\
&= \begin{bmatrix} 1 & y \end{bmatrix} \begin{bmatrix} \sigma_a^2 & \sigma_{ay} \\ \sigma_{ay} & \sigma_y^2 \end{bmatrix} \begin{bmatrix} 1 \\ y \end{bmatrix} \\
var(\mathcal{Z}) &= \sigma_a^2 + 2y\sigma_{ay} + y^2\sigma_y^2
\end{aligned} \tag{C.6}$$

Now, plugging in Equations (C.5) and (C.6) into Equation (C.4), we get

$$\begin{aligned}
\mathbb{E}((v_a - \bar{v}_a)|\mathcal{Z}) &= \frac{\sigma_a + y\sigma_{ay}}{\sigma_a^2 + 2y\sigma_{ay} + y^2\sigma_y^2}\mathcal{Z} \\
\mathbb{E}(v_a|\mathcal{Z}) &= \bar{v}_a + \frac{\sigma_a^2 + y\sigma_{ay}}{\sigma_a^2 + 2y\sigma_{ay} + y^2\sigma_y^2}\left(C'(a) - i(a,y) - \bar{v}_a - y\bar{v}_y\right) \\
\mathbb{E}(v_a|\mathcal{Z}) &= \bar{v}_a + \rho(y)\left(ka - i(a,y) - \bar{v}_a - y\bar{v}_y\right) \qquad \text{(C.7)}
\end{aligned}$$

where

$$\rho(y) = \frac{\sigma_a^2 + y\sigma_{ay}}{\sigma_a^2 + 2y\sigma_{ay} + y^2\sigma_y^2} \qquad \text{(C.8)}$$

Appendix C.1.3. Solve for $\mathbb{E}(v_y|a,y)$ in terms of the known values of a, y and parameters of the distribution of $\mathbf{v} = (v_a, v_y)$, and the unknown value of $i(a,y)$.

Equation (C.3) implies

$$v_a + yv_y = C'(a) - i(a,y)$$

so by subtracting $\bar{v}_a + y\bar{v}_y$ from both sides, we have

$$\begin{aligned}
(v_y - \bar{v}_y) + \frac{1}{y}(v_a - \bar{v}_a) &= \frac{1}{y}\left(C'(a) - i(a,y) - \bar{v}_a - y\bar{v}_y\right) \\
(v_y - \bar{v}_y) &= \frac{1}{y}\left(-(v_a - \bar{v}_a) + (C'(a) - i(a,y) - \bar{v}_a - y\bar{v}_y)\right) \\
\mathbb{E}((v_y - \bar{v}_y)|\mathcal{Z}) &= \frac{1}{y}\left(-\mathbb{E}((v_a - \bar{v}_a)|\mathcal{Z}) + (C'(a) - i(a,y) - \bar{v}_a - y\bar{v}_y)\right) \\
\mathbb{E}((v_y - \bar{v}_y)|\mathcal{Z}) &= \frac{1}{y}\left(-(\rho(y)\left(ka - i(a,y) - \bar{v}_a - y\bar{v}_y\right)) + (ka - i(a,y) - \bar{v}_a - y\bar{v}_y)\right) \\
\mathbb{E}((v_y - \bar{v}_y)|\mathcal{Z}) &= \frac{1}{y}(1 - \rho(y))\left(ka - i(a,y) - \bar{v}_a - y\bar{v}_y\right) \\
\mathbb{E}(v_y|\mathcal{Z}) &= \bar{v}_y + \frac{1}{y}(1 - \rho(y))\left(ka - i(a,y) - \bar{v}_a - y\bar{v}_y\right) \\
\mathbb{E}(v_y|\mathcal{Z}) &= \bar{v}_y + \chi(y)\left(ka - i(a,y) - \bar{v}_a - y\bar{v}_y\right) \qquad \text{(C.9)}
\end{aligned}$$

where

$$\chi(y) = \frac{1 - \rho(y)}{y} \qquad \text{(C.10)}$$

Appendix C.1.4. Express i(a, y) in a differential equation and then solve for i(a, y), the only unknown value in the above conditional expectations.

Equation (C.3) implies

$$i(a,y) = \mu_a \frac{\partial \mathbb{E}(v_a|a,y)}{\partial a} - \mu_y \frac{\partial \mathbb{E}(v_y|a,y)}{\partial a}$$

Then, let y be fixed so

$$i(a,y) = \bar{i}(a) = \mu_a \frac{\partial \mathbb{E}(v_a|a,y)}{\partial a} - \mu_y \frac{\partial \mathbb{E}(v_y|a,y)}{\partial a} \qquad \text{(C.11)}$$

Differentiating Equations (C.7) and (C.9) with respect to a yields:

$$\frac{\partial \mathbb{E}(v_a|\mathcal{Z})}{\partial a} = \rho(y)\,(k - \bar{i}'(a),) \tag{C.12}$$

and

$$\frac{\partial \mathbb{E}(v_y|\mathcal{Z})}{\partial a} = \chi(y)\,(k - \bar{i}'(a))\,. \tag{C.13}$$

Then, plugging in Equations (C.12) and (C.13) into Equation (C.11) yields the following differential equation:

$$\begin{aligned} \bar{i}(a) &= (\mu_a \rho(y) - \mu_y \chi(y))(k - \bar{i}'(a)) \\ \bar{i}(a) &= \lambda(k - \bar{i}'(a)) \end{aligned} \tag{C.14}$$

where

$$\lambda = \mu_a \rho(y) - \mu_y \chi(y) \tag{C.15}$$

The differential equation in Equation (C.14), can be generally solved as follows:

$$\bar{i}(a) = k(\lambda + \xi e^{\frac{-a}{\lambda}}) \tag{C.16}$$

for some constant ξ. You can confirm that Equation (C.16) is the correct solution, by substituting it and its derivative $(-k\xi\frac{1}{\lambda}e^{\frac{-a}{\lambda}})$ back into the differential equation in Equation (C.14). Now, to solve for the constant ξ, simply note that since the exponential term would eventually dominate all others in the objective function, the only value of ξ generating finite solutions is $\xi = 0$. This then implies that:

$$\bar{i}(a) = \lambda k. \tag{C.17}$$

Appendix C.1.5. Solve for individual level of prosocial activity or volunteering, a.

Then, substitute Equations (C.8), (C.10), (C.15) and (C.17) into Equation (C.3) to solve for the individual level of prosocial behavior or volunteering a.

$$\begin{aligned} ka &= v_a + y v_y + \bar{i}(a) \\ ka &= v_a + y v_y + k\lambda \\ a &= \frac{v_a + y v_y}{k} + \mu_a \rho(y) - \mu_y \chi(y) \\ a &= \frac{v_a + y v_y}{k} + \mu_a \frac{\sigma_a^2 + y\sigma_{ay}}{\sigma_a^2 + 2y\sigma_{ay} + y^2\sigma_y^2} - \mu_y \frac{1 - \frac{\sigma_a^2 + y\sigma_{ay}}{\sigma_a^2 + 2y\sigma_{ay} + y^2\sigma_y^2}}{y} \\ a(y, k|A_{pub}O_{pub}R_{prvt}) &= \frac{v_a + y v_y}{k} + \mu_a \frac{\sigma_a^2 + y\sigma_{ay}}{\sigma_a^2 + 2y\sigma_{ay} + y^2\sigma_y^2} - \mu_y \frac{\sigma_{ay} + y\sigma_y^2}{\sigma_a^2 + 2y\sigma_{ay} + y^2\sigma_y^2} \end{aligned} \tag{C.18}$$

Appendix C.1.6. Solve for aggregate level of prosocial activity or volunteering, $\bar{a}$.

The aggregate level of prosocial activity, or volunteering, is as follows.

$$\bar{a}(y,k|A_{pub}O_{pub}R_{prvt}) = \int\int \left[\frac{v_a + yv_y}{k} + \mu_a \frac{\sigma_a^2 + y\sigma_{ay}}{\sigma_a^2 + 2y\sigma_{ay} + y^2\sigma_y^2} - \mu_y \frac{\sigma_{ay} + y\sigma_y^2}{\sigma_a^2 + 2y\sigma_{ay} + y^2\sigma_y^2}\right] f(v_a, v_y) dv_a dv_y$$

$$\bar{a}(y,k|A_{pub}O_{pub}R_{prvt}) = \left[\frac{\bar{v}_a + y\bar{v}_y}{k} + \bar{\mu}_a \frac{\sigma_a^2 + y\sigma_{ay}}{\sigma_a^2 + 2y\sigma_{ay} + y^2\sigma_y^2} - \bar{\mu}_y \frac{\sigma_{ay} + y\sigma_y^2}{\sigma_a^2 + 2y\sigma_{ay} + y^2\sigma_y^2}\right] \tag{C.19}$$

Lastly, after imposing the simplifying assumption that $\sigma_{ay} = 0$, the aggregate level of volunteer behavior when the volunteer behavior is public, the incentive offer is public, and the reputation is private, is

$$\bar{a}(y,k|A_{pub}O_{pub}R_{prvt}) = \frac{\bar{v}_a + y\bar{v}_y}{k} + \underbrace{\bar{\mu}_a \frac{\sigma_a^2}{\sigma_a^2 + y^2\sigma_y^2} - \bar{\mu}_y \frac{y\sigma_y^2}{\sigma_a^2 + y^2\sigma_y^2}}_{\theta(a,y)}, \tag{C.20}$$

where $\theta(a,y)$ denotes the image payoff that agents receive from others observing them volunteering at level a for a given incentive rate y.

Appendix C.2. Solving Under Conditions: $A_{pub}O_{prvt}R_{prvt}$

While Section Appendix C.2 solves BT's model with a public incentive offer, this section will solve a modified version of BT's model assuming a private incentive offer (O_{prvt}). The other two conditions in this section though- public volunteer behavior (A_{pub}) and private reputations (R_{prvt}) - are the same as Section Appendix C.2.

In particular, if the incentive offer is private, then I assume y_b is the fixed incentive rate that others believe an agent receives, although y is the true incentive rate. This assumption may only be reasonable for when others believe an agent did not receive an incentive ($y_b = 0$), which is the relevant assumption for this study as discussed later in the experimental design. Then, her preference is represented by:

$$U(a,y) = (v_a + yv_y)\, a + I(a,y_b) - C(a) \tag{C.21}$$

where

$$I(a,y_b) = \mu_a \mathbb{E}(v_a|a,y_b) - \mu_y \mathbb{E}(v_y|a,y_b) \tag{C.22}$$

Then, differentiating Equation (C.21) with respect to a yields

$$\underbrace{C'(a)}_{=ka} = (v_a + yv_y) + \underbrace{\mu_a \frac{\partial \mathbb{E}(v_a|a,y_b)}{\partial a} - \mu_y \frac{\partial \mathbb{E}(v_y|a,y_b)}{\partial a}}_{i(a,y_b)}. \tag{C.23}$$

However, observers believe that an agent's preferences are represented by:

$$U(a,y_b) = (v_a + yv_{y_b})\, a + I(a,y_b) - C(a) \tag{C.24}$$

where

$$I(a,y_b) = \mu_a \mathbb{E}(v_a|a,y_b) - \mu_y \mathbb{E}(v_y|a,y_b) \tag{C.25}$$

Then, differentiating Equation (C.24) with respect to a yields

$$\underbrace{C'(a)}_{=ka} = (v_a + yv_{y_b}) + \underbrace{\mu_a \frac{\partial \mathbb{E}(v_a|a,y_b)}{\partial a} - \mu_y \frac{\partial \mathbb{E}(v_y|a,y_b)}{\partial a}}_{i(a,y_b)}. \tag{C.26}$$

Using the observer's belief about the first order condition, Equation (C.26), solve for $\mathbb{E}(v_a|a,y_b)$ and then $\mathbb{E}(v_a|a,y_b)$ in terms of the known values and unknown value of $r(a,y_b)$, by using signal extraction problem techniques and basic manipulations. Then, express $i(a,y_b)$ as an ODE and solve for $i(a,y_b)$. These steps for solving mirror those in the previous Sections Appendix C.1.2, Appendix C.1.3 and Appendix C.1.4.

Next, solve for a from the agents true first order condition, Equation (C.23), by substituting in the known values for the conditional expectations. Lastly, you can solve for the aggregate level of volunteer behavior, again imposing the simplifying assumption that $\sigma_{ay} = 0$, when the volunteer behavior is public, the incentive offer is private, and the reputation is private, as

$$\bar{a}(y,k|A_{pub}O_{prvt}R_{prvt}) = \frac{\bar{v}_a + y\bar{v}_y}{k} + \underbrace{\bar{\mu}_a \frac{\sigma_a^2}{\sigma_a^2 + y_b^2\sigma_{y_b^2}} - \bar{\mu}_y \frac{y_b\sigma_{y_b^2}}{\sigma_a^2 + y_b^2\sigma_{y_b^2}}}_{\theta(a,y_b)}, \tag{C.27}$$

where $\theta(a,y_b)$ denotes the image payoff that agents receive from others observing them volunteering at level a for a *believed* incentive rate y_b.

Appendix C.3. Solving Under Conditions: $A_{pub}O_{prvt}R_{pub}$ and $A_{pub}O_{prvt}R_{pub}$

While Sections Appendix C.1 and Appendix C.2 solve for the optimal level of volunteering when reputations are private (R_{prvt}), the models in this section assume a public reputation (R_{pub}). As discussed more in the paper, this study imposes one additional assumption to BT's assumptions:

Additional Assumption (Reputations Assumption)**.** *An agents' reputation perfectly identifies their type* (v_a, v_y).[31]

So, if an agent's reputation is public, then her marginal image payoff, $i(a,y) = \frac{\partial I(a,y)}{\partial a} = 0$. Then, in the case of public volunteer behavior with public incentive offers and public reputations ($A_{pub}O_{pub}R_{pub}$), we can merely substitute $i(a,y) = 0$ into the first order condition shown in Equation (C.3), which then yields the following first order condition that is easily solved as follows.

$$C'(a) = v_a + yv_y \tag{C.28}$$

$$ka = v_a + yv_y$$

$$a(y,k|A_{pub}O_{pub}R_{pub}) = \frac{v_a + yv_y}{k}$$

$$\bar{a}(y,k|A_{pub}O_{pub}R_{pub}) = \frac{\bar{v}_a + y\bar{v}_y}{k} \tag{C.29}$$

In the case of public volunteer behavior with private incentive offers and public reputations ($A_{pub}O_{prvt}R_{pub}$), we can similarly substitute $i(a,y_b) = 0$ into the first order condition shown in Equation (C.23), which then

[31] Note that an agent's reputation can be defined by a variety of characteristics as long as those characteristics perfectly identify their type.

yields the following first order condition and solution that is easily solved as follows.

$$C'(a) = v_a + yv_y \tag{C.30}$$

$$ka = v_a + yv_y$$

$$a(y, k|A_{pub}O_{prvt}R_{pub}) = \frac{v_a + yv_y}{k}$$

$$\bar{a}(y, k|A_{pub}O_{prvt}R_{pub}) = \frac{\bar{v}_a + y\bar{v}_y}{k} \tag{C.31}$$

Appendix C.4. Predictions

This study's predictions easily follow from the optimal aggregate levels of prosocial behavior or volunteering solved under the four different conditions above and shown in Equations (C.20), (C.27), (C.29), and (C.31).

Prediction 1 (The Negative Image Effect)**.** *In the case of incentivized and public volunteer behavior, with private reputations, public incentive offers discourage volunteer behavior, relative to private incentive offers, as long as* $\theta(a, y_b) > \theta(a, y)$ *(for which a sufficient condition is* $y_b \to 0$*).*

Prediction 2 (The Reputations Effect)**.** *In the case of incentivized and public volunteer behavior, with private incentive offers, public reputations discourage volunteer behavior, relative to private reputations, as long as* $\theta(a, y_b) > 0$ *(for which a sufficient condition is* $y_b \to 0$*).*

Prediction 3 (The Interactions Effect)**.** *In the case of incentivized and public volunteer behavior, with public reputations, the observability of the incentive offer has no impact on volunteer behavior.*