Wage Elasticities in Working and Volunteering: The Role of Reference Points in a Laboratory Study

Christine L. Exley,[a] Stephen J. Terry[b]

[a] Harvard Business School, Boston, Massachusetts 02163; [b] Boston University, Boston, Massachusetts 02215
Contact: clexley@hbs.edu, http://orcid.org/0000-0003-0760-8871 (CLE); stephent@bu.edu (SJT)

Abstract. We experimentally test how effort responds to wages—randomly assigned to accrue to individuals or to a charity—in the presence of expectations-based reference points or targets. When individuals earn money for themselves, higher wages lead to higher effort with relatively muted targeting behavior. When individuals earn money for a charity, higher wages instead lead to lower effort with substantial targeting behavior. A reference-dependent theoretical framework suggests an explanation for this differential impact: when individuals place less value on earnings, such as when accruing earnings for a charity instead of themselves, more targeting behavior and a more sluggish response to incentives should result. Results from an additional experiment add support to this explanation. When individuals select into earning money for a charity and thus likely place a higher value on those earnings, targeting behavior is muted and no longer generates a negative effort response to higher wages.

1. Introduction

According to estimates from the Bureau of Labor Statistics, 63 million people in the United States volunteered at least once in 2014, collectively working around eight billion hours. This effort represented about 4% of total hours worked in the United States the same year.[1] Not capturing less formal sources of volunteer activities, however, even these large figures underestimate volunteer behavior. Paid employees of nonprofit and for-profit organizations are known to volunteer in the form of unpaid "overtime" labor (see Gregg et al. 2011). Half of millennial employees have participated in company-sponsored volunteer initiatives at their place of employment (see The Case Foundation 2015). Overall, two-thirds of adults in the United States have engaged in informal volunteer activities, such as completing favors for neighbors.[2]

In considering how to encourage volunteer effort, a robust literature has found that traditional monetary incentives are often ineffective; they may limit volunteers' ability to feel good about themselves or to signal to others that they are prosocial, crowding out their motivation to volunteer.[3] One way to avoid such crowd-out concerns may involve constructing incentives that benefit a charity, instead of the volunteers themselves. Even then, recent experimental evidence from Imas (2014) suggests that increases in "volunteer wages," or the benefits to a charity from each unit of volunteers' effort, are substantially less effective at increasing effort than wage increases in a working context.[4]

We consider a potential source of weak volunteer responsiveness to incentives by appealing to a traditional mechanism from the labor economics literature: targeting. Performance targets are ubiquitous as a means to track and encourage higher outcomes.[5] But the presence of targets may backfire if they render volunteer effort unresponsive to increased incentives. That is, consider an environment in which an individual desires to produce a fixed target amount f of value. If each unit of their effort e results in an output of w units for their nonprofit, then increases in the wage w may pathologically lead to *less* effort, since a targeting individual would simply adjust their labor downward according to the schedule $e = f/w$. Overall value provided to the organization would remain unchanged at f, despite the increased incentives.

In this context, managers face a trade-off. On the one hand, targets may generate increased effort through their very existence. On the other hand, targets may render traditional incentives ineffective for boosting output. The importance of this trade-off depends

crucially on the extent to which targeting behavior is relevant in practice, and there are reasons to suspect it may be more relevant in the volunteering context. In particular, a standard reference-dependent theoretical framework suggests that when individuals place less value or intrinsic weight on earnings, more targeting behavior and a more sluggish response to incentives should result.[6] If individuals simply value earnings for the charity less than earnings for themselves, as suggested by prior literature, individuals in a volunteering context may engage in more targeting behavior and respond more negatively to incentives.[7] From a laboratory experiment where we randomly assign participants to earning money for themselves or to earning money for a charity, we indeed find evidence in support of this possibility. While we observe a positive wage elasticity in the working context, substantial targeting behavior generates a negative wage elasticity in the volunteering context.

However, the random assignment to the working or volunteering context in our laboratory experiment abstracts away from an important element in the field: the role of selection. For instance, individuals who select into volunteering for a nonprofit organization likely place a higher value or intrinsic weight on earnings for a charity, and thus the same reference-dependent theoretical framework suggests that a negative wage elasticity should be less likely. An additional online experiment that varies the recruitment procedure of participants, and allows for a greater role for selection, provides support of this prediction as well.

Our laboratory study follows a similar design to Abeler et al. (2011). In that experiment, the authors vary a reference point rather than the wage itself, remaining within the working context. Participants' effort levels often settle at the reference point exactly, consistent with their model of reference-dependent labor supply.[8] By instead varying wage rates in the presence of a fixed reference point, we provide the first laboratory test of effort response to wage changes in the presence of reference points, to our knowledge. That is, we can investigate whether targeting behavior indeed results in negative wage elasticties.

Participants solve tables in a simple but tedious real effort task that has an expectations-based reference payment of $8; participants earn a "fixed payment" of $8 with 50% probability regardless of how many tables they solve. With the remaining 50% probability, participants earn their "acquired earnings," which equal the number of tables they solve times the wage rate. While participants earn money for themselves in the working context, participants earn money for the American Red Cross (ARC) in the volunteering context. Three wage rates, all of which are chosen to allow participants to earn the reference payment of $8 exactly with an integer number of tables, are explored for each context.

In the working context, 20% of participants reach the reference payment exactly for a wage rate of 25¢. Our finding of targeting behavior in this instance replicates the results from a similar treatment in Abeler et al. (2011).[9] However, when we explore a lower wage rate of 16¢ or a higher wage rate of 50¢, there is less targeting behavior with participants instead responding to the lower and higher wages in the traditional manner—they work less when paid less and work more when paid more. We correspondingly estimate a *positive* and economically significant wage effect on effort. When wages approximately triple, workers complete about 48% *more* tables, relative to the median. We conclude that within the context of this laboratory experiment and our implemented wage variation, targeting behavior fails to overturn the traditional conclusion that effort increases as wages increase.

In the volunteering context, by contrast, 20%–30% of participants reach the reference payment exactly across all three wage rates—25¢, 50¢, and 80¢.[10] Targeting behavior across the entire wider range of wages is consistent in a reference-dependent theoretical framework with relatively lower intrinsic valuations of earnings in the volunteering context. We correspondingly estimate a *negative* and economically significant wage effect on effort: when the wage approximately triples, volunteers complete about 58% *fewer* tables relative to the median.

Our online study follows a similar procedure to the volunteer context in our laboratory study while also varying the recruitment procedure to consider the role for selection. Among participants who are recruited via material that does not highlight the opportunity to earn money for a charity during the study, negative responses to higher volunteer wages are observed, as in our laboratory study. Among participants who are recruited via material that highlights the opportunity to earn money for a charity during the study, negative responses to higher volunteer wages are no longer observed.[11]

The results from our two studies provide insight into when managers seeking to elicit higher effort might justifiably worry that the imposition of targets causes sluggish or negative responses to incentives. In situations where individuals are highly motivated for earnings, targeting behavior will likely be weak. For example, employees earning money for themselves or nonprofit volunteers who have undergone any stringent form of selection may place high value on their earnings. However, if people care intrinsically little about earnings, targeting may be strong and render traditional incentives ineffective. Such people may include experimental participants assigned to volunteering, workers volunteering at company-sponsored events, workers completing unpaid overtime, or volunteers only loosely attached to a nonprofit.

We contribute to the broad targeting literature by highlighting how the relevance of targeting may depend on the underlying intrinsic motivation that likely varies across contexts and across different types of selection into particular contexts. Much of this literature focuses on the role of targeting among workers. For instance, appealing to theories involving loss aversion and reference-dependence in a field experiment involving a delivery service in Zurich, Fehr and Goette (2007) find that higher wages do in fact induce lower effort.[12] Camerer et al. (1997) find observational evidence for negative wage elasticities among New York City taxicab drivers, but this sparked a debate including contributions by Farber (2005, 2008, 2015), Crawford and Meng (2011), Chou (2002), and Doran (2014). Recently, this literature has expanded to investigate the potential explanatory power of targeting for contexts as diverse as the duration of unemployment and performance in sports, such as in Pope and Schweitzer (2011), Allen and Dechow (2013), Allen et al. (2017), and DellaVigna et al. (2017).[13] Beyond the targeting literature, we also contribute to a comparative literature that documents how behavioral motivations may prove more relevant in prosocial settings.[14] Finally, by discussing the potential pitfalls of performance targets, we contribute to a rich literature in labor economics, corporate finance, and macroeconomics that discusses the potential drawbacks or pathological effects of such targets (see, e.g., Oyer 1998, Larkin 2014, Terry 2017).

The remainder of this paper proceeds as follows: Section 2 details our design; Section 3 presents our laboratory results; Section 4 discusses results from an additional online experiment motivated by our laboratory results; Section 5 concludes. In the online appendix, we provide additional results and robustness checks, together with more information on our theoretical predictions.

2. Design for the Laboratory Experiment

Our laboratory study consists of participants earning payments according to two states of the world. First, with probability 0.5, their payments equal acquired earnings that they accumulate by completing an effort task. A wage rate w is given for each unit of effort completed, so acquired earnings for a participant with effort level e equal we. Second, with probability 0.5, participants' payments equal a fixed payment f regardless of how many units of effort they complete. The total payment to a participant in "working" treatments will be awarded to the participant themselves, and in alternative "volunteering" treatments the payment will be awarded instead to a charitable organization.[15]

How does this lottery structure allow us to study the role of targeting behavior? To answer that question, we will first lay out a benchmark theoretical structure

of effort determination that omits a role for targeting before discussing the remaining details of our experimental design. Then, we follow Abeler et al. (2011) and extend the environment to allow for loss aversion and expectations-based reference dependence. In that extended version the fixed payment f, which is controlled and identifiable within the laboratory environment, serves as a target level for participant earnings.

First, consider the following exceedingly simple benchmark model. Let each agent have the following quasilinear preferences in their expected value of earnings c and disutility from provided effort e:[16]

$$\mathbb{E}(\alpha c) - \frac{\gamma}{2}e^2.$$

Here, $\alpha > 0$ represents the weight on participant earnings, which might vary by context. For instance, we would likely expect lower levels of intrinsic payoff from earnings α in a volunteer context than in a working context, since individuals earn money for others rather than themselves. Given our lottery structure, labor supply or effort choice e results in payoffs given by $\frac{1}{2}\alpha we + \frac{1}{2}\alpha f - \gamma/2e^2$.[17] Optimization of these payoffs in effort choice e yields the classical optimal labor supply function e^{class}, where

$$e^{\text{class}}(w, f, \gamma, \alpha) = \frac{\alpha w}{2\gamma}.$$

We immediately see that the fixed payment f does not enter classical labor supply, and further we have that labor supply is uniformly upward-sloping in the wage.[18]

We now consider the implications of introducing another term in preferences that allows for loss aversion in agents indexed by a parameter $\lambda \geq 1$. In general, loss aversion and hence the value of λ may vary across participants. When faced with outcome lottery c, an agent possessing a reference lottery r experiences "gain-loss utility" $\mu(x)$ based on the difference in utility payoffs between the outcome and reference lotteries $x = \alpha c - \alpha r$:

$$\mu(x) = \begin{cases} \lambda x, & x \leq 0; \\ x, & x \geq 0. \end{cases}$$

Therefore, higher values of loss aversion λ for an agent imply that deviations in outcomes below the target or reference lottery r are more painful. To incorporate gain-loss preferences in the presence of loss aversion, we add to payoffs the expression $\mathbb{E}_{c,r}\mu(\alpha c - \alpha r)$, with expectations taking into account uncertainty in both c and r.[19] The reference lottery r can in principle be chosen in many different ways. For instance, the expectations-based approach we follow from Kőszegi and Rabin (2006), which maximizes our comparability

with existing laboratory studies, requires that the reference lottery equals the equilibrium outcome lottery itself.[20] The reference lottery and hence gain-loss utility involves only monetary payoffs in this framework, since effort costs do not vary with the outcome of the fixed payment versus wage lottery. Based on this structure, if the agent chooses an effort level e with $we \leq f$, their payoffs are given by

$$\frac{1}{2}\alpha we + \frac{1}{2}\alpha f - \frac{\gamma}{2}e^2 + 4\left[\frac{1}{2}\left(\frac{1}{2}(\alpha we - \alpha we)\right.\right.$$
$$\left.\left. + \frac{1}{2}\lambda(\alpha we - \alpha f)\right) + \frac{1}{2}\left(\frac{1}{2}(\alpha f - \alpha we) + \frac{1}{2}(\alpha f - \alpha f)\right)\right].$$

Here, the first three terms duplicate the classical payoff, and the four terms in brackets make up the gain-loss term. To understand the gain-loss term, consider the case in which the agent receives we, which occurs with probability $\frac{1}{2}$. With probability $\frac{1}{2}$, the agent expected we and experiences zero gain or loss $\alpha we - \alpha we = 0$. However, with probability $\frac{1}{2}$, the agent expected to receive the larger fixed payment $f \geq we$, and in this case they experience loss in the total amount $\lambda(\alpha we - \alpha f)$. These considerations account for the first two terms in brackets. However, with probability $\frac{1}{2}$ the agent actually receives the fixed payment $f \geq we$. If they expected we, the agent experiences the gain $\alpha f - \alpha we$ (the third term), and if they expected f, the agent experience zero gain or loss with $\alpha f - \alpha f = 0$, the fourth term.

A similar logic applies when the agent chooses effort e with acquired earnings we greater than the fixed payment f; the payoffs for the agent in all cases are provided in the theory appendix (available in Section C of the online appendix). The presence of loss aversion always implies that deviations of acquired earnings we from the fixed payment f involve the possibility of costly disappointment, inducing a kink in payoffs. As discussed in detail in the theory appendix, the resulting segmented labor supply function is

$$e^{ref}(w, f, \gamma, \alpha, \lambda) = \begin{cases} e_1, & e_1 > \dfrac{f}{w}; \\ \dfrac{f}{w}, & e_1 \leq \dfrac{f}{w} \leq e_2; \\ e_2, & e_2 < \dfrac{f}{w}, \end{cases}$$

where we have $e_1 = (\alpha w(3/2 - \lambda))/\gamma$ and $e_2 = (\alpha w(\lambda - 1/2))/\gamma$. We can determine some things about e^{ref} immediately. First, in contrast to the classical case, labor supply responds to the level of the reference or fixed payment f and is in fact weakly increasing in f. Abeler et al. (2011) explicitly state and then provide experimental evidence for this result by varying the fixed payment f. Second, and more directly useful for our purposes, we can also describe the shape of the dependence of labor supply on the wage w.

Figure 1. Optimal Labor Supply Is Segmented
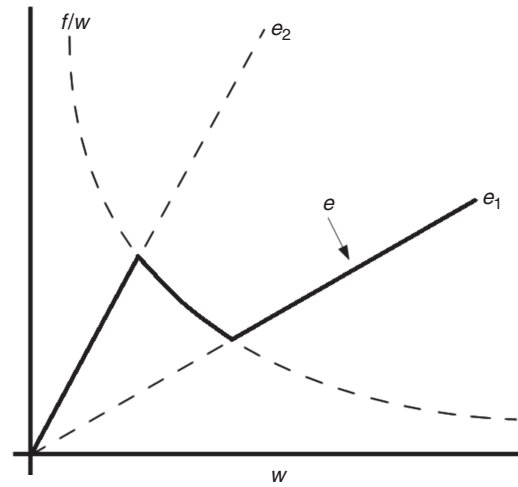
Notes. This figure plots the configuration of optimal segmented labor supply $e^{ref}(w, f, \gamma, \alpha, \lambda)$ as the wage w varies, in the case that $1 < \lambda < \frac{3}{2}$. The case that $\lambda \geq \frac{3}{2}$ is discussed in the theory appendix, and at the boundary $\lambda = 1$, $e^{ref} = e^{class}$. The shaded, dotted lines are the interior labor supply optimizers e_1 and e_2, together with the corner reference point solution f/w. The bold overlaid, segmented line labeled e in the figure is the labor supply curve e^{ref} itself.

In particular, Figure 1 plots a stylized version of this reference-dependent effort supply, e^{ref}.[21,22] For very low wages w, effort increases with w. Similarly, for very high wages w, effort increases with the wage. However, for intermediate wages w, targeting behavior induces $e = f/w$ as acquired earnings hit the reference or fixed payment f. This targeting behavior occurs because for intermediate levels of the wage, earnings in the classical case are not too far from the target level f. Since deviation from the fixed payment involves potential disappointment for loss-averse agents, it is optimal to avoid such disappointment through choice of exactly the target level of labor supply. This yields labor supply that is downward-sloping in w.

The range of intermediary wages for which targeting behavior occurs and negative wage elasticities may be observed will likely differ across contexts. For many parameterizations of the model, the range of wages that induce target behavior by agents is given by $w_1 - w_2$, where $w_1 = \sqrt{f\gamma/(\alpha(3/2 - \lambda))}$ and $w_2 = \sqrt{f\gamma/(\alpha(\lambda - 1/2))}$. In these cases, it is easy to show that $(\partial w_1 - w_2)/\partial\alpha < 0$. More simply, a lower intrinsic value α placed on earnings widens the region over which agents exert exactly the target level of effort f/w, assuming that there are no other changes in the distributions of agent preference parameters.[23] Since agents exhibiting targeting behavior actually reduce their effort in response to higher wage rates w, more targeting can serve to weaken the overall effort response to increased incentives. In summary, contexts in which agents care little about earnings are predicted to feature a high level of targeting, while contexts with

strong intrinsic motivation should exhibit more traditional responses to incentives.

To directly test the effort response to wages in this context, in our experimental design we hold constant the value of the fixed payment f and instead vary the offered wage w as well as the recipient of agent's overall monetary rewards across contexts.

First, we set an expectations-based reference point such that participants expect to earn a reference or fixed payment f of $8 with 50% probability. When a participant enters the lab, they are shown the contents of two envelopes. One envelope contains a sheet of paper that says "Sheet A: Acquired Earnings," while the other envelope contains a sheet of paper that says "Sheet B: Fixed Payment $8." The study leader mixes these envelopes in a bag, and then the participant selects one envelope. The participant does not open the envelope until after the study is complete, so a participant only knows that there is an equal probability that they selected an envelope containing Sheet A or B.[24]

If the participant's envelope contains "Sheet A: Acquired Earnings," their earnings will be equal to their acquired earnings of we. Subjects' acquired earnings result from them solving tables in a simple but tedious real effort task. Successfully solving a table requires participants to correctly count how many 0s are in a randomly-generated series of 150 0s and 1s. Once a participant correctly solves one table, a new table is randomly generated.[25] For each table a participant solves, a participant's acquired earnings increase by a fixed wage rate, w. Participants are allowed to solve tables for as little or as long as they want, up to 60 minutes. Their effort e is the total number of tables they solve. On the other hand, if the participant's envelope contains "Sheet B: Fixed Payment $8," their earnings will be equal to the fixed payment f of $8, irrespective of how many tables are solved.

Second, as noted above we examine both a working and a volunteering environment across subjects so that each participant is only exposed to one of these environments. In the working environment, participants earn money for themselves. In the volunteering environment, by contrast, participants earn money for the ARC. That is, the ARC will receive a participant's acquired earnings of we, or fixed payment f of $8 if their envelope contains Sheet A or Sheet B, respectively. See Online Appendix Figures A.1 and A.2 for screenshots of the main effort task in the working and volunteering environments.

Third, we vary the wage w across subjects so that each participant is only exposed to one of the wage levels. By varying the wage faced by participants, as opposed to the reference payment, we can directly observe the responsiveness of effort to wage changes and offer a new laboratory test of the empirical relevance of expectations-based reference points for labor supply.

There are a few other design features worth noting. Each study session only involves one participant at a time, to ease concerns about peer effects, conformity, and image motivation, such as wanting to appear prosocial.[26] In other words, each experimental participant completed all study tasks within a separate laboratory room not containing any other experimental participants. Prior to completing the real effort task of solving tables, participants must successfully answer several understanding questions and complete a practice round. In the practice round, they also solve tables but are only paid a known and fixed piece rate of 10¢ for each table they solved within four minutes. After completing the real effort task of solving tables, participants complete a short follow-up survey to gather demographic and other relevant information and then are paid in cash.[27]

Single person sessions were run from March to October 2013 in the Stanford Economics Research Laboratory (SERL). When recruiting participants from the laboratory's pool of eligible undergraduate students from Stanford University, participants were neither informed that they may earn money for the ARC nor given details about the decisions they would make. Consistent with standard practice for SERL, participants expected an average compensation around $20 per hour. This resulted in 180 undergraduate students from Stanford University, or 30 participants in each of a total of six treatment groups (2 contexts × 3 wage rates). Across the treatment groups, participants were similar on observables, as shown in Online Appendix Table A.10.

3. Results from the Laboratory Experiment

We first analyze a two-by-two design to investigate if participants respond differently to wages in the volunteering and working environment. Participants face a wage rate w of {25¢ or 50¢} in a {working or volunteering} environment. Both wage rates allow participants to earn the reference or fixed payment f of $8 exactly by putting forth effort e of 32 or 16 tables solved given the wage rates w of 25¢ or 50¢, respectively.

To consider how effort responds to the wage rates in volunteering and working, we thus estimate $Tables_i = \beta_0 + \beta_1 I(Volunteering)_i + \beta_2 I(w = \$0.50)_i + \beta_3 I(Volunteering)_i * I(w = \$0.50)_i + [Controls_i] + \epsilon_i$. The dependent variable is participants' effort level, $Tables_i$, which equals the number of tables they solve. Indicators for the volunteering environment and 50¢-wage rate are $I(Volunteering)_i$ and $I(w = \$0.50)_i$, respectively. Table 1 presents the corresponding median, ordinary least squares (OLS), and Tobit estimates, with and without controls.[28] The coefficient on $I(Volunteering)_i$, while consistently negative, indicates that there are no significant differences between effort for volunteers and workers given the low wage of 25¢. However,

Table 1. Number of Tables Solved

	Median		OLS		Tobit	
$I(Volunteering)$	−2.00	−7.41	−8.87	−7.51	−8.92	−7.56
	(5.16)	(6.03)	(7.28)	(7.50)	(7.42)	(7.37)
$I(w = \$0.50)$	16.00***	12.94**	13.90*	10.00	13.96*	10.09
	(5.16)	(6.13)	(7.28)	(7.63)	(7.42)	(7.51)
$I(Volunteering) *$	−29.00***	−21.18**	−24.30**	−22.25**	−25.01**	−22.87**
$I(w = \$0.50)$	(7.30)	(8.44)	(10.29)	(10.50)	(10.51)	(10.35)
Constant	34.00***	29.00***	40.50***	33.87**	40.07***	32.71**
	(3.65)	(10.60)	(5.15)	(13.18)	(5.25)	(12.97)
Controls	No	Yes	No	Yes	No	Yes
N	120	120	120	120	120	120

Notes. Regression results from $Tables_i = \beta_0 + \beta_1 I(Volunteering)_i + \beta_2 I(w = \$0.50)_i + \beta_3 I(Volunteering)_i * I(w = \$0.50)_i + [Controls_i] + \epsilon_i$. The dependent variable, *Tables*, is the number of tables completed in the up to 60-minute real effort task for participant i. All regressions are at the participant level. $I(Volunteering)_i$ is an indicator for participant i earning money for the charity (as opposed to for themselves), $I(w = \$0.50)_i$ is an indicator for participant i having a wage equal to $0.50 (as opposed to $0.25). Controls include a productivity measure defined as the number of tables completed in the four-minute practice round and indicators for whether or not some participant is a male, a U.S. citizen, a freshman, a sophomore, a junior, has stated volunteer hours above the median of the experimental sample, and feels favorably about the American Red Cross. Standard errors are in parentheses.
*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

doubling the wage to 50¢ is significantly less effective at encouraging effort for volunteers than workers, as shown by the robust and negative coefficient on $I(Volunteering)_i * I(w = \$0.50)_i$. We summarize the following:

> **Working vs. Volunteering Result**: *Increasing wages from 25¢ to 50¢ is substantially less effective at encouraging more volunteering effort than working effort.*

The weaker response of effort to incentives that we observe in the volunteering context relative to working echoes the results in Imas (2014) and more broadly the literature on how incentives in volunteering contexts often fail. Crucially though, our experimental design allows us to dive deeper and investigate targeting as a particular explanation for this observed difference in wage elasticities. The following subsections will therefore consider the role of targeting in effort put forth by volunteers and workers, and in doing so, also introduce one additional wage treatment group for both the working and volunteering contexts. Our experiment's one-person-per-session structure makes additional treatments quite lengthy and costly to run. Therefore, as discussed below, we used reference-dependent theory as a guide for choosing one additional new wage in each context after analyzing the results from the above two-by-two design.

3.1. Working Results
Figure 2 plots the distribution of effort in the working contexts, and the black bars indicate the percentage of participants whose effort level is equal to the reference level, or yields acquired earnings equal to the fixed payment f of $8 exactly. For the low wage rate of 25¢, over 20% of workers have effort equal to the

reference level. In fact, the observed targeting behavior for workers nearly replicates one treatment condition in Abeler et al. (2011).[29] With the higher wage rate of 50¢, however, the frequency with which workers' effort levels equal their reference level exactly is cut in half to only 10% of the time. Nearly all other workers instead exceed their reference level with the 50¢ wage.

Using Figure 1 as a guide, this pattern suggests that while a 25¢ wage may fall on a downward-sloping portion of the labor supply, 50¢ likely falls to the far right on an upward-sloping portion of labor supply. In an attempt to explore the relevant range of targeting behavior for labor supply in the working context, we thus ran an additional treatment with a lower wage of 16¢. The result, as shown in Figure 2, is clustering remains evident in slightly weaker fashion with the lower wage of 16¢.[30]

To consider whether the varying levels of targeting behavior correspond with the responses to wage changes, we estimate $Tables_i = \beta_0 + \beta_1 I(w = \$0.25)_i + \beta_2 I(w = \$0.50)_i + [Controls_i] + \epsilon_i$. The dependent variable is participants' effort level, $Tables_i$, which equals the number of tables they solve. Indicators for the wages of 25¢ and 50¢ are $I(w = \$0.25)_i$ and $I(w = \$0.50)_i$, respectively, while the excluded category is the 16¢ wage. Table 2 presents the corresponding median, OLS, and Tobit estimates, with and without control.[31] As shown by the estimated coefficient on $I(w = \$0.25)_i$, there is positive but insignificant impact of increasing wages from 16¢ to 25¢. Coupled with some observed clustering at both of these wage levels, this insignificant finding leaves room for the possibility that targeting behavior may somewhat reduce wage elasticities in the working environment. Nonetheless, there is no significant evidence for negative wage elasticities. As

Figure 2. (Color online) Working: Number of Tables Solved by Wage
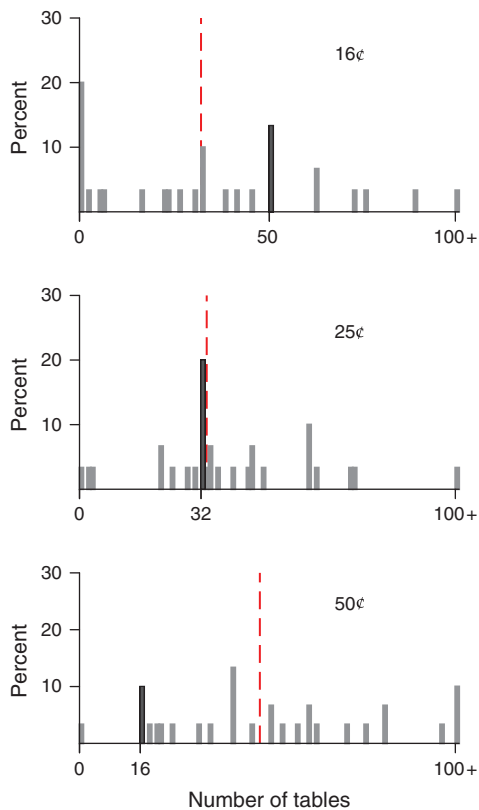
Notes. The figure plots the observed distribution of tables completed by experimental participants for each of the three wages when participants are earning money for themselves. The height of the black bar indicates the percentage of participants who stopped solving tables once they hit the reference level of effort, or the reference payment of earning $8. The location of the dashed line indicates the median number of tables completed within that treatment group. Each treatment includes 30 Stanford University undergraduate participants, for a total of 180 participants. Each bar has a width of 1, except for final bin of "100+," which represents the percentage of participants who solved 100 or more tables.

shown by the estimated coefficient on $I(w = \$0.50)_i$, the overall impact of increasing wages from 16¢ to 50¢ is significantly positive on effort level.[32] We summarize the following:

Working Result: *Effort levels exhibit limited targeting behavior. Increasing wages by approximately threefold from $0.16 to $0.50 leads to a 48% median increase in effort.*[33]

In other words, when participants earn money for themselves, we neither observe a wide band of targeting behavior nor experimentally recover backward-bending effort. Of course, there may still exist some smaller section of wages with downward-sloping labor supply. Given our theoretical framework and experimental results, we conclude that the relevant range of wages for which overall labor supply may be downward-sloping is narrower than $0.16 \leq w \leq \$0.50$, limiting its scope in our context. Stressing caution in

Table 2. Working: Number of Tables Solved

	Median		OLS		Tobit	
$I(w = \$0.25)$	2.00	8.30	6.87	9.54	10.14	13.64
	(8.73)	(9.30)	(8.46)	(9.13)	(9.09)	(9.44)
$I(w = \$0.50)$	18.00**	18.10**	20.77**	19.14**	24.12***	22.69**
	(8.73)	(9.01)	(8.46)	(8.85)	(9.08)	(9.09)
Constant	32.00***	20.25	33.63***	25.29	29.78***	14.02
	(6.18)	(19.34)	(5.98)	(18.98)	(6.49)	(20.06)
Controls	No	Yes	No	Yes	No	Yes
N	90	90	90	90	90	90

Notes. Regression results from $Tables_i = \beta_0 + \beta_1 I(w = \$0.25)_i + \beta_2 I(w = \$0.50)_i + [Controls_i] + \epsilon_i$. The dependent variable, *Tables*, is the number of tables completed in the up to 60-minute real effort task for participant i. All regressions are at the participant level. $I(w = \$0.25)_i$ and $I(w = \$0.50)_i$ are indicators for participant i having a wage equal to $0.25 and $0.50, respectively (with the excluded wage level being $0.16). Controls include a productivity measure defined as the number of tables completed in the four-minute practice round and indicators for whether or not some participant is a male, a U.S. citizen, a freshman, a sophomore, a junior, has stated volunteer hours above the median of the experimental sample, and feels favorably about the American Red Cross. Standard errors are in parentheses.
*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

extrapolation here is warranted. Note that integer constraints on the numbers of tables completed restrict us in most cases to fairly large percentage changes in wages across treatments, and wage variation in practice may naturally involve smaller wage changes.[34]

3.2. Volunteering Results

Figure 3 plots the distribution of effort in the volunteering contexts with the black bars again indicating the percentage of participants whose effort level is equal to the reference level. For both wage rate of 25¢ and 50¢, over 20% volunteers have effort equal to the reference level. In choosing an additional wage, we therefore sought to find an upper bound for the targeting range by more than tripling the low wage so our additional wage is 80¢. Remarkably, however, targeting behavior remains persistent with over 20% of volunteers again having effort equal to the reference level when the wage is 80¢.

To consider whether the persistent targeting behavior corresponds with reduced worker effort in response to higher wages, we estimate $Tables_i = \beta_0 + \beta_1 I(w = \$0.50)_i + \beta_2 I(w = \$0.80)_i + [Controls_i] + \epsilon_i$. The dependent variable is participants' effort level, $Tables_i$, which equals the number of tables they solve. Indicators for the wages of 50¢ and 80¢ are $I(w = \$0.50)_i$ and $I(w = \$0.80)_i$, respectively, while the excluded wage is 25¢. Table 3 presents the corresponding median, OLS, and Tobit estimates, with and without controls.[35] Relative to the lowest wage of 25¢, we observe a statistically significant reduction in effort when the wage is instead 50¢ or 80¢. However, we find an insignificant difference between effort in response to 50¢ or 80¢, suggesting

Figure 3. (Color online) Volunteering: Number of Tables Solved by Wage
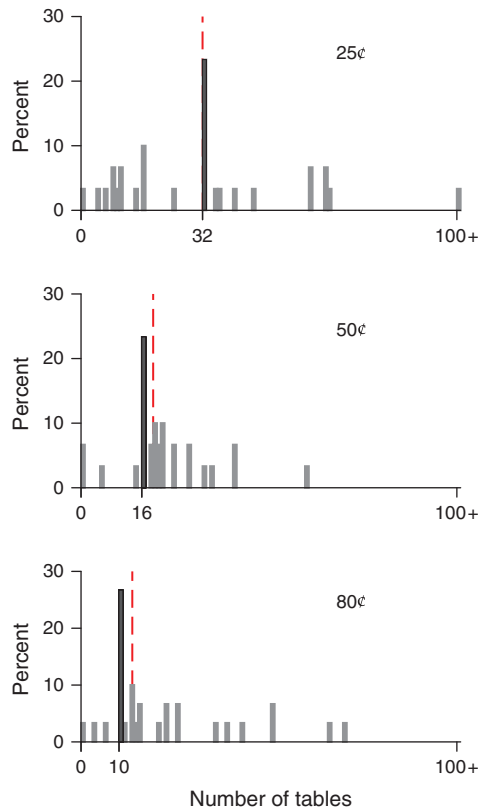
Notes. The figure plots the observed distribution of tables completed by experimental participants for each of the three wages when participants are earning money for the ARC. The height of the black bar indicates the percentage of participants who stopped solving tables once they hit the reference level of effort, or the reference payment of earning $8. The location of the dashed line indicates the median number of tables completed within that treatment group. Each treatment includes 30 Stanford University undergraduate participants, for a total of 180 participants. Each bar has a width of 1, except for final bin of "100+," which represents the percentage of participants who solved 100 or more tables.

that 80¢ may be an upper bound for which volunteering labor supply may be downward-sloping in our setting.[36] We summarize the following:

Volunteering Result: *Effort levels exhibit strong targeting behavior. Increasing wages by approximately threefold from $0.25 to $0.80 leads to a 58% median decrease in effort.*[37]

In other words, the empirical relevance of targeting behavior for effort responses seems very strong in the volunteering environment. Different from the working context, in which we fail to recover evidence of backward-bending labor supply, our volunteering results suggest that targeting is important for the response of effort to incentives over a wide range of parameters when individuals earn money for a charity.

Table 3. Volunteering: Number of Tables Solved

	Median		OLS		Tobit	
$I(w = \$0.50)$	−13.00***	−10.26**	−10.40**	−9.94**	−10.79**	−10.19**
	(4.47)	(4.69)	(4.91)	(4.96)	(5.04)	(4.83)
$I(w = \$0.80)$	−18.00***	−13.85***	−10.07**	−12.28**	−10.12**	−12.62**
	(4.47)	(4.84)	(4.91)	(5.11)	(5.03)	(4.98)
Constant	32.00***	24.18***	31.63***	24.73**	31.36***	24.90***
	(3.16)	(9.06)	(3.48)	(9.57)	(3.55)	(9.31)
Controls	No	Yes	No	Yes	No	Yes
N	90	90	90	90	90	90

Notes. Regression results from $Tables_i = \beta_0 + \beta_1 I(w = \$0.50)_i + \beta_2 I(w = \$0.80)_i + [Controls_i] + \epsilon_i$. The dependent variable, $Tables$, is the number of tables completed in the up to 60-minute real effort task for participant i. All regressions are at the participant level. $I(w = \$0.50)_i$ and $I(w = \$0.80)_i$ are indicators for participant i having a wage equal to $0.50 and $0.80, respectively (with the excluded wage level being $0.25). Controls include a productivity measure defined as the number of tables completed in the four-minute practice round and indicators for whether or not some participant is a male, a U.S. citizen, a freshman, a sophomore, a junior, has stated volunteer hours above the median of the experimental sample, and feels favorably about the American Red Cross. Standard errors are in parentheses.
  *$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

4. Design and Results from the Additional Online Experiment to Consider the Role of Selection

As detailed in Section 2, if individuals place a lower intrinsic value (α) on earnings for the charity than themselves, we would expect a wider region over which agents exhibit targeting behavior in the volunteering context than in the working context. Our findings from the laboratory eriment above are consistent with this logic. However, negative responses to volunteer wages may also be less likely in situations where individuals select into the volunteering context. Individuals selecting into volunteering may have a higher intrinsic valuation on earnings for charities, as seems likely both intuitively and as can be shown formally in an extension of our theoretical framework with selection in our online appendix. Our theory would predict a reduced prevalence of targeting behavior for such individuals.

To consider this potential mechanism of selection on valuations α in our context, we ran an online version of our study on Amazon Mechanical Turk. See Paolacci et al. (2010) and Horton et al. (2011) for details about this platform. Four hundred workers, required to have been in the United States and to possess high approval ratings of at least 95% from 100 or more previous tasks on the platform, participated in our study in response to a "Self Ad" or "Charity Ad."[38] Recruiting participants in the afternoon of February 17, 2016 and morning of February 18, 2016, the Self Ad read *"Academic survey with $2 completion award and additional money for yourself possible!"* Recruiting participants in the morning of February 17, 2016 and the afternoon of February 18, 2016, the Charity Ad read *"Academic survey with $2 completion*

award and additional money for American Red Cross possible!"

While participants choose to complete our study in response to different advertisements, participants view identical study materials after being recruited from these advertisements. Any differences in behavior across the Self Ad condition and Charity Ad condition only reflect the potentially different selection of participants into these conditions. In particular, simple theoretical frameworks such as ours would suggest that the Charity Ad condition likely recruits individuals with higher valuations of money for the ARC. For such selected individuals, we may therefore expect a reduced prevalence of targeting behavior.

After participants are recruited into the online version of our study, the study procedures follow the volunteering context design in Section 2 with a few modifications. First, the instructions, terminology, and tables are simplified as shown via a screenshot in Online Appendix Figure A.3. Second, the payment parameters are lowered to be appropriate for payments on Amazon Mechanical Turk. Third, while the participants still face an equal chance of earning their fixed amount or acquired earnings for the ARC, chance is resolved via computer code.

In particular, the study proceeds as follows. First, participants must successfully answer several under-standing questions and complete a practice round. The practice round requires participants to complete 10 tables and thus earn an additional $1 for themselves. Second, participants learn about the payments to the ARC associated with the real effort task. With a 50% chance, the ARC will receive a fixed amount of 28¢ regardless of how many tables they solve. With a 50% chance, the ARC will receive their acquired earnings of we, where w is their wage rate and e is their effort level that equals the number of tables they choose to solve. Participants are randomly offered either a volunteer wage of 2¢ or 4¢.[39] Third, participants complete as many tables as they choose—up to 100 tables—with the option to stop completing tables at any time by clicking on the button that reads "click here to stop volunteering."[40] Fourth, participants learn how much money the ARC will receive according to the chance resolved by the computer code. Fifth, participants complete a follow-up study to gather demographic and other relevant information and then payments are distributed.

Note that our design allows us to recruit participants under the Self Ad or Charity Ad without engaging in any deception. Participants earn additional payments for themselves in the practice round, a feature highlighted in the Self Ad. Participants may earn additional payments for the ARC in the real effort task, a feature highlighted in the Charity Ad.

Figure 4. (Color online) Volunteering in Online Study: Number of Tables Solved by Wage and Advertisment
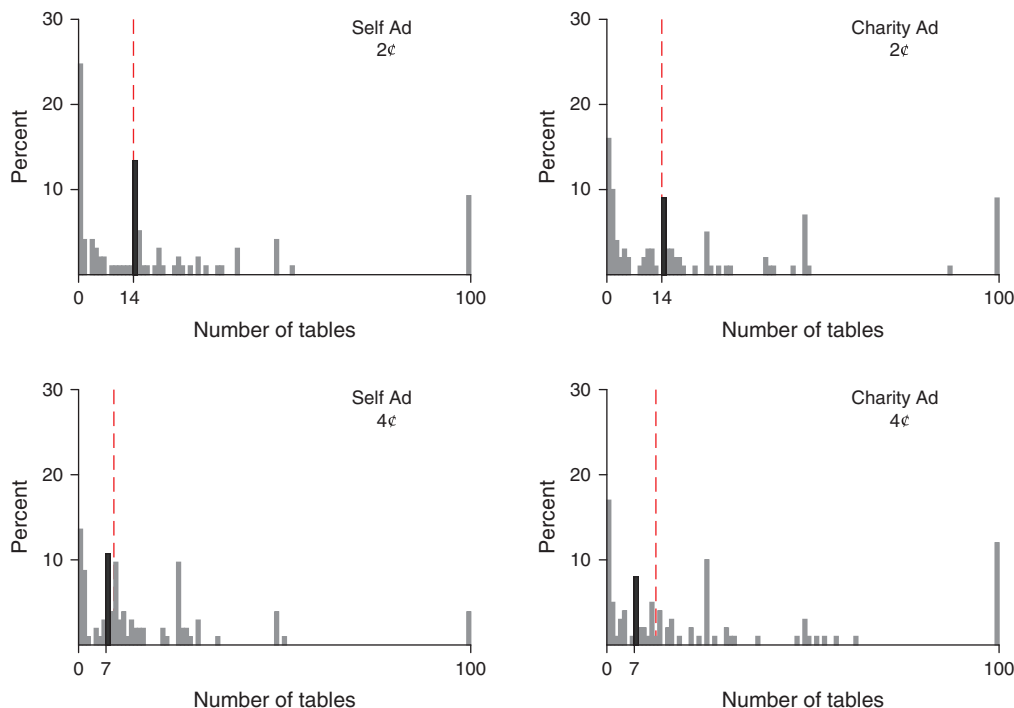
Notes. The figure plots the observed distribution of tables completed by Amazon Mechanical Turk participants according to their offered wage (of 2¢ or 4¢) and whether they were recruited via a Self Ad or Charity Ad. The height of the black bar indicates the percentage of participants who stopped solving tables once they hit the reference level of effort, or the reference payment of earning 28¢. The location of the dashed line indicates the median number of tables completed within that treatment group. Each treatment includes 97–103 participants, for a total of 400 participants. Each bar has a width of 1. Participants were not allowed to solve more than 100 tables.

Among participants recruited via the Self Ad, as shown on the left-hand side of Figure 4, there is substantial clustering around the reference level of 14 tables when the wage is 2¢ and 7 tables when the wage is 4¢. The top panel of Online Appendix Table A.11 indeed confirms that this negative wage elasticity is statistically significant when considering estimates at the median, and qualitatively but not statistically significant when considering OLS and Tobit estimates. Evidence for targeting behavior, however, appears less compelling when instead considering participants recruited via the Charity Ad, as shown on the right-hand side of Figure 4. The bottom panel of Online Appendix Table A.11 reports no significant evidence for a negative wage elasticity when considering estimates at the median, and the OLS and Tobit estimates support a positive, albeit also insignificant, effort response to higher wages. Comparisons across the Self Ad and Charity Ad are therefore qualitatively, but not significantly, supportive of a more negative wage elasticity resulting from the Self Ad. In other words, the results of our online experiment are consistent with a weaker role for targeting behavior when highly motivated individuals self-select into volunteering.

5. Conclusion

In this paper, we experimentally test the labor supply response to wage changes in the presence of a reference point or target. In line with prior targeting literature, we might expect participants to sometimes choose their effort such that they earn the reference payment, working less when they are paid more.

In our laboratory experiment, we find some evidence of targeting behavior in the working context, but we do not find any significant evidence in favor of a negative wage elasticity. Workers solve about 48% more tables, relative to the median, when the wage is approximately tripled. By contrast, we find that higher wages induce lower effort because of strong targeting behavior in the volunteering context. Volunteers solve about 58% fewer tables relative to the median when their effective wage is more than tripled.

A reference-dependent theoretical framework suggests a potential explanation for this differential impact of targets when participants are randomly assigned to the working versus volunteering context. In particular, when agents place less weight on earnings, such as when assigned to earn money for a charity instead of themselves, the model predicts more targeting and a more sluggish or negative response to higher wages.

By the same logic, however, when individuals select into a volunteer opportunity—instead of finding themselves faced with a volunteer opportunity—they may place higher weight on earnings to a charity and thus a negative response to higher wages may be less likely. Results from our additional online study support this

possibility. Among participants who select into the study knowing that they will face a volunteer opportunity, targeting behavior does not generate a negative effort response to higher wages.

Both policy makers and managers seeking to elicit more prosocial behavior through volunteering might do well to take these findings into account when relying on reference points, embodied as explicit or implicit targets and goals, to encourage more effort. When laborers are highly motivated, such as employees or volunteers highly attached to a nonprofit, targets may work well. By contrast, when volunteers are only loosely attached to a charity, or when workers are not compensated for their efforts, targets may backfire and generate a negative response to incentives. Future work may also seek to consider other mechanisms that may influence the degree of targeting behavior across contexts.[41]

Endnotes

[1] Calculation of these aggregate figures is straightforward, drawing on data from the Bureau of Labor Statistics' 2014 release *Volunteering in the United States*, the same agency's *Current Employment Statistics* as of July 2015, as well as the authors' calculations. Note that these figures rely on the Bureau of Labor Statistics' definition of volunteering. Both legally and in practice, the definition of volunteering may be complicated as noted in http://www.dol.gov/elaws/esa/flsa/docs/volunteers.asp (accessed November 11, 2017) and Musick and Wilson (2007).

[2] See https://www.nationalservice.gov/vcla/national (accessed November 11, 2017).

[3] Related works on intrinsic motivation for prosocial behavior include Titmuss (1970), Andreoni (1989, 1990), Bénabou and Tirole (2003), Frey and Oberholzer-Gee (1997), Gneezy and Rustichini (2000), Frey and Jergen (2001), and Meier and Stutzer (2008). Studies considering image motivation or signaling include Bénabou and Tirole (2006), Ariely et al. (2009), Goette et al. (2010), Gneezy and Rustichini (2000), Mellström and Johannesson (2008), Carpenter and Myers (2010), Meer (2011), Lacetera et al. (2012, 2014), and Exley (2017). For a nice survey related to when incentives may succeed or backfire, see Gneezy et al. (2011).

[4] Relatedly, Karlan and List (2007) and Null (2011) document in field experiments that charitable donations also appear unresponsive to the social benefit of giving.

[5] In fact, consulting firms routinely advise nonprofits on the judicious choice of such targets (see Sawhill and Williamson 2001 for an example). Note also that in this paper when we refer to volunteer targets, we are predominantly referring to goals set for the volunteers themselves within charitable organizations rather than the paid employees of charitable organizations.

[6] Other factors could also be at work, such as lower loss aversion parameters in volunteering relative to working, or a potentially correlated shift between loss aversion parameters and intrinsic valuations.

[7] The subsequent laboratory study discussed in this paper involves Stanford undergraduates as does the study in Exley (2015), which finds that over 90% of participants value money for a charity less than money for themselves. The subsequent online study discussed in this paper involves participants from Amazon Mechanical Turk, as does the study in Exley and Kessler (2017), which also finds that over 90% of participants value money for a charity less than money for themselves.

[8] Other studies find that more nuanced predictions of reference-dependent theory for labor supply may not always hold up well when tested experimentally (Gneezy et al. 2017), suggesting that further investigation is warranted. Also, note that a negative wage elasticity of labor supply can be rationalized by behavioral theory on loss aversion and reference dependence including Bell (1985), Gul (1991), Loomes and Sugden (1986), and Kőszegi and Rabin (2006).

[9] We are therefore consistent with a body of laboratory experiments confirming targeting behavior and loss aversion, such as Gneezy et al. (2017), Gill and Prowse (2012), and Ericson and Fuster (2011).

[10] In the volunteer context, the wage received *by the participant* is always equal to 0. However, our experimental notion of a volunteer wage involves the wage offered to a charitable organization, the ARC, for every unit of effort completed by the participant.

[11] Interestingly, recent studies do not find evidence of student selection in laboratory studies influencing the degree of prosocial behavior (Cleave et al. 2013, Abeler and Nosenzo 2015). However, a large empirical and theoretical literature, including recent field evidence in Ashraf et al. (2015), shows how selection can influence the extent to which individuals respond to incentives. Also, our thanks to anonymous referees for suggesting we further consider this possibility.

[12] In this paper, effort or labor supply should be understood as referring to the intensive margin, as our experimental variation does not allow for an explicit participation margin. However, as Fehr and Goette (2007) notes, the implications of reference-dependence for the extensive margin of labor supply are nuanced. For a summary of the theoretical implications of loss aversion for labor supply, as well as a review of the observational literature on targeting and labor supply, see Goette (2004). For an extension of the theory in Section 2 to include the extensive margin, see the theory appendix.

[13] Although not in the volunteering context, there is some related literature on targeting behavior with respect to charitable giving. For instance, Harbaugh (1998a, b) shows that donors may give amounts equal to the lower bound of a reporting bin.

[14] Such comparative literature includes Exley (2015), Bernheim and Exley (2015), Imas (2014), and Exley and Kessler (2017).

[15] In considering this study through the lens of effort provision, as in Abeler et al. (2011), we will use the framing of volunteering as opposed to charitable giving. In doing so, we follow previous laboratory studies on volunteer behavior, such as Ariely et al. (2009). Brown et al. (2018), in fact, show that within the laboratory context, participants respond very differently, indeed more generously, to volunteer frames (i.e., when exerting effort in a task to earn money for a charity) versus donating frames (i.e., when deciding how much to donate after earning money for themselves by exerting effort in a task). In considering time or effort an important feature of volunteering, it is also interesting to note that Craig et al. (2017) confirm in a field study that individuals are sensitive to the time costs of their giving.

[16] The overall monetary payments from the experiment are small and temporary, so the quasilinear specification ruling out income effects seems to be a reasonable approximation for our context.

[17] The quadratic specification for the cost of effort function is chosen for notational convenience only, although generalizing the convexity of the cost function would not qualitatively change the results in this section. By contrast, allowing for a nonzero intercept in the effort cost function does imply a nontrivial extensive margin choice for labor

supply. In the theory appendix, we discuss the details of a version of the model with participation costs and demonstrate that the essential targeting implications of the model remain unchanged.

[18] Note that if preferences α vary by context (working or volunteering) this may effect labor supply. Unsurprisingly, we do in fact later observe mean differences in effort by context, although such variation is not our focus.

[19] To simplify the resulting expressions for labor supply in the presence of loss aversion, we will actually multiply by 4 and add the term $4\mathbb{E}_{c,r}\mu(\alpha c - \alpha r)$ to preferences. This innocuous choice affects only the scaling of the units in which an agent's loss aversion parameter λ is expressed. In particular, inspection of the simplified payoffs for agents reveals that identical preferences can always be generated with a different multiple on gain-loss utility and appropriate re-normalization of the loss aversion parameter $\lambda \geq 1$.

[20] See Loomes and Sugden (1986), Shalev (2000), or Gul (1991) for other treatments of expectations-based endogenous reference points. Sugden (2003) and Farber (2008) are agnostic about the source of the reference point, and Masatlioglu (2005) together with Sagi (2006) consider the "status quo" as a reference point.

[21] In the theory appendix, we discuss the robustness of this figure's implications as the loss aversion parameter λ varies for an individual. Note that high enough levels of loss aversion lead to a two-segment labor supply function, for which targeting behavior occurs at all wages past a certain threshold. We view this result as qualitatively similar to the predictions of Figure 1 and hence omit it from the main discussion in the text.

[22] Note that Figure 1 plots the labor supply of a single agent with a fixed level of α and λ. Average labor supply across a large sample of agents, the outcome measured empirically, will reflect a smoothed version of Figure 1 given well behaved distributions of α and λ.

[23] Note that these expressions hold for the case $\lambda \in (1, \frac{3}{2})$. In the theory appendix, we discuss labor supply in the case that $\lambda \geq \frac{3}{2}$, where labor supply curves will instead consist of two segments and exhibit infinitely large targeting regions for any value of α.

[24] In fact, after a participant selects an envelope, the envelope is taped shut and the participant signs the envelope.

[25] This differs slightly from Abeler et al. (2011), who give the participants a total of three chances to solve a table correctly, after which the participants face a financial penalty if they still have not correctly solved a table.

[26] For instance, Falk and Ichino (2006) find that peer effects can lead to lower variance in behavior and higher productivity; Bernheim (1994) develops a theory where people care about others' perceptions of them; Andreoni and Bernheim (2009) show that people like to appear to be fair; Harbaugh (1998a, b), Bénabou and Tirole (2006), Ariely et al. (2009), and Exley (2017), among many other papers, show that people like to appear to be prosocial.

[27] All participants receive their earned payments from the practice round, and workers receive an additional compensation from their effort task. To ensure compensation across workers and volunteers are expected to be comparable, participants also receive their show-up fee of $20 if they are in the volunteering context or $13 if they are in the working context. The comparable effort in the working and volunteer context when the wage equals 25¢, as discussed later, helps to ease potential concerns related to this difference in show-up fees.

[28] A full distribution of labor supply is implied by theory, given a distribution of loss aversion, so the median regressions are independently interesting, and truncation of the tables completed at 0 from below suggests the use of a Tobit specification as a robustness check. Also, as a robustness check, we note that the dependent variable in our main specifications from Table 1 is a count variable, and Online Appendix Table A.1 contains the qualitatively similar results from

a negative binomial regression. The results are also robust to the use of the alternative outcome measures of time spent solving tables or acquired earnings, as shown in Online Appendix Tables A.2 and A.3. Note that when interpreting the alternative measures, the study was run using an online survey software called Qualtrics, which we have discovered measured time spent solving tables with some error. However, it is interesting to note that this measure indicates that the median time spent solving tables is 995 seconds, the median time spent solving the first table is 43 seconds, and the median time spent on the table where participants choose to instead stop is 7 seconds.

[29] The most comparable condition in Abeler et al. (2011) involves their treatment where participants' reference level is 35 tables since the wage rate is 20¢ and the reference payment is 7€. In this condition, 17% of their participants stop exactly at the reference level.

[30] It should also be evident from Figure 2 that in the 16¢-wage treatment participants are more likely to choose an effort level of 0 tables exactly. Although the baseline theoretical environment laid out in Section 2 implies strictly positive effort $e > 0$, the theory appendix extends the model to consider a nonzero fixed cost of participation. In this case, with a nontrivial extensive margin choice for labor supply, it is easy to show that lower wages predict more nonparticipation, although effort and targeting results conditional upon participation go through unchanged. Consistent with these predictions, as the wage increases in the lower panels of Figure 2 fewer participants choose to provide zero effort.

[31] As with the earlier tables, we obtain similar results when considering a negative binomial regression or alternative outcome measures of time spent solving tables or acquired earnings, as shown in Online Appendix Tables A.4, A.5, and A.6.

[32] The wage elasticity from 25¢ to 50¢ is also positive, and in the first column of Table 2, significantly so as we reject equality of coefficients on $I(w = \$0.25)_i$ and $I(w = \$0.50)_i$ ($p = 0.0704$).

[33] The median effort level for the \$0.16 wage is 36 tables and the median effort for the \$0.50 wage is between 45 and 50 tables. This calculation therefore uses the median effort for the \$0.50 wage as 47.50 tables, while the median regression output assumes 50 tables and would thus imply an increase of 56%.

[34] For example, the hourly wages reported by Farber (2008) for New York City taxi drivers, a population long-studied for evidence of targeting behavior, exhibit a standard deviation of around 20% relative to their mean. This empirical variation is smaller than the difference across our treatment levels of wages.

[35] As with the earlier tables, we obtain similar results when considering a negative binomial regression or alternative outcome measures of time spent solving tables or acquired earnings, as shown in Online Appendix Tables A.7, A.8, and A.9.

[36] In the first column of Table 3, we fail to reject the equality of coefficients on $I(w = \$0.50)_i$ and $I(w = \$0.80)_i$, with $p = 0.2660$.

[37] The median effort level for the \$0.25 wage is 32 tables and the median effort for the \$0.80 wage is between 13 and 14 tables. This calculation therefore uses the median effort for the \$0.80 wage as 13.50 tables, while the median regression output assumes 14 tables and would thus imply a decrease of 52%.

[38] This sample of 400 workers reflects us dropping 3 workers from the 403 workers who started our study. In particular, when these 3 workers (2 recruited via the Self Ad and 1 recruited via the Charity Ad) did not complete the study within the allotted two hours, three new "slots" became available and were completed by 3 new workers.

[39] These parameters allow participants to reach the reference level exactly and the reference level seems reasonable as Exley (2017) finds that 74% of mTurk participants are willing to solve seven similar-type questions to earn money for the ARC.

[40] Since we could not directly monitor participants time spent solving tables in the online study, we chose the limit of 100 tables instead of a time limit as in our laboratory study.

[41] For instance, variation in loss aversion or risk aversion across the working and volunteer contexts is possible, as is variation in the weight attached to any informational component of the reference point itself. Indeed, Bracha et al. (2015) document how effort responds to purely informational reference points about what other earns, and if adherence to norms is more likely in the volunteering than working context, as supported by the private treatment conditions in Bernheim and Exley (2015), the value placed on such information may be particularly strong in the volunteering context.

References

Abeler J, Nosenzo D (2015) Self-selection into laboratory experiments: Pro-social motives versus monetary incentives. *Experiment. Econom.* 18(2):195–214.

Abeler J, Falk A, Goette L, Huffman D (2011) Reference points and effort provision. *Amer. Econom. Rev.* 101(2):470–492.

Allen EJ, Dechow P (2013) The "rationality" of the long distance runner: Prospect theory and the marathon. Working paper, University of Southern California, Los Angeles.

Allen EJ, Dechow PM, Pope DG, Wu G (2017) Reference-dependent preferences: Evidence from marathon runners. *Management Sci.* 63(6):1657–1672.

Andreoni J (1989) Giving with impure altruism: Applications to charity and ricardian equivalence. *J. Political Econom.* 97(6): 1447–1458.

Andreoni J (1990) Impure altruism and donations to public goods: A theory of warm-glow giving. *Econom. J.* 100(401):464–477.

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

Ariely D, Bracha A, Meier S (2009) Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *Amer. Econom. Rev.* 99(1):544–555.

Ashraf N, Bandiera O, Lee SS (2015) Do-gooders and go-getters: Career incentives, selection, and performance in public service delivery. Working paper, London School of Economics, London.

Bell D (1985) Disappointment in decision making under uncertainty. *Oper. Res.* 33(1):1–27.

Bénabou R, Tirole J (2003) Intrinsic and extrinsic motivation. *Rev. Econom. Stud.* 70(3):489–520.

Bénabou R, Tirole J (2006) Incentives and prosocial behavior. *Amer. Econom. Rev.* 96(5):1652–1678.

Bernheim BD (1994) A theory of conformity. *J. Political Econom.* 102(5):841–877.

Bernheim BD, Exley CL (2015) Understanding conformity: An experimental investigation. Working paper, Stanford University, Stanford, CA.

Bracha A, Gneezy U, Loewenstein G (2015) Relative pay and labor supply. *J. Labor Econom.* 3(2):297–315.

Brown AL, Meer J, Williams JF (2018) Why do people volunteer? An experimental analysis of preferences for time donations. *Management Sci.* Forthcoming.

Camerer C, Babcock L, Lowenstein G, Thaler R (1997) Labor supply of New York City cabdrivers: One day at a time. *Quart. J. Econom.* 112(2):407–441.

Carpenter J, Myers CK (2010) Why volunteer? Evidence on the role of altruism, image, and incentives. *J. Public Econom.* 94(11–12): 911–920.

Chou YK (2002) Testing alternative models of labour supply: Evidence from taxi drivers in Singapore. *Singapore Econom. Rev.* 41(1):17–47.

Cleave BL, Nikiforakis N, Slonim R (2013) Is there selection bias in laboratory experiments? The case of social and risk preferences. *Experiment. Econom.* 16(3):372–382.

Craig AC, Garbarino E, Heger SA, Slonim R (2017) Waiting to give: Stated and revealed preferences. *Management Sci.* 63(11): 3672–3690.

Crawford VP, Meng J (2011) New York City cab drivers' labor supply revisited: Reference-dependent preferences with rational expectations targets for hours and income. *Amer. Econom. Rev.* 101(5):1912–1932.

DellaVigna S, Lindner A, Reizer B, Schmieder J (2017) Reference-dependent job search: Evidence from Hungary. *Quart. J. Econom.* 132(4):1969–2018.

Doran K (2014) Are long-term wage elasticities of labor supply more negative than short-term ones? *Econom. Lett.* 122(2):208–210.

Ericson KMM, Fuster A (2011) Evidence on reference-dependent preferences from exchange and valuation experiments. *Quart. J. Econom.* 126(4):1879–1907.

Exley CL (2015) Excusing selfishness in charitable giving: The role of risk. *Rev. Econom. Stud.* 83(2):587–628.

Exley C (2017) Incentives for prosocial behavior: The role of reputations. *Management Sci.*, ePub ahead of print March 24, https://doi.org/10.1287/mnsc.2016.2685.

Exley CL, Kessler JB (2017) The better is the enemy of the good. Working paper, Harvard University, Cambridge, MA.

Falk A, Ichino A (2006) Clean evidence on peer effects. *J. Labor Econom.* 24(1):39–57.

Farber HS (2005) Is tomorrow another day? The labor supply of New York City cabdrivers. *J. Political Econom.* 113(1):46–82.

Farber HS (2008) Reference-dependent preferences and labor supply: The case of New York City taxi drivers. *Amer. Econom. Rev.* 98(3):1069–1082.

Farber HS (2015) Why you can't find a taxi in the rain and other labor supply lessons from cab drivers. *Quart. J. Econom.* 130(4):1975–2026.

Fehr E, Goette L (2007) Do workers work more if wages are high? Evidence from a randomized field experiment. *Amer. Econom. Rev.* 97(1):298–317.

Frey BS, Jergen R (2001) Motivaiton of crowding theory. *J. Econom. Surveys* 15(5):589–611.

Frey BS, Oberholzer-Gee F (1997) The cost of price incentives: An empirical analysis of motivation crowding-out. *Amer. Econom. Rev.* 87(4):746–755.

Gill D, Prowse V (2012) A structural analysis of disappointment aversion in a real effort competition. *Amer. Econom. Rev.* 102(1):469–503.

Gneezy U, Rustichini A (2000) Pay enough or don't pay at all. *Quart. J. Econom.* 115(3):791–810.

Gneezy U, Meier S, Rey-Biel P (2011) When and why incentives (don't) work to modify behavior. *J. Econom. Perspect.* 25(4):191–210.

Gneezy U, Goette L, Sprenger C, Zimmermann F (2017) The limits of expectations-based reference dependence. *J. Eur. Econom. Assoc.* 15(4):861–876.

Goette L (2004) Loss aversion and labor supply. *J. Eur. Econom. Assoc.* 2(2–3):216–228.

Goette L, Stutzer A, Frey BM (2010) Prosocial motivation and blood donations: A survey of the empirical literature. *Transfusion Medicine and Hemotherapy* 37(3):149–154.

Gregg P, Grout PA, Ratcliffe A, Smith S, Windmeijer F (2011) How important is pro-social behaviour in the delivery of public services? *J. Public Econom.* 95(7–8):758–766.

Gul F (1991) A theory of disappointment aversion. *Econometrica* 59(3):667–686.

Harbaugh WT (1998a) The prestige motive for making charitable transfers. *Amer. Econom. Rev.* 88(2):277–282.

Harbaugh WT (1998b) What do donations buy? A model of philanthropy based on prestige and warm glow. *J. Public Econom.* 67(2):269–284.

Horton JJ, Rand DG, Zeckhauser RJ (2011) The online laboratory: Conducting experiments in a real labor market. *Experiment. Econom.* 14(3):399–425.

Imas A (2014) Working for the "warm glow": On the benefits and limits of prosocial incentives. *J. Public Econom.* 114(June):14–18.

Karlan D, List JA (2007) Does price matter in charitable giving? Evidence from a large-scale natural field experiment. *Amer. Econom. Rev.* 97(5):1774–1793.

Kőszegi B, Rabin M (2006) A model of reference-dependent preferences. *Quart. J. Econom.* 121(4):1133–1165.

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

Lacetera N, Macis M, Slonim R (2014) Rewarding volunteers: A field experiment. *Management Sci.* 60(5):1107–1129.

Larkin I (2014) The cost of high-powered incentives: Employee gaming in enterprise software sales. *J. Labor Econom.* 32(2):199–227.

Loomes G, Sugden R (1986) Disappointment and dynamic consistency in choice under uncertainty. *Rev. Econom. Stud.* 53(2): 271–282.

Masatlioglu Y (2005) Rational choice with status quo bias. *J. Econom. Theory* 121(1):1–29.

Meer J (2011) Brother, can you spare a dime? Peer pressure in charitable solicitation. *J. Public Econom.* 95(7):926–941.

Meier S, Stutzer A (2008) Is volunteering rewarding in itself? *Economica* 75(297):39–59.

Mellström C, Johannesson M (2008) Crowding out in blood donation: Was titmuss right? *J. Eur. Econom. Assoc.* 6(4):845–863.

Musick MA, Wilson J (2007) What is volunteering? *Volunteers: A Social Profile* (Indiana University Press, Bloomington), 11–35.

Null C (2011) Warm glow, information, and inefficient charitable giving. *J. Public Econom.* 95(5–6):455–465.

Oyer P (1998) Fiscal year ends and nonlinear incentive contracts: The effect on business seasonality. *Quart. J. Econom.* 113(1):149–185.

Paolacci G, Chandler J, Ipeirotis PG (2010) Running experiments on amazon mechanical turk. *Judgment Decision Making* 5(5): 411–419.

Pope DG, Schweitzer ME (2011) Is tiger woods loss averse? Persistent bias in the face of experience, competition, and high stakes. *Amer. Econom. Rev.* 101(1):129–157.

Sagi JS (2006) Anchored preference relations. *J. Econom. Theory* 130(1):283–295.

Sawhill J, Williamson D (2001) Measuring what matters in non-profits. *McKinsey Quart.* (May), https://www.mckinsey.com/industries/social-sector/our-insights/measuring-what-matters-in-nonprofits.

Shalev J (2000) Loss aversion equilibrium. *Internat. J. Game Theory* 29(2):269–287.

Sugden R (2003) Reference-dependent subjective expected utility. *J. Econom. Theory* 111(2):172–191.

Terry S (2017) The macro impact of short-termism. Working paper, Boston University, Boston.

The Case Foundation (2015) 2015 Millenial impact report: Cause, influence, and the next generation workforce. The Case Foundation, Washington, DC.

Titmuss RM (1970) *The Gift Relationship* (Allen and Unwin, London).