Ex-ante Commitments to "Give if you Win" Exceed Donations After a Win*

Christian Kellner[a]
David Reinstein[b]
Gerhard Riener**[c]

[a] *University of Southampton*
[b] *University of Exeter*
[c] *Heinrich Heine University Düsseldorf*

December 9, 2018

Abstract

Should fundraisers ask a banker to donate "if he earns a bonus" or wait and ask after the bonus is known? Standard EU theory predicts these are equivalent; loss-aversion and signaling models predict a larger commitment *before* the bonus is known; theories of affect predict the reverse. In five experiments incorporating lab and field elements (N=1363), we solicited charitable donations from small lottery winnings, varying the conditionality of donations between participants. Pooling across experiments, participants are 23% more likely to commit to donate from the winning income and commit 25% more when asked *before* the lottery's outcome is determined—relative to those asked to donate *after* they learn they have won. These differences are strongly statistically significant. This represents the first evidence on how pro-social behavior extends to conditional commitments over uncertain income, with implications for charitable fundraising, giving pledges, and experimental methodology.

Keywords: Social preferences, contingent decision-making, signaling, field experiments, charitable giving.

JEL codes: D64, C91, C93, L30, D01, D03, D84.

* Acknowledgments: We would like to thank the Universities of Essex, Düsseldorf, Mannheim and Bonn for providing financial support for our research. We would also like to thank the many colleagues who have offered us valuable advice and comments, including Johannes Abeler, Michelle Brock, Gary Charness, Miguel Fonseca, Jean Hindriks, Brit Grosskopf, Zack Grossman, Sandra Ludwig, Friederike Mengel, Graeme Pearce, Timothy Rakow, Uri Simonsohn, Paul Smeets, and Jeroen van de Ven, and conference and seminar participants at the Advances in Field Experiments, the British Academy, CAGE/CMPO, Einaudi, ESA, Essex, Exeter, FUR, J-LAGV, London Experimental Workshop, M-Bees, Sheffield, the Science of Philanthropy Initiative, SUNY Albany, and UCSB. We are also grateful to Samuel Dexter, Edward Dickerson, Saga Eriksson, Jonathan Homola, Louis Philipp Lukas, Kajetonas Orlauskas, Stavros Poupakis, and Agata Siuchninska for excellent research assistance. This paper benefited strongly from detailed comments from JPubE referees and from Co-editor Keith Marzilli Ericson.

** Corresponding author (riener@uni-duesseldorf.de).

1 Introduction

Most research on other-regarding behavior considers choices under certainty. However, decisions in this domain often involve risk, uncertainty, and contingent plans. We provide unique evidence on how other-regarding behavior extends to income uncertainty and to contingent commitments. This is motivated by a particular practical question: what is the best way to ask for a charitable donation from an individual who may get an uncertain bonus income? Should you ask her *before*—to make a contingent commitment to donate if she wins the bonus or ask her *after*—to donate after her bonus has been revealed?

There are important differences between these two modes of asking which may impact behavior: (i) *Before* commitments are from uncertain income. (ii) *Before* commitments to donate are not realized with certainty. (iii) *After* commitments follow an experience of winning. If she is a standard expected utility maximizer only caring about outcomes, this will not matter. In contrast, certain theories of *affect* predict that she will donate more *after* winning. However, we show that under particular specifications, loss-aversion and signaling models predict a larger commitment for giving *conditionally*, in advance of learning the outcome. This latter prediction is largely substantiated by our evidence from a series of lab and web-based experiments, discussed below.

This is an important issue for policymakers and fundraisers. Many employees receive windfall payments in supplement to their regular income. In the 2011/12 tax year, bonuses to UK workers totaled £37 billion, of which £13 billion was in the financial sector, at an average rate of £12,000 per worker (ONS, 2012). In the USA, Wall Street banks distributed $26.7 billion in bonuses in 2013 (NY Comptroller, 2014). Anecdotal evidence suggests that a significant share of this bonus income was not fully anticipated.[1] In the wake of recession and scandals in the financial markets, bankers have been encouraged to give back their bonuses, or donate them to charity.[2] Our evidence suggests it may be more effective to ask bankers to commit to donate from *future* bonuses. Moreover, this question is relevant to situations in which individuals are asked or volunteer to donate from actual or potential financial gains. Lottery organizers may include a tick-box to make a conditional pledge. Ethical investment accounts could automatically donate gains that exceed expectations.[3] This is not merely hypothetical: several prominent movements and organizations ask students, workers, and entrepreneurs to publicly pledge a share of their future income and profits.[4]

This also has implications for experimental methodology, in particular, the random lottery incentive scheme, where only one stage is chosen randomly for payment (see Cubitt et al., 1998). Particularly where signaling is relevant, subjects may not treat each stage independently, and may integrate their choices into a global decision frame.

1. From our personal correspondence: "Most people at the top or the bottom of the performance level will know they're (not) getting a bonus—people in the middle will be unsure until they're announced. Among the people who know they're going to get a bonus, the size of the bonus is uncertain until announced.", Raj C: Hedge Fund Manager, London (2015). See also forum posts <*http://www.quora.com/Bonuses/How-accurately-can-an-employee-predict-his-or-her-annual-bonus-in-advance-e-g-in-the-banking-industry*>, accessed 7 Feb, 2015.

2. "Johnson: Bankers should assuage guilt by giving bonuses to homeless scheme." – *The Guardian*, 13 Feb 2009.

3. E..g., Triodos Bank offers a "Save and Donate" account <http://www.triodos.co.uk/en/personal/savings-overview/charity-saver/>, accessed 29 Sep. 2017; however this currently involves *fixed* interest rates and a fixed donation share.

4. "Giving What We Can," founded by Tony Ord, asks people to make a *giving pledge* to donate roughly 10% of their future income. According to their website <*https://www.givingwhatwecan.org/about-us/history/*> *(accessed 25 Sep 2017*) they have over 3000 members who have donated over $24 million (and pledged to donate far more); a large share of whom are students, who presumably face great uncertainty over their lifetime earnings. Thelifeyoucansave.org, promotes a smaller-percentage pledge with rates adjusted by income and by country. The Founders pledge (https://founderspledge.com/about-us , accessed 12 Sept. 2016) asks tech entrepreneurs who have yet to 'cash out' to make a legally-binding commitment to donate 2% or more of their potential proceeds to a social cause of their choice. Motivated by our research, the London-based City Philanthropy recently held a "Bonus pledge think tank" to explore this idea (See: http://www.cityphilanthropy.org.uk/events/1-bonus-pledge-think-tank and http://www.cityphilanthropy.org.uk/news/call-city-firms-help-cabinet-office-research-%E2%80%98windfall%E2%80%99-giving, accessed 12 Sept. 2016).

To the best of our knowledge, there is no direct economic evidence on the effect of the resolution of income uncertainty on other-regarding behavior. Tonin et al. (2014) and Reinstein (2010) each ran experiments involving charitable donation in uncertain environments, where subjects knew that only one of a series of decisions would be implemented; both found that donations declined over time. According to Reinstein "if individuals are not [expected utility] maximizers over outcomes but gain warm glow utility from unrealized commitments, this decline could be attributed to satiation of warm glow". In laboratory dictator games Brock et al. (2013) and Sandroni et al. (2013) each found that social preferences and fairness concerns appear to depend on a combination of *ex post* and *ex ante* concerns. Smith (2011) found that giving (to other subjects who had incurred an income loss) was higher when giving decisions were made using the strategy method than when subjects were asked ex-post. These results argue against a model where an individual maximizes expected utility with a consistent utility function that considers only outcomes.

Grossman (2015), focuses on measuring and differentiating self and social signaling, and varies the probability that a subject's chosen "gift" is implemented.[5] His laboratory experiments (with a standard student sample) involved binary-choice dictator games where an individual's choice may be randomly overruled with a given probability. The treatments varied this probability, and whether the "observer" (another subject or the experimenter) saw the outcome, the choice, and the probability of overrule. In general, he found that the probability that the choice was overruled had neither a large nor a significant effect on choices in *any* of the observability conditions. Our approaches differ substantially (from Grossman as well as from Andreoni and Bernheim); in particular, our focal treatment involves a conditional commitment from uncertain *income*.[6]

Generosity involving *unconditional future* commitments is a distinct but related issue. Breman (2011) ran a field experiment asking donors to commit to increase their regular donation either immediately, in one months time, or in two months time. She found the longest delay led to the greatest increase in contributions. Her explanation is that the cost of giving occurs at the time of payment, while "the warm-glow … will be experienced at the time of committing to giving."[7] Commitments in Breman's experiment could be reversed but they rarely were. While Andreoni, Serra-Garcia (2016) also find greater "Give-Later" commitments (in longitudinal laboratory experiments), they find frequent reneging on previous pledges, as well as heterogeneous dynamic inconsistency.

Our evidence abstracts from the issue of *delay*: in our experiments the uncertainty is resolved almost immediately and there is no difference in when the donations are realized. However, in real-world fundraising applications (e.g., in the bankers' bonuses context), asking for conditional donations of uncertain income is likely to also mean asking for a *delayed* donation.

Our paper and experiments do not directly consider *unconditional* donation choices made before uncertain income is resolved. There are several arguments for allowing people to vary their donation according to their

5. (Andreoni, Bernheim, 2009 similarly vary this probability, but their experiment is strongly focused on "audience effects" and involves a known recipient observer throughout.

6. Our experiments, described in section 3, differ from Grossman's along several other dimensions (binary vs. continuous choices, different sets of probabilities, with/without a real-effort task, session-level versus within-session variation, oral versus computer instructions), and we also provide evidence from several field contexts. For our context that most resembles his—the laboratory Uncertain Collection treatment—we also find a null result, reported in table D.11. In general, for our experiments involving *students*, we find small effects, which a standard laboratory sample size would have limited power to detect.

7. Breman draws on Thaler et al. (2004), whose "Save More Tomorrow" experiment found that individuals save more when asked to pre-commit a portion of future pay raises towards retirement savings. She extends the "pre-commitment for the future" part of their treatment to the charitable domain; our experiments extend the *commit uncertain salary increases* effect (which the authors argue is driven by loss aversion). These results largely support Andreoni, Payne (2003), who write that "a commitment to a charity may yield a warm-glow [benefit] to the givers before ... the costs are paid". This raises the question "when does the benefit of giving occur and how long does it last?" By this logic we might expect to see charitable giving exclusively through end-of-life bequests, which would yield warm glow that could be savored over one's entire life. However, bequest giving is rare (Cabinet Office, 2011, Giving White Paper, HM Stationary Office.)

realized income, either by asking them after the uncertainty has been resolved or by allowing them to make a conditional commitment. Individuals might be more generous if they can hedge in this way, effectively smoothing their consumption. Note also that standard optimization predicts that people are better off when they can make decisions after uncertainties are resolved. This might offer people a reason to delay making a commitment whenever they are asked to donate and facing uncertain income; such delays and "transactions costs" are seen to reduce contributions (Knowles et al., 2015). As most people face a lifetime of financial uncertainties, a fundraiser may thus be tempted to ask people to commit to making donations *conditionally* on certain financial outcomes.

However, conditional commitments do involve a sort of uncertainty. Some previous work suggests that uncertainty and ambiguity about recipients and the outcomes of donations may lead to more self-interested behavior (e.g., Brock et al., 2013, Small et al., 2003). In particular Exley's (2016) experimental subjects make several series' of choices between certain and uncertain payoffs for oneself and a charity. Her results suggest that where subjects must choose between lotteries/certainties for themselves and a charity, they practice *motivated reasoning* (see Gino et al., 2016), overweighting or underweighting the probability of a loss in order to justify less-generous behavior. Our evidence suggests that this "excuse-driven" response to risk does not extend to conditional commitments (or at least, this does not outweigh the other factors favoring the Before mode). Indeed, a simple extension of Exley's model predicts that conditional donations should *not* be vulnerable to motivated reasoning over probabilities. In considering a "give if you win" commitment, the tradeoff between one's own and the charity's income is not affected by the probability of a win.

We present the results of a series of five experiments in distinct contexts, with complementary strengths, each combining lab and field characteristics in a different way. Our experiments offer the first systematic insight into contingent giving from known or uncertain income. To avoid contrast and experimenter-demand effects, all experiments involve only between-subject treatment variation and only a single charitable ask; thus we use large samples to detect moderate-sized effects. Table 1 in section 3 offers a brief summary of the differences between our five experiments.

Our laboratory experiments offer a more controlled environment, where we can be sure subjects are making decisions on their own, they cannot communicate, and they can tangibly verify that outcomes are randomly determined (see Maniadis et al., 2014). Our web-based evidence offers environmental validity and is less prone to experimenter demand. Our lab experiments varied the presentation of the earnings as random (the "bonus" being awarded with 50% probability) and the timing of the contribution request. Depending on the treatment, we observed conditional pre-commitments for (losing and) winning states or decisions after winning (or losing) a lottery. While the lab experiment involved two levels of earnings and five treatments, we focus on the two treatments that most parallel those in our web-based experiments.[8]

In all of the experiments the results are in the same direction: people committed to donate more when asked immediately *before* they knew if they had won. Although these differences are statistically significant in only some of these experimental contexts (two at $p<0.01$ and one at $p<0.10$), they are strongly significant when we pool across all experiments. Overall, conditional donations ("if you win") were 25% higher (3.8 percentage points higher) in the *Before* treatments (Table 4). The effect was particularly strong and significant in our

8. Given the large variance in donation choices we had limited power to detect moderate-sized differences among all of these. Thus, we only report them in the appendix for completeness. The other three treatments were the following. *Before Both*: a separate ex-ante decision for losing and winning states. *Uncertain Collection*: the income was known and certain, but there was only a 50% chance that a committed donation was to be collected. *Benchmark*: either low or high income, but with no income uncertainty. Our results are very similar if we pool this *Benchmark* treatment with the *After* treatment reported below, and also similar (but stronger) if we pool the *Before* and *Before Both (if win)* treatments.

web-based experiment using non-student participants: giving nearly doubled, and its incidence increased by 50%. Further modelling suggests a non-linear relationship: the *Before* treatment has a larger effect for those predicted (by pre-determined observables like age and gender) to donate less.

2 Basic Setup and Predictions

In this section we offer a theoretical perspective on giving when income is uncertain. Consider two income levels, w and ℓ, where $w > \ell \geq 0$. The decision maker knows she is facing a lottery with a non-degenerate probability p of winning w and a probability $1-p$ of losing and earning ℓ. Consider the following settings. Suppose, at least for now, that a donation is only feasible or reasonable after *winning* the lottery.

***After* setting (A):** Income is unknown until a lottery is resolved, and she learns whether she has won or lost the lottery. If she has won, she learns her income and then is immediately asked to donate to a specific charity. She donates $g^a \geq 0$.[9]

***Before* setting (B):** Income is unknown until a lottery is resolved. Before she learns the outcome, she is asked to make a binding commitment to donate to a specific charity *if she wins w*. She commits to give $g^b \geq 0$ if she wins (if she loses nothing is collected).

Our main question is: "how does her commitment or pledge in the *Before* setting, to donate if she wins, compare to her donation in the *After* setting when she has already won?", i.e. "what is the relationship between g^b and g^a?" (To motivate future work, we define and derive predictions for additional settings and for donations from the lower income in Appendix A.)

2.1 Expected utility maximization over outcomes

In most previous models of charitable giving, only an individual's *realized* contribution (and consumption) affects her utility. She may care about the total amount of the public good provided (Becker, 1974), she may gain warm glow from the amount of her own income she has actually given away (one interpretation of Andreoni, 1990), she may care about her impact on outcomes (Duncan, 2004) or on an individual she identifies with (Atkinson, 2009), or she may receive a prestige benefit from observed realized donations (Harbaugh, 1998). Although these theoretical papers generally do not consider uncertain environments, they have been applied to such contexts (e.g., DellaVigna et al. 2012; Sandroni et al. 2013; Vesterlund 2003, as well as in numerous laboratory experiments) using the expected utility framework.[10] For any model that can be expressed in terms of expected utilities over outcomes the timing and uncertainty of the decision (i.e., whether it is a sure thing or a prospect) is irrelevant to the individual's choice. This is stated in prediction 1 and is trivially proven in Appendix A.1.

2.2 Signaling and self-signaling

When signaling or self-signaling is considered as a motive for donations (as in Benabou et al., 2006; Grossman, 2015; Tonin et al., 2014), utility may not only depend on outcomes, but also on beliefs. Hence, the utility

9. If she loses, she may also be asked to give. However, in three of five of our experiments we did not ask losers to donate, and we have limited statistical power our power to make relevant comparisons here. Thus we focus on the (conditional) donations from the "winning" income g^t for treatments $t \in \{a, b\}$. We consider the donations from the losing income (labeled g^t_ℓ) in the appendix and very occasionally below.

10. Other non-EU *procedural*, such as reciprocity (Sugden, 1984) and the *Kantian* motive (Roemer, 2010; Sugden, 1982) have also been modeled solely in terms of *actual* donations, and do not have an explicit role for unrealized commitments. Some more recent work has argued that intentions and commitments may yield direct utility and signaling value; we return to this below.

obtained in a given state of the world may depend on unrealised donation commitments from the past, as such commitments may still have an impact on beliefs. Also, the findings of Sandroni et al., 2013 could be understood as evidence for the presence of (self-)signaling or other non-outcome based motives.[11]

To analyse this point, we offer a simple signaling model (see Appendix A.2 for derivations and extensions to additional settings) with two types of agents (or two types of self): one who gets an inherent benefit from donating to the charity (a "good type"), and one who does not (a "bad type").[12] We demonstrate that uncertain collection of committed donations can lead to larger (conditional) donations. We focus on parameter values where, at the good types' preferred donation (ignoring signaling), the bad types have an incentive to pool to gain reputation; i.e., "the separation constraint binds" (see appendix: condition 4). Here, as the probability of collection decreases below one, the level of conditional-on-collection donations that can be sustained as an equilibrium satisfying the intuitive criterion increases. Essentially, as the intent to donate can still be demonstrated, while the cost of actually donating will only be paid with a probability less than one, the (minimum) conditional-on-collection donation must increase in order to separate types.

If an individual only recalls his or her own true type with error, the signaling model can be considered as an equivalent *self*-signaling model, as noted in Benabou et al. (2006; 2011).

Prediction 1. *Signaling generosity, where the separation constraint binds*

$g^b > g^a$*, for good types, while bad types are unaffected by the treatment.*

2.3 Loss Aversion and Reference Points

When making—even riskless—choices, it is often argued that decisions are influenced by anticipated gains and losses relative to a reference point (see Tversky et al., 1991). Thaler et al. (2004) claim "… once households get used to a particular level of disposable income, they tend to view reductions in that level as a loss." In considering this model, we suppose the individual has a reference point over her own consumption, not including charitable giving, and her utility function sums a standard reference-independent term and a *gain-loss* component. Her donation decision, whether stochastic or certain, anticipates how the donation will reduce the remaining wealth available for her own consumption. If this will fall below her reference point, she will incur a psychological loss. We assume there is no gain-loss utility over the donation itself (i.e., a single target, as in Camerer et al., 1997).[13]

11. Nearly a third of Sandroni et al. (2013)'s dictator subjects demonstrated a strict preference for a coin flip between increasing their own and another's payoff, preferring this to getting either alternative with certainty. This choice may be driven by a diminishing returns to both private consumption and the benefits of the pro-social choice, (e.g., self-signaling, impact, or warm glow), if a *commitment* to donate with some probability *itself* yields these benefits.

12. This model is distinct from Grossman (2015). He models a continuum of types with binary choices, where the outcome is not entirely deterministic: either choice may be overruled by nature with a known probability. He further solved for cases varying the observer's information about the choice and environment. As in our model, both the signaling value and the material cost of a particular donation increase in the probability the donation is realized. In his model, where the observer sees only the choice and not the probability, a donation commitment (of a specified size) is more likely where the donation is collected with a lower probability, because signaling is "cheaper". However, in considering a varying probability of realization, Grossman only compares environments where the externally-observed probability of realization—and thus the external signaling value—is constant. He does not model a case where the observer (or future self) sees both the choice and the probability, as in our model, and these vary together. This case is particularly relevant to donations from uncertain income. For the "after" ask, the probability of realization is one, and this is common knowledge to the fundraiser and potential donor. For the "before ask" we consider, it is common knowledge that this probability is less than one. However, in the real world, the exact probability of a bonus may not be common-knowledge; the impact of this may be considered in later work.

13. This may hold if donating nothing and using all income for own-consumption is typically seen as the default, thus the basis for a reference point. Note that this model's predictions would be qualitatively the same if there were two targets, but the gain-loss utility were far more salient for consumption targets than for giving targets.

While the reference point may change over time, we assume here that she is *myopic* in the sense that when making a decision she does not anticipate these changes. For simplicity, we consider a utility function embodying a linear loss function, i.e.,

$$v(x,g,\pi) = \begin{cases} u(x) + \omega(Dg) & \text{if } x \geq \pi \\ u(x) - \delta[u(\pi) - u(x)] + \omega(Dg) & \text{if } x < \pi; \end{cases}$$

subject to the budget constraint $x + g \leq E$,

where x represents own consumption, g is the committed donation expenditure and D indicates whether it is realized, π is a reference point, specified below, and δ is a (strictly) positive constant. Here $u(\cdot)$, the sub-utility of own-consumption, is assumed to be strictly increasing and concave, as is the "warm glow" function $\omega(\cdot)$. All derivations are in the appendix.

Suppose the reference point always corresponds to the expected future income at the point of the decision, the maximum own-consumption one could achieve if one's investments paid their expected value. This implies that contributions incur a *loss* in all cases expect for conditional contributions from an expected win, implying greater contributions in the *Before* than in the other treatments, i.e.:

Prediction 2. *Loss Aversion, expected income, immediate adjustment*

$$g^b > g^a, \text{ provided } g^b < w - (pw + (1-p)\ell)),$$

recalling that w and ℓ were the winning and losing incomes, respectively.

This prediction extends to alternate assumptions over the reference consumption basket.[14] However, it does not apply to *any* model with loss-aversion; it depends on several key assumptions. It is crucial that in the *After* treatment the reference point for own consumption corresponds to the (new) expected wealth, or adjusts at least partly to the realized income. If we assume the reference point is unchanged throughout the relevant decision period, the *Before* and *After* donations will be equivalent. We might alternatively allow reference points and gain-loss utility for *both* own-consumption and charitable giving, and also allow these to adjust to the planned donation after the income is realized. Then, under a *preferred personal equilibrium* (Koszegi et al., 2006) even loss-averse individuals may choose $g^a = g^b$ (see online appendix B.1).

2.4 Affective state (unanticipated) and generosity

A favorable realization of a lottery may put people in a good mood, while an unfavorable outcome may do the opposite. Theories and evidence on the interaction of affective state and generosity point to more giving when an individual is in a positive mood (Levin et al., 1975; Weyant, 1978; Underwood et al., 1976; Kidd et al., 2013; Drouvelis et al., 2016). On the other hand Fishbach et al. (2007) offer mixed evidence, and Kuhn et al. (2011), find "greater lottery winnings do not raise the likelihood that a household will donate its fee for completing our survey to charity". Putting this together we might predict greater generosity *after* a prize has

14. The reference consumption basket might be based on her expectations before being asked to donate, thus deducting no donation; alternately, it may have anticipated a small probability of an ask, or it might immediately subtract the expected value of the conditional donation after the ask. For any of these the reference consumption is still less than the higher earnings w, and the above prediction will attain. Note that we can consider the "ask" as a special shock motivating giving by changing the environment/context or temporarily adjusting the utility function to make the utility slope of giving particularly steep; see Andreoni, Rao (2011) and models in Reinstein (2011) and Kotzebue et al. (2009). Hence we may predict individuals give a larger share of their experimental "winnings" than the normal share of their income that they donate.

been won, relative to *before* the prize outcome is known, provided individuals do not anticipate their mood changes when making conditional donations.[15]

Prediction 3. *Affective state:* $g^a > g^b$

2.5 Further alternative explanations

Participants may have non-standard beliefs about probabilities and randomness. In particular, they may believe that their commitment to contribute will increase their likelihood of winning. This may stem from "magical thinking", an illusion of control (see seminal article by Langer, 1975, and the literature following it) or exhibiting "just world beliefs" (Rubin et al., 1975). An individual who believes in Karma (cf. Levy et al., 2006) may believe she will be rewarded for good acts (or good commitments) and punished for bad ones.[16] While we can not rule this out, we emphasize in each experiment that stochastic outcomes have been determined by random draw *prior* to their donation choices. We also differentiate our results by measures of stated religious affiliation, finding no significant differences (however, our sample yielded limited power to detect an effect). Several other behavioral models and concerns could also predict donation behavior distinct from the standard expected utility model, including adaptation, tangibility, present-bias, a status-quo reference point, and uncertainty aversion. We discuss these in the online appendix, arguing that these are less relevant than the models presented above, and are not supported by our evidence.

3 Designs and Implementation

We ran a series of five experiments across over different contexts, demographics, framings, and rules. For transparency, we report on *all* of the experiments that we ran as part of this project.[17] We present and analyze (i) the entire project and (ii) the two experiments that we preregistered with the AEA RCT registry, declaring our study parameters, hypotheses, and analysis plan.[18]

The contextual variations impart a benefit. Gathering evidence over domains varying in the "distribution of the characteristics of the units" (participant demographics) and the "specific nature of the treatments [and] treatment rate" helps us to explore the "credibility of generalizing to other settings", i.e., the sensitivity of the results to the context, the framing, and the subject pool (Athey et al., 2017).

Our lab experiment (involving a field commodity: the charitable donation) permits greater control and design flexibility. The four web-based experiments have many of the usually cited advantages of field experiments (see Harrison et al., 2004). In particular, they offer less risk of experimenter demand effects and reflect more naturalistic, less self-conscious behavior: participants were unlikely to know they were participating in an experiment, and still less likely to consider that they were in an experiment focused on charitable giving. There is also strong environmental validity. These experiments resemble real-world contexts that participants

15. Similar predictions could arise out of an (indirect) reciprocity model (see Simpson et al., 2008), e.g., if the lottery's sponsor were the charity itself, or were believed to be sympathetic to the charity; the reciprocity motive would also have to depend on the realization of the "gift" and not only its probabilistic implementation.

16. Participants may donate more before if they believe that a spiritual force affects their winning probabilities; but it is not clear whether in the *Before* treatment she will give conditional or unconditionally. She may want to appease the gods by saying "I will donate anyway," or she may want to give them an incentive to make her a winner by making her donation conditional on a win. Similarly, she may donate more *After* out of a sense of gratitude towards this spiritual force. (As this is difficult to pin down, we did not include it in the table.)

17. Chronologically, we ran the experiments in the reverse order presented below; we discuss the logic of this sequence in the appendix section C.1. We exclude one experiment we ran at a Cat Fair in 2010 which we ended after about 12 observations, after it became clear that this was an environment unlikely to provide much variation in donation behavior.

18. As noted in Button et al., 2013, "pre-registration clarifies whether analyses are confirmatory or exploratory, encourages well-powered studies and reduces opportunities for non-transparent data mining and selective reporting."

may be accustomed to: universities often run employability promotions, researchers run broad surveys and often give participants the option to donate their earnings, many promotions involve a prize lottery, and web sites often ask for donations.

Table 1 presents a summary of the experiments highlighting the most important differences between our experiments. These variations were either part of our initial design (e.g., sensitivity to a non-student sample, and to a varying the probability of realization), responses to referees' suggestions (e.g., the opportunity to reverse a *Before* commitment), or reflect feasibility concerns and calibration adjustments to achieve an intermediate baseline rate of contributions.

However, key elements were the same across experiments. Each experiment involved very similar *Before* and *After* treatments, resembling the settings discussed in section 2; each participant made this charitable decision (at most) once, with between-subject variation. The chance to win a prize was always tied to participation in an encouraged activity or performing a task. Uncertainty over the prize was always resolved immediately after the *Before* donation choice or immediately before the *After* donation choice. All charitable "asks" required participants to make an active decision, with a choice architecture that weakly suggested donating. We always used well-known charities including an international-poverty-related charity, and we always provided a means to verify that donations were actually passed to the charities. For all experiments other than *Valentine's*, participants could choose among two chats, committed donations were automatically deducted from earnings/prizes, and we supplemented each donation with a 10% or 25% match.

Table 1. Summary of Experiments

Label	Context	Date	Participants		Population	Location	Donation Fulfillment	Base pay	Don. stake	Prob(Win)	Charity(s)	Match %
			total	usable								
Prolific	Employment choices	2017	320	240	General, non-student population	UK	Automatic (reversible)	£1	£10 Money	50% or 10%	Oxfam, British Heart Fdn	25%
Omnibus	Internet survey, ESSEXLab recruitment	2017	734	460	Students & staff	Essex	Automatic	0	£10 Amazon voucher	50%	Oxfam, British Heart Fdn	25%
Employability	Employability survey	2013–14	592	375	Students	Essex	Automatic	0	£20 Restaurant or Amazon	25%	Oxfam, WWF	10%
Laboratory	Fluid intelligence measure	2013–16	433	131	Students	Mannheim, Dusseldorf	Automatic	€7	€14 Money	50%	Bread for the World, WWF	-
Valentine's	Valentine's cards	2012	205	159	Students & staff	Essex, Bristol, Warwick	Active follow-up	0	£20 (£30) Restaurant	Ambiguous	Right to Play	-

Notes: This table summarizes the most important features of the experiments reported. The Prolific study also enabled us to use a rich set of previously collected background variables. "Usable" refers to participants in the *Before* and *After* treatments described above, excluding losers in the *After* treatment and excluding additional Laboratory treatments.

In addition to the web links to our experimental instruments (below), we provide further details on each experiment, including several key screen shots, and information on other treatment arms in the appendix.

3.1 Pre-registered experiments

3.1.1 Prolific. From July 29 through August 1 2017, we recruited 320 participants through Prolific Academic (prolific.ac), a widely-used recruitment platform for researchers and startups (an additional five participants began but did not complete the survey). We advertised our study as “Employment choices (basic payment plus bonus opportunities)”, promising a base pay of £1 and a duration of about ten minutes (mean actual response time was 5.9 minutes, median 4.3 minutes). We screened for only non-student UK residents, native-English speakers, aged 18 and older, and who had not done any of our previous studies. This left 4212 eligible participants, who were randomly selected to be invited via batched emails from Prolific. The entry page for eligible participants is given in the appendix figure C.0.

The study began with our non-deception rules and a consent form. We next announced "If you complete this survey, you have a 50% [alt: 10%] chance of winning a £10 Amazon voucher. After you complete this survey, we will reveal whether you have won this prize and explain how to claim it"; corresponding to the treatments shown below.

Next, mainly as a “non-deceptive obfuscation” (Zizzo, 2010) we presented a vignette involving a job interview for an “Assurance Trainee” position, and a series of hypothetical questions about requested and anticipated salary. We next elicited self-reported happiness, followed by a series of unincentivized verbal crystallized intelligence questions. The Before and After donation treatments (and prize realizations) followed this. Finally they were asked non-incentivized risk-preference, trust elicitation, and (again) happiness questions.

We next presented the following (randomly allocated) treatments.

Before-50% These 80 participants were first asked to make a conditional commitment, on the screen shown in figure 6. On the next screen those who commited to donate chose whether they wanted their donation to go to Oxfam or the British Heart Foundation. They next learned if they had won (winners’ screen: figure 8). Next, “Before-winners” (after learning they had won) were reminded of their donation (or non-donation) choice, and asked “Would you like to revise your donation decision?” If they chose to revise, they were presented the donation choice once again (with the language from the *After* treatment, as in screen 9).[19]

Before-10% 80 participants were assigned to this treatment, which was identical to *Before-50%*, but with a 10% chance of winning the £10 (which was announced upfront).

After-50% 160 participants were assigned to this treatment, with a 50% chance of winning the £10 prize. After the above, the 80 winners saw the screen in figure 8. Next, (winners), were asked to donate, on the screen depicted in figure 9. On the next screen those who commited to donate chose their preferred charity.

A copy of the full experimental instrument can be viewed and tested at https://goo.gl/xZWDqg.[20] In addition, we outline key implementation details in the annotated pre-registration provided in the additional online materials.

While participants in Prolific are in general aware that they are being paid to participate in *research* and product testing, it seems unlikely that our participants realized that our study concerned charitable giving. Participants spent the large majority of their time on the employment and intelligence questions, and essentially only saw a single screen involving charitable giving. Note that Prolific *already* gives their participants the opportunity to donate their base pay to one of two charities after each experiment. In a separate survey of 190 participants (details available by request) we found that 97% of them knew about Prolific’s donation option

19. As the opportunity to revise was a surprise, and the geographically-dispersed Prolific participants were unlikely to have communicated (even if they had been in the same treatment), we consider the initial donation decision in the Before-50% and After-50% to to be comparable to the *Before* and *After* treatments in our other experiments. All results presented below are based on the initial (rather than the revised) donation choices; however, results are not sensitive to this exclusion.

20. This link will cycle through each of the treatments. You can type any characters in the box requesting a “Prolific ID”; typing “skip” will skip the vignettes and verbal intelligence questions.

0.4
(a)
b

Employment Choices (Basic payment plus bonus opportunities)

Researcher Profile

This study is hosted by David Reinstein

Org: exeter.ac.uk

Study Description

In this study you will be asked to answer a series of questions and give your personal opinions, beliefs, and impressions.

We will not try to deceive you. Everything stated is accurate to the best of our knowledge, and we will pay amounts exactly as we promise.

This study is part of academic research, run by a researcher at the University of Exeter, with other academic partners. The study will take about 10 minutes to complete and does not have any associated risks beyond what you would normally experience in day-to-day life.

If you complete this study you will earn £1.

There will also be opportunities for additional rewards and bonuses based on chance.

All of the data will be anonymous; we will not ask you to provide your name or any other personally identifying information.

Details

- Available Places : 0/160
- Avg. Completion Time : 5mins
- Avg. Reward Per Hour : £12.00/hr
- Maximum Allowed Time : 30mins
- Reward: £ 1.00

Device Compatibility

0.4
(a)
b

You have a 50% chance of winning a £10 bonus. Before we reveal if you have won this prize...

We are giving you the opportunity to donate from your prize to one of two charities: either Oxfam or the British Heart Foundation.
For every pound you donate, we will add an extra 25p. Please click on the images below for further information about these charities (links will open in a new tab).

IF you win the £10 bonus prize, **WOULD** you be willing to donate to one of the above charities?

This will not affect your chance of winning, as the prize winners have already been chosen through a random draw.

If you win, your donation will be automatically deducted from your prize and passed on to the charity of your choice, plus an additional 25% from our own funds. Donations will be made within 14 days and receipts will be kept at the Exeter Business School academic support office. We will not pass your personal information on to the charity.

You have a 50% chance of winning a £10 bonus. Before we reveal if you have won this prize...

We are giving you the opportunity to donate from your prize to one of two charities: either Oxfam or the British Heart Foundation.
For every pound you donate, we will add an extra 25p. Please click on the images below for further information about these charities (links will open in a new tab).

IF you win the £10 bonus prize, **WOULD** you be willing to donate to one of the above charities?

This will not affect your chance of winning, as the prize winners have already been chosen through a random draw.

If you win, your donation will be automatically deducted from your prize and passed on to the charity of your choice, plus an additional 25% from our own funds. Donations will be made within 14 days and receipts will be kept at the Exeter Business School academic support office. We will not pass your personal information on to the charity.

Please enter the amount you would like to donate (if anything) if you win the prize, in the box below. (Enter a number between 0 and 10).

Figure 6. Prolific, Before 50% treatment: Ask screen

You have a 50% chance of winning a £10 bonus. On the next screen we will reveal whether you have won.

>>

Congratulations you have WON a £10 bonus prize!

You must continue to the end to be certain your response is recorded, and to be certain that you can claim your prize.

Continue

Figure 7. Prolific, After 50% treatment (winner): Screens preceding and then announcing win

Congratulations you have WON a £10 bonus prize!

You must continue to the end to be certain your response is recorded, and to be certain that you can claim your prize.

Continue

Figure 8. Prolific, Before (win) treatment: Screen announcing win

Before we continue...

We are giving you the opportunity to donate from your prize to one of two charities: either Oxfam or the British Heart Foundation.
For every pound you donate, we will add an extra 25p. Please click on the images below for further information about these charities (links will open in a new tab).

Now that you have won the £10 bonus prize WOULD you be willing to donate to one of the above charities?

If you donate, your donation will be automatically deducted from your prize and passed on to the charity of your choice, plus an additional 25% from our own funds. Donations will be made within 14 days and receipts will be kept at the Exeter Business School academic support office. We will not pass your personal information on to the charity.

Please enter the amount you would like to donate (if anything) if you win the prize, in the box below. (Enter a number between 0 and 10).

Figure 9. Prolific, After treatment: Ask screen

(although only 17% could correctly identify either of the charities Prolific works with). This context may further reduce the extent to which our charitable treatments were perceived as experimental. None of our participants donated the base payments from our study, and only 1 of 240 had ever previously donated via Prolific.

This experiment involved two features we did not use in any of the other experiments: the *Before-10%* treatment allowed us to consider the sensitivity of this treatment to the probability of realization; and the opportunity for *Before* winners to adjust their donations allowed us to consider the strength of these commitments and the importance of making them binding. We return to this in section 4.3.

3.1.2 Omnibus . We ran our "Omnibus" field experiment in June 2017.[21] Existing ESSEXLab participants (mainly University of Essex students) were awarded £10 Amazon prizes with 50% probability in exchange for completing a wide-ranging omnibus survey. On June 1, nearly the entire pool of 2736 ESSEXLab participants were emailed a direct personalized link with an invitation to take the Omnibus; roughly half of these were randomly assigned to our study, of which we included the first 600 to respond. Non-responders were reminded one and two weeks later. We gave them a deadline of June 18, and provided further information about the Omnibus.[22] Roughly 89% of those who began the survey completed it; median total response time was 12.8 minutes among completers.

After clicking the link, the first screen noted "[...] you have a 50% chance of winning a £10 Amazon voucher. After you complete this survey, we will reveal whether you have won this prize and explain how to claim it." Chosen donations were automatically deducted from prizes, and participants could choose between two charities, here Oxfam or the British Heart Foundation. The *Before* and *After* treatments were block-randomized (stratified) by gender. The language for the *Before* and *After* treatments was extremely similar to the comparable treatments–Before-50% and After–in the Prolific experiment, shown in the previous subsection.

3.1.3 AEA registration of design, hypotheses, and analysis plans. We registered our Omnibus and Prolific experiments with the AEA registry in advance of conducting these studies. Our registered plans, both initial and revised, can be found at https://www.socialscienceregistry.org/trials/2180; in the online appendix to this paper, we provide an annotated excerpt.[23] We registered a descriptive pre-analysis plan, including power calculations, and including our plans to bifurcate our estimates by gender, indicated religiosity, and stated risk-aversion (for the Omnibus) and also by initial stated-happiness (for the Prolific study). As shown below, some of these interactions proved significant (as well as some not pre-registered), but the heterogeneity appears to be entangled with a nonlinear treatment response.

3.2 Non-preregistered experiments

3.2.1 Employability. Our "Employability" field experiment was run in 2013/14 in the context of a career-awareness promotion funded and announced by the University of Essex Faculty of Social Sciences. Participants had a known (25%) probability winning a prize worth £20 (an Amazon or dinner voucher). As noted, donation commitments were *automatically* deducted from the value of the voucher. We obtained 352 valid responses that involved a donation choice. Further details of this experiment are given in appendix C.6.[24]

21. A copy of the experimental instrument can be tested at https://goo.gl/vmHEK6; this will cycle through each of the treatments.

22. This Omnibus was funded as an ESSEXLab innovation. Survey questions were unincentivised, and included demographics, psychometrics, political attitudes, and economic beliefs and preferences. The text of the invitation email sent to participants is given in our online appendix.

23. Some details were revised and registered after the initial registry but before the experiment began. The history of changes to the registration can be seen at https://www.socialscienceregistry.org/trials/2180/history/19934. Additional common sense small changes were made after the experiment began for feasibility reasons, as noted.

24. A copy of the experimental instrument can be tested at https://goo.gl/qSvhi1; this will cycle through each of the treatments.

3.2.2 Laboratory Experiment (Before and After treatments only).

The laboratory environment permitted us to run a variety of treatments; however most of these comparisons proved to be under-powered. We discuss these in more detail in appendix C.3), and also give more complete details of the lab experiments. We focus here on the *Before* and *After* lab treatments. Subjects in these treatments first performed a real-effort task, and were told they would be rewarded €7 for this independently of their performance. They were next told "with a probability of 50 percent you will be rewarded a bonus of €7 on top of your already acquired income of €7." This was followed by *Before* and *After* donation treatments, including a 25% match rate, which closely resembled those in our other studies. The laboratory permitted strong control: subjects could not communicate with others, and we took strong measures to convincingly guarantee anonymity and demonstrate to the subjects that neither their performance nor their donation choices could affect their chances of winning.[25]

These experiments were run in Düsseldorf and Mannheim on a standard experimental subject pool (recruitment was conducted via ORSEE; Greiner, 2004), using virtually identical protocols and zTree code at each lab (Fischbacher, 2007). We ran nine sessions over five days in January–February 2013 and November 2014, and 24 sessions over 7 days in March–April 2016. A complete set of relevant screen shots and translations are available in our online appendix.

3.2.3 Valentine's .

In 2012 we ran an experiment tied to a St. Valentine's Day E-card web site accessible at three UK universities (Bristol, Essex, and Warwick); full details are in Appendix C.5, with further implementation materials provided in the online appendix. We offered a lottery for restaurant vouchers (worth roughly £20-30) as a participation incentive; participants were told the total number of restaurant vouchers to be given away, but not the exact chances of winning.

In this experiment pledges to donate were *not* binding in either treatment; a student who pledged had to make an effort actively follow with this donation. However, our evidence from the Prolific experiment suggests that these pledges were made sincerely. We discuss this in more detail in Appendix C.5. Another unique element: in the Valentine's experiment we asked those "losers" who did *not* win the prize if they would like to donate; however, 0/47 did so. We return to this in section 4.3.

3.3 Randomization tests and summary statistics

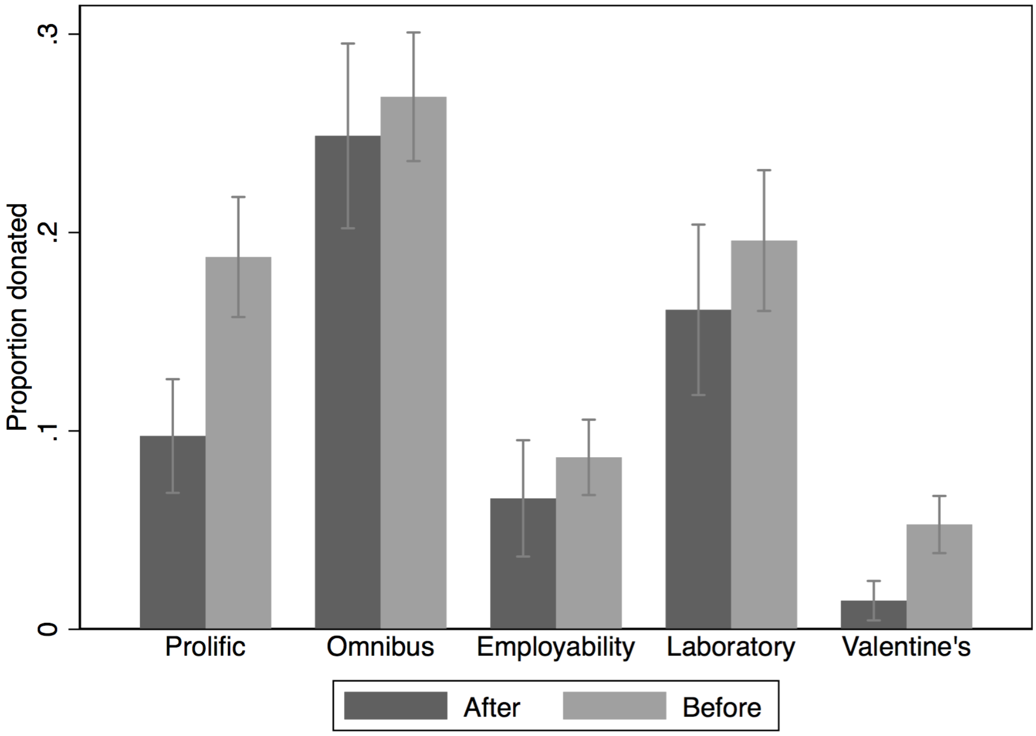
In Appendix D.1 we present summary statistics along with evidence that our randomization successfully balanced the treatments, for all experiments. Tables D.2 and D.3 compare the two treatments for the web-based experiments and the four treatments from the laboratory experiment. The mean values of observable variables are similar across treatments, and we do not detect significant differences for the great majority of tests. In the Prolific experiment, there is some imbalance by age: those in the Before trial are significant older.[26] However, our results are barely affected by including controls for age (cf. Table 6), and robust to other reasonable control strategies, as table D.10 illustrates.

4 Results

We first compare the donations in the Before and After treatments in each experiment. As seen in Figure 10, in each experiment average donations were higher in the Before treatment relative to the After treatment,

25. However, even in this context, we cannot rule out a signaling motive, including (probablistic) signalling to experimenters, to oneself, or to peers in later conversations, given an aversion to lying. We discuss this further in the appendix.

26. The difference is likely due to an unlucky draw and multiple hypothesis testing. Attrition in Prolific was tiny: only 5 of 325 participants who began this survey quit (typically before reaching the prize/donation stage) or timed out.



Bars: estimated standard deviations.

Figure 10. Mean share committed by experiment, by Before vs. After

although the differences are sometimes within the conventional standard-error margin. Table 2 reports the incidence of donating a positive amount, as well as the shares of the endowments donated by treatment, for each experiment, and pooling across experiments.[27]

Table 3 reports ordinary least squares regressions on donation levels (above) and incidence (below) for the Before treatment versus the After treatment in each experiment, excluding participants who earned the lower level of income (or who failed to win the prize).

Each of our experiments suggest an effect in the same direction–a positive effect of "asking before" on donation behavior. The impact is particularly large and highly-significant in the *Prolific* experiment, suggesting that the *Before* treatment might be most effective for non-students. However, in both regression analysis (Table 3) and in simpler tests (Table 2), in about half of our experiments this fails to reach conventional levels of statistical significance. This may stem from a lack of power to detect smaller effects in some of our experiments[28] As shown in Table D.13, the between-subject variance in donation behavior is large, both here and in previous work. For all regression tables we present 95% confidence intervals to convey the precision of our estimates, and to allow inference about the bounds of our effect. In many cases these bounds are wide; e.g., for the Omnibus trial the effect is bounded above at over 1/3 of the mean donation.

We pool our data across all of our experiments to perform a meta-analysis, allowing greater statistical power. For statistical inference, we consider this as a draw from a population composed of likely participants

27. All donations are reported in Euros. Donations from the UK experiments are evaluated at an exchange rate of 1.1971 EUR/GBP (October 1, 2013 rate). For comparability across experiments, in this subsection we do not report on donations from the lower income in the lab (recall that losers in the most of the field experiments were not asked to donate).

28. Button et al., 2013 argue that low-powered studies reduce the *positive predicted value* of our published evidence base. Our experiments are, in general, strongly powered to detect "medium" effect sizes (Cohen's d=0.5) or larger, and the pooled experiments have strong power to detect even "small effect" sizes (d=0.2). We give these calculations in appendix D.3).

Table 2. Summary Statistics: Shares of endowments donated by treatment; Nonparametric tests

	Pooled	Prolific	Omnibus	Employability	Laboratory	Valentine's
After						
Incidence	0.42	0.40	0.50	0.31	0.64	0.13
Mean	0.15	0.10	0.25	0.07	0.16	0.01
Median	0.00	0.00	0.05	0.00	0.14	0.00
75^{th} pctl	0.20	0.20	0.50	0.05	0.29	0.00
Std. dev.	(0.26)	(0.15)	(0.35)	(0.15)	(0.19)	(0.04)
Before						
Incidence	0.48	0.60	0.56	0.31	0.74	0.35
Mean	0.17	0.19	0.27	0.09	0.20	0.05
Median	0.00	0.10	0.10	0.00	0.14	0.00
75^{th} pctl.	0.25	0.30	0.50	0.05	0.29	0.05
Std. dev.	(0.27)	(0.23)	(0.34)	(0.20)	(0.19)	(0.09)
Total						
Incidence	0.46	0.53	0.54	0.31	0.70	0.28
Mean	0.16	0.16	0.26	0.08	0.18	0.04
Median	0.00	0.10	0.10	0.00	0.14	0.00
75^{th} pctl.	0.20	0.20	0.50	0.05	0.29	0.05
Std. dev.	(0.26)	(0.22)	(0.35)	(0.19)	(0.19)	(0.08)
Non-parametric tests						
Diff. in incidence	-0.06	-0.20	-0.06	-0.00	-0.10	-0.22
p-value (Fisher)	0.05	0.00	0.28	1.00	0.25	0.00
Diff. in means	-0.02	-0.09	-0.02	-0.02	-0.03	-0.04
p-value (rank sum)	0.05	0.00	0.32	0.81	0.14	0.00
Observations	1363	240	460	375	129	159

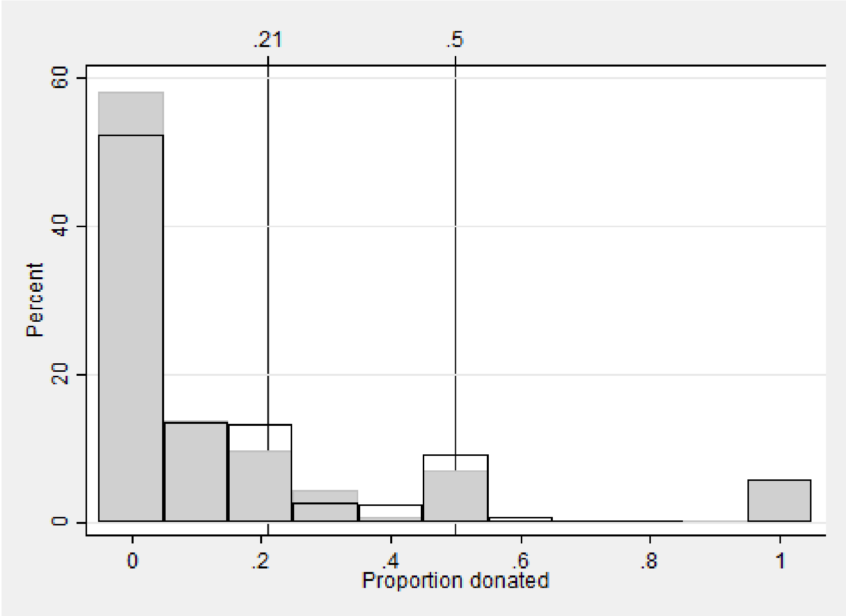
Notes: Average proportion of earnings donated for the Before treatment versus the After treatment in each experiment. For the Lab experiment, this excludes data from the *Certain*, *Before-Both*, and *Uncertain* treatments. We exclude all participants with the lower earnings (lab) and all those in the *After* treatments who did not win the prizes (and thus were not asked to donate). "Incidence": share donating a positive amount. At the bottom we report p-values for tests of differences between outcomes in Before and After treatments, from exact-tests (for incidence) and for rank-sum and t-tests (for proportion donated).

Table 3. Ordinary Least Squares Regressions: Donations levels and incidence by experiment (winners only)

Panel A: Levels	Prolific	Lab	Employability	Omnibus	Valentines
Before	0.90***	0.85*	0.49	0.24	0.92***
[95% CI]	[0.40,1.40]	[-0.16,1.86]	[-0.50,1.49]	[-0.57,1.05]	[0.42,1.42]
(p-value)	(0.0004)	(0.099)	(0.330)	(0.568)	(0.0004)
Constant	0.97***	1.88***	1.58***	2.98***	0.35**
	[0.63,1.32]	[1.29,2.47]	[0.75,2.42]	[2.31,3.64]	[0.061,0.63]
Observations	240	129	375	460	159

Panel B: Incidence	Prolific	Lab	Employability	Omnibus	Valentines
Before	0.20***	0.097	0.0033	0.056	0.21***
[95% CI]	[0.068,0.33]	[-0.099,0.29]	[-0.11,0.12]	[-0.041,0.15]	[0.080,0.34]
(p-value)	(0.003)	(0.328)	(0.956)	(0.257)	(0.002)
Constant	0.40***	0.65***	0.31***	0.50***	0.13***
	[0.29,0.51]	[0.50,0.80]	[0.20,0.41]	[0.42,0.58]	[0.041,0.23]
Observations	240	129	375	460	159

Notes: This table reports coefficients, 95% confidence intervals, standard errors, and t-test p-values from ordinary least squares regressions on donation levels (top) and incidence (bottom) for the Before treatment versus the After treatment in each experiment. Here we also include p-values for future meta- and p-curve analysis (see Simonsohn et al., 2014). These results exclude participants in the After treatment who did *not* win. The Lab column also includes hidden de-meaned dummy and treatment interactions for the laboratory location; no other columns include controls. We report Eicker-Huber-White heteroskedasticity-robust standard errors. Significance levels: * p<0.1, ** p<.05, *** p<.01. Results are very similar for nonparametric tests (table 2), and across a variety of specifications, clusterings, and control strategies; see appendix, table D.10.



Black lines denote the Before treatment. Grey bars denote the After treatment.

Figure 11. Histograms of shares donated, pooled over all experiments

Notes: Average proportion of earnings donated for the Before treatment versus the After treatment, pooling across all experiments. For the Lab experiment, this excludes data from the *Certain*, *Before-Both*, and *Uncertain* treatments. We exclude all participants with the lower earnings (lab) and all those who did not win the prizes. The vertical lines indicate the 75^{th} and 90^{th} percentile of the pooled donation share.

in each of our experiments, with shares corresponding to our relative sample sizes of UK students, German students, and UK nonstudents.[29]

In Figure 11 we present a histogram of the shares of endowment donated, pooled over all experiments. This reveals a small shift away from zero donations towards moderate donations.

All regressions (except where noted) include de-meaned dummies for each experiment, and the interactions of these with the *Before* treatment. This estimator recovers the *average* treatment effect (ATE) for our source population in the presence of heterogeneity. In contrast, OLS estimators are more efficient if effects are homogenous, but they achieve this efficiency by (over)weighting observations (relative to shares of the source population) with higher conditional variance in the treatment and less residual variance in the outcome variable. With heterogeneous treatment effects this yields an arbitrarily weighted estimate of treatment effects (Angrist J. D. and J. S. Pischke, 2008, p. 58), while the "fully interacted" estimator recovers the ATE (see Athey et al., 2017, equation 5.4). However, our results are similar with or without these interactions, as well as with additional interactions by specific pre-determined variables (table D.7). In order to take into account potential session-specific correlated errors (for the lab experiments) and date-specific correlated errors (for the field experiments) we use cluster-robust standard errors at these levels.[30]

As shown in Table 4, the difference for the Before and After treatments is strongly statistically significant in the pooled data and the pooled 95% confidence interval is between 1% and 6% percent of the endowment, implying a proportional increase of 6%-38% of the average donation rate.

29. In appendix D.5, we consider alternate meta-analytic approaches to our data, and address issues of "p-hacking" and potential biases from "optional stopping".
30. We provide robustness checks with specifications of all models in the appendix, table D.10.

Table 4. Ordinary least squares regressions on donations shares; Pooled over experiments

	Pooled			Pooled, preregistered only		
	(1)	(2)	(3)	(4)	(5)	(6)
	Proportion	Level	Incidence	Proportion	Level	Incidence
	b/ci95/se/p	b/ci95/se/p	b/ci95/se/p	b/ci95/se/p	b/ci95/se/p	b/ci95/se/p
Before	0.035***	0.526***	0.097***	0.071***	0.723***	0.162***
	[0.011,0.060]	[0.199,0.854]	[0.049,0.144]	[0.031,0.112]	[0.300,1.146]	[0.062,0.263]
	(0.012)	(0.167)	(0.024)	(0.021)	(0.216)	(0.051)
	(0.004)	(0.002)	(0.000)	(0.001)	(0.001)	(0.002)
Constant	0.138***	1.898***	0.433***	0.138***	1.513***	0.425***
	[0.117,0.158]	[1.610,2.186]	[0.389,0.478]	[0.109,0.167]	[1.207,1.820]	[0.343,0.508]
	(0.011)	(0.147)	(0.023)	(0.015)	(0.156)	(0.042)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Experiment controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1363	1363	1363	700	700	700
Clusters	742	742	742	700	700	700

Notes: This table reports coefficients, 95% confidence intervals, standard errors, and t-test p-values from ordinary least squares regressions. Dependent variables are donations as shares of endowment, donation levels, and donation incidence for the Before treatment versus the After treatment. Results exclude participants in the After treatment who did *not* win. We run pooled analysis comparing all experiments (Columns 1 to 3) and for the pre-registered experiments only (Omnibus and Prolific). All regressions include experiment-specific demeaned dummy controls, which subsume a control for differing endowments, as well as the interactions of these with the Before treatment. We account for potential correlated errors at the session level for the lab experiments using cluster-robust standard errors, while assuming independence across observations for the internet based experiments. Significance levels: * p<0.1, ** p<.05, *** p<.01. Results are very similar for nonparametric tests (table 2), and across a variety of specifications, clusterings, and control strategies; see appendix, table D.10. Results excluding the Valentine's experiment only were similar (available by request)

Result 1. *Overall, the Before treatment increased the average amount donated relative to the After treatment.*

Result 2. *Overall, participants were more likely to give in the Before relative to the After treatment.*

Columns 3 and 6 of Table 4 report the coefficients from a linear probability model for the incidence of giving (logistic specifications lead to similar significance levels and estimated marginal effects; see table D.10 in the appendix. Pooling across all experiments, the *Before* treatment has a significant impact on the extensive margin response, whether or not we limit the sample to experiments with automatic deduction only, i.e. excluding the Valentine's experiment.

Revisiting our theoretical predictions, the greater giving in the *Before* versus *After* treatments is consistent with both loss-aversion (under an expected-wealth or intermediate reference point, which immediately adjusts), confirming prediction 3 and with the signaling model (prediction 2.2). It is inconsistent with expected utility over outcomes (prediction 1), with the standard affective mood argument (Prediction 8, or with loss-aversion with a slowly-changing (or unchanging) reference point (prediction 7).

4.1 Robustness checks and quantile effects, evidentiary value, power

To provide strong evidence that our results are robust and not driven by "p-hacking" (Simonsohn et al., 2014), in appendix Table D.10 we report a matrix of results of reasonable alternative modeling choices over experiment selection, outcomes, error structure, specification, and control variables.

Across these specifications, our results are significant at the p=.05 level or better in 155 of 169 regressions. The notable exceptions are several Probit and Logit specifications on the pre-registered sample, with controls and demeaned interactions, and with robust but not clustered standard errors; 7/9 of the regressions in this category reported p-values above $p = .05$. However, with clustered standard errors, all of the comparable regression coefficients of interest were strongly significant.

Table 5. Quantile regressions on Donation shares: For all experiments pooled; and for preregistered only

Panel A: Pooled	Quantile				
	50	60	70	80	90
	b/ci95/se/p	b/ci95/se/p	b/ci95/se/p	b/ci95/se/p	b/ci95/se/p
Before	0.032**	0.038*	0.084**	0.040	0.047*
	[0.004,0.060]	[-0.003,0.079]	[0.018,0.150]	[-0.019,0.100]	[-0.005,0.098]
	(0.046)	(0.080)	(0.283)	(0.121)	(0.123)
	(0.026)	(0.069)	(0.013)	(0.187)	(0.078)
Observations	1363	1363	1363	1363	1363
Panel B: Prereg.	Quantile				
	50	60	70	80	90
	b/ci95/se/p	b/ci95/se/p	b/ci95/se/p	b/ci95/se/p	b/ci95/se/p
Before	0.073	0.146**	0.127*	0.146	0.146**
	[-0.015,0.161]	[0.023,0.269]	[-0.011,0.265]	[-0.066,0.358]	[0.017,0.275]
	(0.329)	(0.916)	(0.894)	(1.577)	(0.960)
	(0.104)	(0.020)	(0.072)	(0.176)	(0.026)
Observations	700	700	700	700	700

Notes: This table reports coefficients, 95% confidence intervals, standard errors, and t-test p-values from quantile regressions of donations as shares of the endowment; quantiles indicated at top. Panel A reports the results for pooled data over all experiments. Panel B for the preregistered experiments. We exclude donations from the lower income level. All regressions include experiment-specific de-meaned dummy controls, as well as the interactions of these with the Before treatment. The constant is not reported. Significance levels: * p<0.1, ** p<.05, *** p<.01.

In Table 5 we present quantile regression results, allowing us to infer the effects of the *Before* treatment on the donation outcome *distribution* (but, as widely noted, this does not necessarily identify the "distribution of the treatment effect"). For the overall pooled data, we find an increase for each of the 5th to 9th deciles, statistically significant at p<0.10 or better for all but the 8th), and the strongest effect on the 7th decile. (The 7th decile donation share is 14.2% of the endowment in the *After* treatment and 20% in the *Before* treatment; see figure 11 histogram.)[31]

In the appendix sectionD.5 we also discuss the "p-curve approach"; although our results easily "pass" the standard p-curve analysis, we explain why this is not quite the right approach. In this same appendix we discuss issues of statistical inference in the presence of data augmentation/sequential analysis, and explain why our results constitute strong evidence in light of our data collection chronology (the timeline discussed in section C.1).

4.2 Heterogeneity and nonlinearity of effects

Pooled data: Gender, age, religiosity. In Table D.7 in the appendix we report regressions with the Before treatment interacted with key pre-determined observables.[32] As we de-mean each of the binary interacted terms, the base coefficient remains a consistent estimator of the average treatment effect, while the interaction terms represent treatment effect differences between groups (see Athey et al., 2017).

Much previous work has found gender differences in the levels and determinants of other-regarding behavior, in their sensitivity to “price” and cost/benefit ratio (Andreoni, Vesterlund, 2001; Cox et al., 2006), in their response to the time delay (Breman, 2011), and in their sensitivity to reporting, prestige, competition, and previously-reported donations (Jingping, 2013; Jones et al., 2014; Meier, 2007; Pan et al., 2011). In our

31. For the pooled preregistered data, we find an increase for each of the 6th to 9th deciles, strongly statistically significant for the 6th and 9th deciles.

32. Where missing, these variables are linearly imputed using other variables in this regression. The results (available by request) are not sensitive to this imputation. The results are also similar when we introduce each interaction in isolation.

After treatments, women donate more than men, and the effect of the Before treatment is somewhat smaller for women.

Overall, Table D.7 suggests substantial heterogeneity in treatment effects by age and gender, and these interactions are sometimes significant.[33] However, most categories with a higher baseline dummy (representing a higher mean in the After treatment) have a negative interaction term, and vice-versa. When baseline outcome levels vary between groups, it is difficult to distinguish heterogeneity from nonlinear treatment effects. Our evidence is also consistent with a smaller impact of the *Before* treatment for more generous individuals. This may be explained by a steeply diminishing marginal utility of donations in this context. In the appendix Table D.8 we provide results from a power model offering evidence of this nonlinear relationship. Table D.9 reveals that the demographic interaction coefficients lose their significance in this nonlinear specification.[34] This is also consistent with the histogram (Figure 11), which suggests shifts from the smallest donations to medium donations, but no shift towards the largest donations.

4.3 Further treatment comparisons and experiment-specific results

Happiness/affect, and donations. In the Prolific and Omnibus experiments, after (but not immediately after) the prize determination and donation questions we elicited a standard measure of happiness; we report relevant results in table D.5. Unsurprisingly, those who failed to win stated a significantly lower level of happiness (about 3/4 of a standard deviation). However, it does not appear that this increased happiness is driving those (winners) in the After treatment to donate more. In the Prolific experiment we also asked the same happiness question near the beginning of the survey. We find a tightly bounded near-zero relationship between this earlier happiness measure and the chosen donation; the 95% confidence interval is less than 1/4 of a standard deviation.

Other lab treatments. Our laboratory design allowed a richer set of treatment comparisons. However, other than the result discussed above (Before versus After) we find insignificant differences in giving from the higher level of earnings between individual treatments, and the large variance in responses and the wide and overlapping confidence intervals suggests a lack of statistical power to distinguish among these. We report on these results in the appendix (Figure D.10 and Table D.11). Note that the *Before* treatment also yields greater donations than the *Benchmark* treatment (with no income or donation uncertainty), and our results are similar if we pool the *Before-both* and *Before* treatments; further details available by request.

Probabilities of winning (Prolific). The Prolific experiment also included a *Before* treatment with a 10% chance of winning. We find (Table 6) very strong effects of both *Before* treatments relative to the *After* treatment (£0.96 in the 50% treatment and £0.60 in the 10% treatment, significant at the 1-percent and 5 percent levels, respectively). The difference between these two *Before* treatments is not statistically significant, and the confidence intervals reveal limited power to distinguish these.[35]

33. Note that the first three columns of these models include treatment-experiment interactions to avoid confounding heterogeneity with differences in the range of demographics by experiment. These interactions are hidden to save space; several of these experiment-treatment interactions are significant even after the demographic controls; details available by request. Note that the *religiosity* interaction is not significant when we include experiment interactions and dummies. This weakly suggests that "magical thinking" is not driving our result; however, the confidence intervals for this interaction are large, implying limited power to detect a difference.

34. We also ran a two-step procedure (i) regressing giving on pre-determined observables and generating a prediction using After-treatment data, and (ii) regressing giving in the Before treatment on this predicted value and demographics. Again the results (available by request) are consistent with a diminishing-returns treatment effect, and show little evidence of direct heterogeneity by demographics.

35. Recall that the signaling model predicts that if a 50% realization of a donation choice yields a larger (good-type) commitment than does an ex-post (100%) choice, a 10% probability of realization must lead to a donation that is still larger. Thus, to the extent that

We control for two background variables—these come from Prolific "screener" questions, which participants are asked to answer when they first sign up for Prolific, and throughout the months and years they are registered. Thus, these were likely answered well in advance of our study, minimizing any possible contamination. We see a near-zero relationship to de-meaned age. However, we find a strong correlation between giving in our experiment and the participant's response to the question "how much, if anything, did you donate to charity in the last 12 months?"; this supports the relevance and generalizability of our results to external environments.

Table 6. Prolific: Linear models of donation levels and incidence

	Levels	Incidence
	(1)	(2)
	b/ci95/se/p	b/ci95/se/p
Before	0.958***	0.168**
	[0.298,1.617]	[0.014,0.322]
	(0.335)	(0.078)
	(0.005)	(0.032)
Before 10%	-0.359	0.043
	[-1.092,0.374]	[-0.108,0.194]
	(0.372)	(0.076)
	(0.336)	(0.574)
Age (centered)	0.039	0.001
	[-0.020,0.097]	[-0.010,0.012]
	(0.030)	(0.006)
	(0.192)	(0.872)
Not previous donor (de-meaned)	-1.037***	-0.337***
	[-1.570,-0.504]	[-0.481,-0.194]
	(0.270)	(0.073)
	(0.000)	(0.000)
Constant	1.010***	0.393***
	[0.660,1.360]	[0.284,0.502]
	(0.178)	(0.055)
	(0.000)	(0.000)
Before-10% summed	0.599	0.211
s.e.	(0.295)	(0.078)
Observations	230	230

Notes: This table reports coefficients, 95% confidence intervals, standard errors, and t-test p-values for regressions on donations by Treatment for the Prolific experiment. Results exclude participants in the After treatment who did *not* win. The "Before" coefficient is the impact of the Before-50% treatment, and "Before 10%" the adjustment to this impact for the Before-10% treatment. Regression controls for Prolific background variables: age (de-meaned and imputed if missing) and self-reported non-giver; results are similar without controls. Dependent variables are (1) the levels of donations in Euros (2) donation incidence. We report robust standard errors. Stars indicate significance levels * p<0.1, ** p<.05, *** p<.01.

Dynamic consistency and fulfillment of pledges. As noted earlier, in the Valentine's experiment, only 12 of 20 winners who pledged a donation followed this up by fulfilling it. However, in the Prolific experiment, which allowed Before-winners to revise their donation choice, of the 30 who had pledged a positive amount, none chose to reduce this after winning, and four chose to increase it. Furthermore, 2/19 of the Before-winners who had *not* pledged chose to revise their decision to a positive amount. Considering these as a random draw from a larger population, we can infer that there is less than a 5% probability that data as extreme as ours

we can rule out a substantially *larger* average commitmentin the Before-10 treatment relative to Before-50 (noting that even the upper 95% bound is less than 40% greater), this speaks against the signaling model.

would arise if 9.5% or more of the source population would cancel or reduce their donation (exact binomial test, $n = 30$, $K = 0$).

Our evidence from the Prolific experiment suggests that a *legally-binding* pledge may not be necessary for *give-if-you-win* to be a successful fundraising strategy. This finding is consistent with Breman's (2011) field experiment, in which very few donors deviated from their previously committed (increases in) contributions, even though deviating only required a small effort. While this contrasts with Andreoni and Serra-Garcia (2016), contextual differences suggest that our *Before* treatment, (relative to these authors' Pledge treatment) was less likely to induce donations from those with dynamically-inconsistent preferences, and those who donated faced greater "moral pressure" not to renege and be inconsistent.[36]

4.4 Overall giving, implications for fundraising and policy

Fundraising organizations will primarily care about the *overall* effect on giving. In some contexts, it is relevant to consider asking people to donate *both* in good and bad states. We might consider asking people for conditional donations *before* the uncertain income is resolved, or asking them after this, whether or not they are "winners." We have limited evidence on donations from the "losing" state, from laboratory and Valentine's experiments.

An alternate policy would be to ask people to make a commitment "if they win", and then ask them again after (and if) they lose the lottery. The standard outcome-based expected utility model predicts that the former ask (for one state) will have no impact on donations after another state has occured. This implies that donations in an "After-lose" state would be the same whether or not we already gave them the *Before* ask. If this holds, our previously stated results would imply that this alternate policy would also raise more than asking *After* both wins and losses. We have no evidence in this context on whether later asks are affected by earlier asks for unrealized states; however, a moral-licensing effect is suggested by the declining giving across randomly-realized stages in Reinstein (2010) and Tonin et al. (2014).

We could also consider asking in advance for a complete contingent plan, or asking *After* for both winners and losers. Here we have some evidence, but limited power to detect differences. As reported in appendix C.3, earlier runs of the laboratory experiment included a *Before-both* treatment, where we asked subjects, in advance, to make a donation choice for both winning and losing states). As noted, we also asked losers to donate in the Laboratory and Valentine's experiments. The regressions in table 7 consider the expected value of the donation for the analogue each of the policies mentioned above. Here, the "expected donation" outcome is coded as half the commitment in the *Before* treatment (the regression base group), half of the sum of the commitments for each state in the *Before Both* treatment, and the actual donation in the After-Lose and After-Win treatments/outcomes. Column 1 reports on the earlier laboratory sessions, which included the *Before Both* treatment. Here, *Before Both* raised 0.206 more than *Before* in expectation but this is statistically insignificant. *After* raised .993 more than *Before* after a win, but 0.340 *less* after a loss. In net, After-both raised 0.121 more than *Before-Both* and 0.326 more than *Before*; neither of these differences are significant. Column 2, which considers the latter difference for *all* lab experiments finds similar results. Column 3 reports on the Valentine's

36. Andreoni, Serra-Garcia, 2016 found that nearly half of students who pledged to donate $5 in a "Pledge" treatment chose to renege on this a week later. However, our experiments differed in important ways. We presented an opportunity to adjust the donation i. in either direction, ii. after experiencing a win, iii. very soon after the initial commitment, iv. in the same environment as the initial commitment and v. as a surprise—the initial commitment was not a tentative one. In contrast, Andreoni, Serra-Garcia, 2016's initial "Pledge" was worded softly ("Ask me again next week and I will make my final decision"), reneging (not increasing) was allowed a full week later, and the week 1 (but not week 2) sessions were accompanied by a slide show read by the experimenter in support of the charity.

experiment, in which none of the losers donated. The After treatment raised £0.459 less than the Before treatment in expectation, and this is strongly significant.

Table 7. Linear models: Comparing expected amounts raised

	(1) Lab, earlier sessions b/ci95/se/p	(2) Lab, all b/ci95/se/p	(3) Valentine's b/ci95/se/p
After (Win or lose, Valentine's)			-0.459*** [-0.711,-0.208] (0.127) (0.000)
Before Both	0.206 [-0.283,0.694] (0.235) (0.392)		
After, Win	0.993** [0.058,1.927] (0.449) (0.038)	0.900** [0.102,1.697] (0.385) (0.029)	
After, Lose	-0.340 [-0.902,0.222] (0.270) (0.222)	-0.397 [-0.900,0.105] (0.243) (0.116)	
EV-gain: After-both vs Before-both	0.121		
se	(0.256)		
EV-gain: After-both vs Before-win	0.326	0.251	
se	(0.250)	(0.229)	
Observations	248	189	159

Notes: This table reports coefficients, 95% confidence intervals, standard errors, and t-test p-values from ordinary least squares regressions on the "Expected revenue raised" (described in text) for each treatment/outcome . *Before* is the base category. The first column reports on the laboratory experiment for the earlier sessions, where we continued to use the Before Both treatment. The Lab columns also include hidden de-meaned dummy and treatment interactions for the laboratory location. We report Eicker-Huber-White heteroskedasticity-robust standard errors. Significance levels: * p<0.1, ** p<.05, *** p<.01.

However, there are many real-world fundraising environments in which an appeal to "losers" (before or ex-post) is likely to be impractical or unsuccessful, and where only winners are targeted. A prime example is the the Founders Pledge, which asks entreprenuers and investors to make a legally-binding "commit[ment] to donating a chosen % of your personal proceeds upon exit to charity."[37] Asking a similar commitment at or anticipating the point of failure could be awkward and uninspiring. It also may be infeasible: what quantity would one target a percentage of, and at what point would failure be publicly visible?

In the conclusion we discuss several underexplored and potentially fruitful applications involving lotteries, tournaments, and windfall gains; in most of these, there is natural way to ask for donations from "nonwinners".

5 Conclusion

Our experiments are the first to document the effect of the resolution of income uncertainty on other-regarding behavior, augmenting existing evidence that such behavior may not be well-explained by outcome-based expected utility theory. As noted, this also has an important implication for experimental methods: many experiments used a "random problem selection mechanism" (Azrieli et al., 2012), selecting only a single decision

37. (From founderspledge.com). As of March 2018, they have recorded 1200 pledges worth $419 million, and 31 exit donations worth $13 million. The site notes "and don't worry - if you don't exit, you don't give."

stage for payment, arguing that this ensures no feedback between stages. This may be violated: e.g., in a dictator game, if an earlier stage's incentives prompted a generous commitment, this might satiate the desire for signaling and lead to lower commitments in later stages (as seen in Tonin et al., 2014 and Reinstein, 2010). The implications for the *strategy method* (Selten, 1967) are similar; in making such decisions, subjects may trade off the costs and benefits of signaling between contingencies.

We find higher donations and a greater propensity to commit to donate when individuals are asked to conditionally commit before learning if they have won a prize or bonus, relative to those asked after they have won. This result is statistically significant ($p<.01$) in pooled data and in three of five distinct contexts ($p<.01$ for two contexts and $p<.10$ for one context) across several different populations (UK nonstudents, British students, German students). The magnitude of this effect is within the range of effects estimated in other charitable giving experiments (see appendix D.8). The effect is stronger for those groups predicted to donate less in the After treatment. However, it may not carry over into every situation; in other environments asking after a bonus may be more effective; perhaps affect/mood may dominate. Allowing for heterogeneous motivations, the theory presented is ambiguous, suggesting that results may vary according to the environment. Still, our evidence strongly suggests that in relevant environments contributions involving uncertain realization and/or uncertain income do not follow the predictions of standard expected utility models.

5.1 Further potential applications

Our results may be relevant to a variety of environments, and the "give if you win" model appears to be catching on to some extent. In addition to the aforementioned Founders Pledge and EA giving pledges, other areas are being explored. In each of these cases, fundraisers will most reasonably ask for commitments or donations from the "winning" state or states, asking for donations/commitments from "losers" is likely to be demotivating or impractical.

Some sectors, most famously financial services, offer substantial bonuses, the exact magnitude of which are often unclear ahead of time. Our findings suggest that asking workers to commit to give a share of their bonus (or their “bonus in excess of a specified expectation”) could be an effective revenue generator for charities. For example, City Philanthropy ran a think tank exploring a "Bonus Pledge"[38] Note that they most naturally focused in on pledging a share of any *bonus* received; they did not discuss or consider asking for a donation "in the event that you do not receive a bonus."

The Dartmouth Founders Project[39] brings this to an alumni fundraising context. This could be extended to solicit pledges to volunteer to mentor future students, as well as pledges to give to charities more generally. Universities and student philanthropic groups may want to target particular groups of students facing a large but variable immediate payoff, or a great deal of lifetime income uncertainty which will be largely determined by theirs initial job placement. E.g., MBA students may be facing a large but variable immediate payoff (MBA "signing bonuses"). Business and law students face a great deal of lifetime income uncertainty, much of which will be determined by the initial job placement (see Oyer 2008, and the perceived "all or nothing" nature of landing a job at a "big-4" financial services firm or a London magic circle" law firm.)

Lotteries could be targeted directly: e.g., those purchasing Powerball tickets could be asked to simply tick a box to pledge to donate a share of their winnings; winners who pledged could be reminded of their commitment, or have this commitment automatically deducted. Similar pledges could be solicited from educational

38. http://www.cityphilanthropy.org.uk/events/1-bonus-pledge-think-tank, accessed 9 Mar 2018.
39. http://giving.dartmouth.edu/founders/?q=about-founders-project

and "opportunity" lotteries; e.g., admission to medical school in the Netherlands involves an explicit random lottery component and winning yields a strong boost to lifetime income Ketel et al., 2016; applicants could take a pledge to either donate if they win, or to do volunteer service. Applicants to the US immigration "diversity lottery" from low income countries could be asked to tick a box to pledge to give back or volunteer in their home country or community if they win. we could also solicit a general pledge from certain categories of unexpected windfall gains (e.g., inheritances from one's extended family, large "punitive" civil legal judgements, surprisingly high tax rebates, and extreme unforseen business profits due to regulatory changes and international events). No similar pledge from *losing* lottery tickets would be feasible; there would be millions of people to contact at a not particularly opportune time, and no income that would be easily to "automatically deduct" from.

This could also apply to processes involving merit and effort as well as randomness. Such voluntary pledges could extend to applicants for international scholarships (especially from countries suffering from "brain-drain"), bidders for large public contracts, participants in high-stakes poker tournaments[40], as well as the aforementioned alumni/first-job pledges.

If binding, these pledges will also allow pledgers make a moral statement; those who feel the wealthy should be more generous can credibly commit that if they become wealthy, they will be generous.

Many governments are deeply involved in promoting private charitable giving, through tax incentives, promotional activities, and encouraging particular donation channels. E.g., the UK government put out an official "Giving White paper seeking to renew Britain's culture of philanthropy" [41], and has actively encouraged "Payroll Giving", to limited success.[42] As suggested in a recent paper funded by the UK Cabinet office Team, 2013, governments may want to integrate "windfall giving" into these promoted giving schemes, and help provide a clear legal environment for charitable commitments of uncertain income.

5.2 Suggestions for future work

Although we find that the *Before* ask raises more than the *After* ask across several contexts, these all involved small participation rewards, and all but our Prolific study used mainly university students. This basic comparison should be tested and trialled in larger-stakes settings, involving relevant participants, environmental and social contexts. In particular, we would encourage trials involving employee giving at companies that offer bonuses and incentive pay (we describe and promote this at giveifyouwin.org).

Our laboratory design and theoretical discussion in appendices A and C.3 lay out a program for unpacking the drivers of differences between *Before* and *After* giving. In particular, the *Uncertain* treatment can isolate the impact of uncertain *collection* of pledges without *income* uncertainty, differentiating our loss-aversion and signaling models. This has practical, as well as theoretical relevance. Contributions or volunteers may only be relevant after a particular unlikely outcome; E.g., a bone marrow or organ donor may only be useful after they are determined to be a biological match. Specialist experts or owners of unusual tools (e.g., a marine biologist, or the owner of a helicopter in a remote village) may be urgently required to volunteer only after rare events (such as a beached whale on a secluded coast or a defibrilator in a mountain village). Furthermore, some people may only care to donate in response to particular events, e.g., an earthquake hitting their home town, or a civil war in their country of origin. If the uncertain *collection* of pledges (i.e., need to realize promises) is

40. see, e.g. https://reg-charity.org/introducing-the-reg-tournament-pledge/
41. See https://www.gov.uk/government/news/giving-white-paper
42. (https://www.gov.uk/government/publications/charities-detailed-guidance-notes/chapter-4-payroll-giving), accessed 9 Mar 2018

what matters, it will be more effective to ask these people to make conditional pledges in advance, even if they are not pledging from uncertain income. E.g., organizations like the Disaster Emergency Committee might find it productive to solicit conditional donations, e.g., "I pledge $1000 in the event of a tsunami or other natural disaster killing more than 50,000 people in the Philippines in the next ten years".

References

Andreoni, J., A. Payne (2003), Do government grants to private charities crowd out giving or fund-raising? American Economic Review 792–812.

Andreoni, J., J. Rao (2011), The power of asking: How communication affects selfishness, empathy, and altruism, Journal of Public Economics 95 (7) 513–520.

Andreoni, J. (1990), Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving, The Economic Journal 100 464–477.

Andreoni, J., 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., M. Serra-Garcia (2016), Time-Inconsistent Charitable Giving, URL: https://papers.ssrn.com/sol3/papers.cfm?abstract%7B%5C_%7Did=2868923.

Andreoni, J., L. Vesterlund (2001), Which is the fair sex? Gender differences in altruism, Quarterly Journal of Economics 116 (1) 293–312.

Angrist J. D. and J. S. Pischke (2008), Mostly Harmless Econometrics : An Empiricist ' s Companion, March 290.

Athey, S., G. W. Imbens (January 1, 2017), Chapter 3 - The Econometrics of Randomized Experimentsa, in: *Handbook of Economic Field Experiments*, edited by A. V. Banerjee, E. Duflo, volume 1, Supplement C volumes, Handbook of Field Experiments, DOI: 10.1016/bs.hefe.2016.10.003, North-Holland 73–140, URL: http://www.sciencedirect.com/science/article/pii/S2214658X16300174 (visited on 09/25/2017).

Atkinson, A. B. (2009), Giving overseas and public policy, Journal of Public Economics 93 (5-6) 647–653.

Azrieli, Y., C. P. Chambers, P. J. Healy (2012), Incentives in experiments: A theoretical analysis, technical report, Working Paper, Ohio State University.

Becker, G. (1974), A Theory of Marriage: Part II.

Benabou, R., J. Tirole (2006), Incentives and Prosocial Behavior, American Economic Review 96 (5) 1652–1678.

Benabou, R., J. Tirole (2011), Identity, morals, and taboos: Beliefs as assets, The Quarterly Journal of Economics 126 (2) 805–855.

Breman, A. (2011), Give more tomorrow: Two field experiments on altruism and intertemporal choice, Journal of Public Economics 95 (11) 1349–1357.

Brock, J. M., A. Lange, E. Y. Ozbay (2013), Dictating the risk: Experimental evidence on giving in risky environments, The American Economic Review 103 (1) 415–437.

Burns, P. (1985), Experience and Decision Making, Research in Experimental Economics,(ed. V. Smith) 3.

Burrows, P., G. Loomes (1994), The impact of fairness on bargainin, Empirical Economics 19 (2) 201–221.

Button, K. S., J. P. A. Ioannidis, C. Mokrysz, B. A. Nosek, J. Flint, E. S. J. Robinson, M. R. Munafò (2013), Power failure: why small sample size undermines the reliability of neuroscience, Nature Reviews Neuroscience 14 (5) 365–376.

Camerer, C., L. Babcock, G. Loewenstein, R. Thaler (1997), Labor supply of New York City cabdrivers: One day at a time, The Quarterly Journal of Economics 112 (2) 407–441.

Cherry, T., P. Frykblom, J. Shogren (2002), Hardnose the Dictator, American Economic Review 92 (4) 1218–1221.

Cialdini, R. B., B. L. Darby, J. E. Vincent (1973), Transgression and altruism: A case for hedonism, Journal of Experimental Social Psychology 9 (6) 502–516.

Comptroller, N. S. (March 2014), Wall Street Bonuses Went Up In 2013, Press release. online, URL: http://www.osc.state.ny.us/press/releases/mar14/031214.htm.

Cox, J. C., C. A. Deck (2006), When are women more generous than men? Economic Inquiry 44 (4) 587–598.

Cubitt, R. P., C. Starmer, R. Sugden (1998), On the validity of the random lottery incentive system, Experimental Economics 1 (2) 115–131.

DellaVigna, S., J. List, U. Malmendier (2012), Testing for Altruism and Social Pressure in Charitable Giving, The Quarterly Journal of Economics 127 (1) 1–56.

Drouvelis, M., B. Grosskopf (2016), The effects of induced emotions on pro-social behaviour, Journal of Public Economics 134 1–8.

Duncan, B. (2004), A Theory of Impact Philanthropy, Journal of Public Economics 88 (9-10) 2159–2180.

Eckel, C. C., P. J. Grossman (October 1, 1996), Altruism in Anonymous Dictator Games, Games and Economic Behavior 16 (2) 181–191, URL: http://www.sciencedirect.com/science/article/pii/S0899825696900810 (visited on 08/21/2017).

Exley, C. L. (2016), Excusing selfishness in charitable giving: The role of risk, Review of Economic Studies 83 (2) 587–628.

Fischbacher, U. (2007), z-Tree: Zurich toolbox for ready-made economic experiments, Experimental Economics 10 (2) 171–178.

Fishbach, A., A. A. Labroo (2007), Be better or be merry: How mood affects self-control, Journal of Peronality and Social Psychology 93 (2) 158.

Gino, F., M. I. Norton, R. A. Weber (2016), Motivated Bayesians: Feeling moral while acting egoistically, The Journal of Economic Perspectives 30 (3) 189–212.

Gourieroux C., A. M., A. Trognon (1984), Pseudo Maximum Likelihood Methods: Applications to Poisson Models, Econometrica 52 701–720.

Greene, W. H. (1994), Accounting for excess zeros and sample selection in Poisson and negative binomial regression models.

Greiner, B. (2004), An online recruitment system for economic experiments, GWDG Bericht 63, edited by K. K. Macho 79–93.

Grossman, Z. (2015), Self-signaling versus social-signaling in giving, Journal of Economic Behavior & Organization.

Harbaugh, W. T. (May 1998), The Prestige Motive for Making Charitable Transfers, The American Economic Review 88 (2) 277–282, URL: http://links.jstor.org/sici?sici=0002-8282(199805)88:2%3C277:TPMFMC%3E2.0.CO;2-9.

Harrison, G., J. List (2004), Field Experiments, Journal of Economic Literature 42 1009–1055.

Helson, H. (1964), Adaptation-level theory.

Hoffman, E., K. McCabe, V. Smith (1996), Social Distance and Other-Regarding Behavior in Dictator Games, American Economic Review 86 653–660.

Hoffman, E., M. Spitzer (1985), Entitlements, Rights, and Fairness: An Experimental Examination of Subjects' Concepts of Distributive Justice, Journal of Legal Studies 14 259.

Huck, S., I. Rasul (2011), Matched fundraising: Evidence from a natural field experiment, Journal of Public Economics 95 (5) 351–362.

Jingping, L. (2013), Four Essays on the Economics of Pro-Social Behaviors.

John, L. K., G. Loewenstein, D. Prelec (2012), Measuring the prevalence of questionable research practices with incentives for truth telling, Psychological science 23 (5) 524–532.

Jones, D., S. Linardi (2014), Wallflowers: Experimental Evidence of an Aversion to Standing Out, Management Science 60 (7) 1757–1771.

Karlan, D., J. A. List (December 2007), Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment, American Economic Review 97 (5) 1774–1793, URL: https://www.aeaweb.org/articles?id=10.1257/aer.97.5.1774 (visited on 08/21/2017).

Ketel, N., E. Leuven, H. Oosterbeek, B. van der Klaauw (2016), The returns to medical school: Evidence from admission lotteries, American Economic Journal: Applied Economics 8 (2) 225–54.

Kidd, M., A. Nicholas, B. Rai (2013), Tournament outcomes and prosocial behaviour, Journal of Economic Psychology 39 387–401.

Knowles, S., M. Servátka (2015), Transaction costs, the opportunity cost of time and procrastination in charitable giving, Journal of Public Economics 125 54–63.

Koszegi, B., M. Rabin (2006), A MODEL OF REFERENCE-DEPENDENT PREFERENCES*, The Quarterly Journal of Economics 121 (4) 1133–1165, URL: http://www.personal.ceu.hu/staff/Botond%7B%5C_%7DKoszegi/refdep.pdf.

Kotzebue, A. v., B. U. Wigger (2009), Charitable Giving and Fundraising: When Beneficiaries Bother Benefactors, in: *XVI Encuentro de Economia Publica: 5 y 6 de febrero de 2009: Palacio de Congresos de Granada* 48.

Kuhn, P., P. Kooreman, A. Soetevent, A. Kapteyn (2011), The effects of lottery prizes on winners and their neighbors: Evidence from the Dutch Postcode Lottery, American Economic Review 101 (5) 2226–2247.

Lakens, D. (2014), Special issue article : Methods and statistics in social psychology : Re fi nements and new developments Performing high-powered studies ef fi ciently with sequential analyses, 710 (March) 701–710.

Langer, E. J. (1975), The illusion of control. Journal of personality and social psychology 32 (2) 311.

Levin, P. F., A. M. Isen (1975), Further studies on the effect of feeling good on helping, Sociometry 141–147.

Levy, G., R. Razin (2006), A Theory of Religion: Linking Individual Beliefs, Rituals, and Social Cohesion, Department of Economics WP, LSE.

Loewenstein, G., D. Schkade (1999), Wouldn't it be nice? Predicting future feelings, Well-being: The foundations of hedonic psychology 85–105.

Maniadis, Z., F. Tufano, J. A. List (2014), One Swallow Doesn't Make a Summer: New Evidence on Anchoring Effects, American Economic Review 104 (1) 277–90, URL: http://www.aeaweb.org/articles.php?doi=10.1257/aer.104.1.277.

Meier, S. (2007), Do women behave less or more prosocially than men? Evidence from two field experiments, Public Finance Review 35 (2) 215–232.

Oberholzer-Gee, F., R. Eichenberger (2004), Fairness in extended dictator game experiments, Manuscript.

Oyer, P. (2008), The making of an investment banker: Stock market shocks, career choice, and lifetime income, The Journal of Finance 63 (6) 2601–2628.

Pan, X. S., D. Houser (2011), Competition for trophies triggers male generosity, PloS one 6 (4) e18050.

Raven, J. (1936), Mental tests used in genetic studies: The performance of related individuals on tests mainly educative and mainly reproductive, Unpublished master's thesis, University of London.

Reinstein, D. (2010), Substitution Among Charitable Contributions: An Experimental Study.

Reinstein, D. A. (2011), Does One Charitable Contribution Come at the Expense of Another? The BE Journal of Economic Analysis and Policy 11 (1).

Reinstein, D., G. Riener (March 1, 2012a), Decomposing desert and tangibility effects in a charitable giving experiment, Experimental Economics 15 (1) 229–240, (visited on 08/17/2017).

Reinstein, D., G. Riener (February 2012b), Reputation and influence in charitable giving: an experiment, Theory and Decision; Dordrecht 72 (2) 221–243, URL: https://search.proquest.com/docview/916001281/abstract/507947BE86B14E51PQ/1 (visited on 08/17/2017).

Roemer, J. E. (2010), Kantian equilibrium, The Scandinavian Journal of Economics 112 (1) 1–24.

Rubin, Z., L. A. Peplau (1975), Who Believes in a Just World? Journal of Social Issues 31 (3) 65–89.

Sagarin, B. J., J. K. Ambler, E. M. Lee (2014), An Ethical Approach to Peeking at Data, Perspectives on Psychological Science 9 (3) 293–304.

Sanborn, A. N., T. T. Hills (2014), The frequentist implications of optional stopping on Bayesian hypothesis tests, Psychonomic bulletin & review 21 (2) 283–300.

Sandroni, A., S. Ludwig, P. Kircher (2013), On the difference between social and private goods, The BE Journal of Theoretical Economics 13 (1) 151–177.

Selten, R. (1967), Die Strategiemethode zur Erforschung des eingeschrankt rationalen Verhaltens im Rahmen eines Oligopolexperiments, in: *Beitrage zur Experimentellen Wirtschaftsforschung*, edited by H. Sauerman, JCB Mohr (Paul Siebeck), Tubingen.

Simmons, J. P., L. D. Nelson, U. Simonsohn (2011), False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant, Psychological science 22 (11) 1359–1366.

Simonsohn, U. (2009), Direct risk aversion evidence from risky prospects valued below their worst outcome, Psychological Science 20 (6) 686–692.

Simonsohn, U., L. D. Nelson, J. P. Simmons (2014), P-curve: a key to the file-drawer. Journal of Experimental Psychology: General 143 (2) 534.

Simpson, B., R. Willer (2008), Altruism and Indirect Reciprocity: The Interaction of Person and Situation in Prosocial Behavior, Social Psychology Quarterly 71 (1) 37–52.

Small, D. A., G. Loewenstein (2003), Helping a victim or helping the victim: Altruism and identifiability, Journal of Risk and uncertainty 26 (1) 5–16.

Smith, A. (2011), Group composition and conditional cooperation, The Journal of Socio-Economics 40 (5) 616–622.

Sugden, R. (1982), On the Economics of Philanthropy, Economic Journal 92 (366) 341–350.

Sugden, R. (December 1984), Reciprocity: The Supply of Public Goods Through Voluntary Contributions, The Economic Journal 94 (376) 772–787, URL: http://links.jstor.org/sici?sici=0013-0133(198412)94:376%3C772:RTSOPG%3E2.0.CO;2-H.

Team, B. I. (2013), Applying behavioural insights to charitable giving, Cabinet Office.

Thaler, R. H., S. Benartzi (2004), Save More Tomorrow: Using behavioral economics to increase employee saving, Journal of political Economy 112 (S1) S164–S187.

Tonin, M., M. Vlassopoulos (2014), An experimental investigation of intrinsic motivations for giving, Theory and Decision 76 (1) 47–67.

Tversky, A., D. Kahneman (1991), Loss aversion in riskless choice: A reference-dependent model, The Quarterly Journal of Economics 106 (4) 1039–1061.

Underwood, B., J. F. Berenson, R. J. Berenson, K. K. Cheng, D. Wilson, J. Kulik, B. S. Moore, G. Wenzel (1976), Attention, negative affect, and altruism: An ecological validation, Personality and social psychology bulletin 3 (1) 54–58.

Vesterlund, L. D. (2003), The Informational Value of Sequential Fund-raising, Journal of Public Economics 87 1278–1293.

Weyant, J. M. (1978), Effects of mood states, costs, and benefits on helping. Journal of Personality and Social Psychology 36 (10) 1169.

Zizzo, D. J. (2010), Experimenter demand effects in economic experiments, Experimental Economics 13 (1) 75–98.

Appendix

A Appendix: Theoretical Predictions; extensions and proofs

To isolate the drivers of potential differences in donations in the above settings, we consider donations from the lower ("losing") income, which we label as $g_\ell^a \geq 0$, as well as the following additional environments. Individuals in setting (D) face no uncertainty, as they know from the beginning whether they have a high or low income; we label their donations g_w^d and g_ℓ^d, indexed by income. Those in setting U also have a definite income (w or ℓ), but their chosen donation (labelled g_w^u or g_ℓ^u) will only be collected with probability p, and otherwise it is returned to them. Finally, we considered a variation of the *Before* setting, where the individual makes *two* donation choices before the uncertainty is resolved; one for low income (g_ℓ^{bb}) and one for high income (g_w^{bb}).

A.1 Expected Utility over outcomes

Consider an individual maximizing a Bernoulli utility function $v(x,g)$, where x represents consumption and g the charitable contribution, subject to non-negativity constraints and to the budget constraint $x+g \leq z$; Wealth or purchasing power (in a given state state of the world) is here denoted by $z \in \{w,\ell\}$.

Let us assume her utility satisfies the standard expected-utility properties, so that the utility of a prospect is the probability-weighted sum of the utility of each element. Suppose she is asked to make a conditional decision, choosing g_w and g_ℓ before learning the realization of z. Assuming non-satiation, we can substitute in the budget constraints and express her problem as

$$g_w^b,\, g_\ell^b \;:=\; \underset{g_w, g_\ell}{\operatorname{argmax}}(1-p)v(l-g_l, g_l) + pv(w-g_w, g_w),$$

where p is her probability of winning the prize. As explained in the main text, this characterizes the most widely cited models of giving, including a warm glow model where, as we assume throughout, the warm glow derives only from the amount *actually* donated. It is trivial to see that the same choices obtain when the donation decision is made after any uncertainty about income has resolved, and for the *Uncertain Collection* case. Thus, a standard model will predict $g_z^a = g_z^b = g_z^d = g_z^u$ for $z \in \{w,l\}$ or for any level of income. This remains true for g_w^b, if, as in our field experiment and in setting B, we constrain $g_l^b = 0$. In other words, the timing of the decision (i.e., whether it is a sure thing or a prospect), is irrelevant to the individual's choice.

The full prediction:

$$\begin{aligned} g_w^d &= g_w^a = g_w^b = g_w^{bb} = g_w^u \qquad and \\ g_\ell^d &= g_\ell^a = g_\ell^{bb} = g_l^u \end{aligned}$$

A.2 Signaling Model of Reputation with uncertain collection

We define an individual's Bernoulli utility as an additively separable function:

$$v(x,g) \;=\; u(x) + \theta\omega(Dg) + R(\phi), \tag{1}$$

where x is an individual's own consumption, g is the amount committed to donate, and D is an indicator variable taking the value one if the committed donation is collected, and zero otherwise. $\theta\omega(\cdot)$ is his intrinsic utility from donating, and $\theta \in \{0,1\}$ reflects his type, "bad" or "good", respectively, drawn by nature with $pr(\theta = 1) := \mu \in (0,1)$.[43] The function $u(\cdot)$ represents the sub-utility of own-consumption, and $\omega(Dg)$ represents his private benefit from actually giving Dg (akin to a warm-glow function, but equally representing the private benefit from augmenting a public good). $R(\phi)$ is his utility from his reputation, a function of ϕ, which represents the posterior probability he and others put on him being of type $\theta = 1$, where $R(0) = 0$, $R(\mu) = \lambda r$, $R(1) = r$; $r > 0$, $0 \leq \lambda < \leq 1$. Note that ϕ may depend on $\mathbf{g_{-i}}$ and g in equilibrium, where $\mathbf{g_{-i}}$ is the vector of others' committed contributions.[44]

As in Benabou et al. (2006), we consider a direct payoff from reputation (in a social or self-signaling context, "which may be instrumental … or purely hedonic"). We focus now on the setting where the individual faces income uncertainty and is asked about a contingent donation only for the state with high income, w. By standard assumptions, she will maximize the expected value of this Bernoulli utility function subject to the budget constraint

$$x + g \ \leq \ z,$$

where z denotes wealth. As donation commitments are only made for one income level, we omit the income superscript for g. The expected value of the utility can be restated as

$$U^{\theta}(g) \ = \ u(l) + p[u(w-g) - u(l)] + p\theta\omega(g) + R(\phi),$$

where p is the the probability (at the time the donation decision is made) that the income is w and the donation will be collected. We consider equilibria where someone is assumed to be a potential good type only if he donates some amount which we will define as g_1. Note that in a separating equilibrium reputation benefits are 0 for the bad types and r for good types. As we are only allowing positive donations ($g \geq 0$ is an implied constraint), it is trivial to show that in a separating equilibrium bad types donate nothing, i.e., $g_0 = 0$, which we assume henceforth. In a pooling equilibrium, everyone will get reputation benefit $R(\mu) = \lambda r$, i.e., some share of the reputation benefit of being known to be a good type.[45]

Separating equilibrium: constraints. We next state the constraints for a separating equilibrium. The relevant constraint of the good type is that

$$U^1(g_1) \ \geqslant \ U^1(g) \ \ \forall g. \tag{2}$$

The relevant incentive compatibility condition of the bad type requires:

43. Our key insights generalize to a model in which types have continuous support, and the probability distribution may condition on a set of observable variables including gender and previous actions, as long as some uncertainty remains.

44. Note that we are assuming he knows his own type θ at the point he makes his decision. To make this a model where *self*-signaling is important, he must have limited memory of θ but better memory of past actions, as in Benabou et al. (2011). These authors write: "This self-assessment or signal, however, may not be perfectly recalled or 'accessible' later on —in fact, there will be strong incentives to remember it in a self-serving way. Actions, by contrast, are much easier to quantify, record and remember than their underlying motivation, making it rational for an agent to define himself partly through his past choices …"

45. λ may depend on the *actual* share of good types in the population, but this will not affect our results unless we are comparing across distinct populations.

$$U^0(0) \;\geqslant\; U^0(g_1). \tag{3}$$

Let g^* represent a good type's preferred donation *net of reputation,* i.e.[46]

$$g^* \;=\; \underset{g}{\operatorname{argmax}}\, \{u(w-g) + \omega(g)\}.$$

Solutions .

Case 1 . Suppose at g^* the bad type will not deviate even if that brings him reputation benefit r, i.e., suppose

$$-p[u(w-g^*) - u(w)] \;\geq\; r. \tag{4}$$

Then, in the separating equilibrium with the lowest level of contributions (which is also the one that maximizes welfare for the good type, and the only one satisfying the intuitive criterion), $g_1 = g^*$, independent of p. The bad type's incentive constraint does not bind in this case, while the good type chooses her warm-glow maximizing donation level, satisfying condition 2. Note that there cannot be a pooling equilibrium here. Summing up, for the intuitive equilibrium in this parameter space, *conditional* donations do not change in the probability that they are collected; hence the intuitive criterion predicts that the *expected* donation will increase in p. Conversely, the expected contribution will decrease as p decreases up until the point at which Condition 4 no longer holds, i.e., up to the point where the collection probability is low enough to tempt bad types to imitate the good types.

Case 2 . Suppose condition 4 fails, i.e, $-p[u(w-g^*)-u(w)] < r$.

Thus if $g_1 \leq g^*$ the bad type would have an incentive to deviate and donate, i.e., the IC constraint is binding for bad types. Thus $g_1 = g^*$ cannot be part of equilibrium play. There are multiple separating equilibria. Consider the separating equilibrium with the lowest level of contributions, which is the only equilibrium that will survive the intuitive criterion. Here, a good type's contribution $g_1^{\min}$ solves:

$$-p[u(w-g_1^{\min}) - u(w)] \;=\; r. \tag{5}$$

In this case, if the collection probability p decreases, the minimum level of conditional donations that separates types ($g_1^{\min}$) increases.[47]

Summarizing Cases 1 and 2. Thus, beginning at a value of p where the separation constraint does not bind, i.e., (4) holds with inequality, reducing p a small amount has no effect on conditional donations ($g_1 = g^*$) but lowers expected donations (pg^*). Reducing it further causes (4) to no longer hold, but permits only an

46. Note that, excluding reputation, the probability of collection does not matter for the optimal decision here.

47. We may also have pooling equilibria where both types contribute g^{pool} satisfying $g* \leq g^{pool} < g_1^{\min}$. These are possible where bad types are willing to contribute g^* even to gain the lower reputation $\tilde{R}(g^{pool}|pooling)$. I.e.,

$$-p[u(w-g^*) - u(w)] \;<\; \lambda r \;<\; r, \tag{6}$$

where the latter inequality is given to highlight that a pooling equilibrium could be ruled out under a weaker condition than condition 4. For lower values of p this equation holds for a wider range of preferences. However, the pooling equilibrium also does not survive the intuitive criterion. There is always a deviation that is only profitable for the good type, as he enjoys not only the reputational gain $(1-\lambda)r$ but also, unlike the bad type, a warm glow.

intuitive separating equilibrium where h's donate $g_1^{\min} > g^*$. Further reducing p increases $g_1^{\min}$ but lowers the probability the contribution is realized.[48]

The analysis extends to situations where income is not uncertain but the collection of donations is (by interpreting the collection probability as p and setting both w and ℓ to the realised income).

We can now compare across settings. For illustration—and resembling our lab experiment—assume that the probability of winning in the *Before* and *After* settings, and the probability the donation is collected in the *Uncertain Collection* setting are all $p = 1/2$. Suppose that the reputational benefit is such that case 2 applies for $p = 1/2$ while case 1 obtains if $p = 1$. This would imply that in the *Before* and *Uncertain Collection* setting good types will commit to donate $g^u = g_1^{\min} > g^*$. In the *After* setting (with the same income) corresponding to $p = 1$, good types will donate $g^a = g^* < g_1^{\min}$. Alternately, suppose case 2 held for both $p = 1$ and $p = 1/2$. Here donations in the *After* setting would be above g^*, but still below g_1^{min}, the commitment in the settings *Before* and *Uncertain Collection*.

Summarizing the above, where parameters are consistent with case 2 (under the *Before* or *Uncertain Collection* settings) this model yields Prediction 5.

Prediction 4. *Signaling generosity, where the separation constraint binds*

$$g_w^u \;=\; g_w^b \;>\; g_w^d \;=\; g_w^a$$

for good types, while bad types are unaffected by the treatment. A similar relationship will hold for donations from the lower level of income if condition 4 also fails at income ℓ , which need not be the case.[49]

A similar relationship will hold for donations from the lower level of income if the separation constraint binds (i.e., condition 4 fails) at ℓ, which need not be the case. Under the standard assumption that $u(\cdot)$ is concave, the parameter space where this holds at income ℓ is a proper subset of the parameter space where this holds at income w.

Note the arguments above do not automatically carry over to the *Before Both* setting: if the reputation takes into account both types of conditional donations, then a bad type who donates only conditional on winning would fully reveal his type. Here, we do not make explicit predictions for this setting.

A.2.1 Signaling model, considering heterogeneity. This model can be extended to reflect signaling where individuals can be publicly identified by a certain characteristic, e.g., gender, and the groups are known to have different type distributions and utility parameters. If we allow all the individual parameters in Equation 1 to differ by the group's observable characteristic, case (1) is more "likely" to hold for groups with a smaller reputation motive (smaller r) relative to the warm glow term (of good types in that group). I.e., as r declines the parameters move towards case 1 above, and if $R(\cdot)$ is not present the results are as in the expected utility model. Thus, under some background environments case (1) may hold for one group, e.g., women, while case

48. The net effect on expected contributions $pg_1^{\min}$ depends on the concavity of the material sub-utility function $u(\cdot)$ We have $-p\frac{dg_1^{min}}{dp} = \frac{u(w)-u(w-g_1^{min})}{u'(w-g_1^{min})} \geq g_1^{min}$ (where the latter inequality follows iff u is convex), implying $\frac{d(pg_1^{min}(p))}{dp} \leq 0$ if and only if u is convex. Thus, under a standard assumption of diminishing returns to own-consumption (concave $u(\cdot)$), lowering p will reduce *expected* contributions.

49. Under the standard assumption that $u(\cdot)$ is concave, the parameter space where this holds at income w is a proper subset of the parameter space where this holds at income l.

(2) may hold for another group and the donation commitments will respond to the uncertain collection for the latter group only.[50]

A.3 Loss Aversion and Reference Points

We now show that our main model of loss averse preferences (as introduced in Section) yields an extended version of Prediction 3 . Suppose indeed the reference point always corresponds to the expected future income at the point of the decision, the maximum own-consumption one could achieve if one's "investments" paid their expected value. In the certain income treatment, the reference point corresponds to the certain income, ℓ or w, implying, $x \leq \pi$. Hence, the optimal donation g_E^*, for income $E \in \{\ell, w\}$, is given by the first order condition:

$$(1+\delta)u'(E-g_E) \;=\; \omega'(g_E).$$

The same holds for the After treatment, as uncertainty has already been resolved, so we can note that for income $E \in \{\ell, w\}$ it is true that $g_E^d = g_E^a = g_E^*$

However, *Before* donations may differ: Here, $\pi = z \equiv \ell + p(w-\ell)$. Consider the case where $w - g_w > z$. In this case, the derivative of the utility function v at g_w^* (as defined above) will be positive, since $\frac{dv(g_w,\pi)}{dg} = -u'(w-g_w) + \omega'(g_c) > -(1+\delta)u'(w-g_w) + \omega'(g_w) = 0$.[51]

Hence only *conditional* donations differ from the benchmark, as donations up to $w-z$ are donations from gains, so that it becomes possible that $g_w^{bb} = g_w^b < g_w^*$ as such donations incur further losses with respect to the own-consumption reference point. However, there is no effect for the case of low income, nor if $w - g_w^* \leq z$, as an ex-ante commitment to donate from ℓ, or to donate past z, *will* incur a psychological loss. The income net-of-donations will then fall below the reference point, so that the donation amount will satisfy $(1+\delta)u'(\ell - g_\ell^b) = \omega'(g_\ell^b)$.

Prediction 5. *Loss Aversion, expected income, immediate adjustment*

$$g_w^{bb} \;=\; g_w^b \;>\; g_w^a \;=\; g_w^d \,\textit{(provided } g_w^a \;<\; w-(pw+(1-p)l)\textit{) and}$$
$$g_\ell^{bb} \;=\; g_\ell^b \;=\; g_\ell^a \;=\; g_w^d.$$

Note that where we observe $g_w^a > w - z$ for a particular individual this model implies $g_w^a = g_w^b > w - z$ would hold for the same subject.

Robustness to intermediate reference points (with adjustment)

The analysis above generalizes to any intermediate reference point. Suppose it always satisfies $\pi : \ell \leq \pi < w$ in the *Before* treatment, while it is given by ℓ or w in the other treatments. If $w - g_w > \pi$, for high income, there will be higher donations in the *Before* treatments ($g_w^b > g_w$) than in the *After* or *Benchmark* treatments. The lower is π, the larger the set of preferences over which $w - g_w > \pi$ will hold. For the "minimum income" reference point $\pi = \ell$, $g_w^b > g_w$ as long as the individual prefers to choose a positive level of consumption under certain income w.

50. Note that if a greater *share* of one group are good types, perhaps implying a larger λ, this will only affect the conditions for a pooling equilibrium but will not affect our conditions for cases 1 and 2.

51. The maximization problem is concave, but not necessarily differentiable. Hence, the optimal solution might be $g_w^b = w - z$.

$\pi < \ell$ implies that the individual's reference point is below the lowest possible outcome. $\pi = 0$ might be interpreted as a "status quo" reference point if the individual does not count any unresolved income in her reference point. However, this would seem paradoxical, as only *part* of the income $(w - \ell)$ is unresolved, and the income ℓ can be seen as certain. One might argue that until income is held "in hand" it is less *tangible* and thus easier to part with (see Reinstein, Riener, 2012a, on this point). However, in the experiments of the present paper, it is hard to see how the base income ℓ in the *Before* treatments is less tangible than the income in the *Benchmark* and *After* treatments; both are promises on a computer screen.

Loss averse: Expected income, no adjustment

Here we modify the above and assume that subjects' reference point corresponds to the original expected value income throughout the relevant decision period, and obtain a slightly different prediction. With this modification, *Benchmark* donations are unaffected, while the *After* levels will now correspond to the aforementioned *Before* levels, as they have the same reference points. This is summarized below.

Prediction 6. *Loss Aversion, expected income, no adjustment*

$$g_w^{bb} = g_w^b = g_w^a > g_w^d \text{ and}$$
$$g_\ell^{bb} = g_\ell^b = g_\ell^a = g_\ell^d.$$

If the reference point does not adjust rapidly, then donations from an anticipated or actual *win* (loss) will be higher (lower) than donations from income that was not subject to uncertainty.

A.4 Affect

We continue the discussion from section 2.4.

Putting this together we might predict greater generosity *after* a prize has been won, relative to *before* the prize outcome is known, and relative to a certain income. We might also predict *lower* generosity after *failing* to win the prize relative to after a certain income (although the "negative state relief" model of Cialdini et al., 1973, predicts the reverse). If individuals in the *Before* setting do not anticipate their change in mood from winning or losing, and if neither non-random earnings nor facing a lottery directly affects mood, then (ignoring other effects) the conditional commitments in the *Before* setting will equal the *Benchmark* donations for the corresponding income levels.[52]

Prediction 7. *Affective state*

$$g_w^a > g_w^d = g_w^b = g_w^{bb},\ g_\ell^a < g_\ell^d = g_\ell^{=} g_\ell^{bb}$$

B Appendix: Alternative models

B.1 Koszegi-Rabin Preferred Personal Equilibrium model

Prediction 3 does not extend to the Koszegi et al. (2006) *preferred personal equilibrium (PPE)* model. In a PPE for the *After-win* case, (even assuming the reference point forms *after* the lottery) the donor does not "give from

52. Similar predictions could arise out of an (indirect) reciprocity model (see Simpson et al., 2008), e.g., if the lottery's sponsor were the charity itself, or were believed to be sympathetic to the charity; the reciprocity motive would also have to depend on the realization of the "gift" and not only its probabilistic implementation.

gains"; the personal equilibrium anticipates her donation, and the standard utility-maximizing bundle yields a preferred personal equilibrium. Similarly, in the *Before* setting the same standard expected-utility maximizing choice yields a PPE. Thus, a straightforward application of the PPE model predicts $g_w^b = g_w^a$. We demonstrate this below.

Consider the PPE model, with the two dimensions consumption income (c) and donation g. In line with their model, denote the standard utility by:

$$m(c,g) \;=\; m_c(c) + m_g(g) \;=\; u(c) + \omega(g)$$

and assume, for the gain-loss component, $\mu(x) = \eta x$ if $x \geq> 0$ and $\mu(x) = \eta\lambda\eta x$ if $x < 0$ so that utility given reference point r becomes:

$$u(c,g|r_c,r_g) \;=\; u(c) + \omega(g) + \mu(u(c) - u(r_c)) + \mu(\omega(g) - \omega(r_g)).$$

Consider the *After-win* context. Assume that the decision maker forms a reference point *after* she has learned if her income is w or ℓ. By proposition 3 in Koszegi-Rabin, with respect to the *preferred* personal equilibrium (PPE) the utility maximizing donation level is the one that maximizes standard utility:

$$u(w-g) + \omega(g).$$

Thus, the PPE model suggests that loss-aversion has no effect on the *After* donation. Denote the solution by g_w^a, from FOC: $-u'(w - g_w^a) = \omega'(g_w^a)$.

Consider now the *Before* case. Here one might apply the model by assuming that the decision maker forms reference lotteries before she knows her income. We investigate now if setting the same donation, i.e., $g_w^b = g_w^a$ remains a personal equilibrium (PE). (If so, it is also a PPE as it maximises ex ante EU.) To specify the conjectured equilibrium reference point lottery, notice that donating g_w^a contingent on winning implies that with probability p the consumption component c and hence r_c is $w - g_w^a$, and the donation component g, hence r_g, is g_w^a. With probability (1-p) the consumption component is ℓ and the donation component is zero (recall we are considering the *Before* treatment, where the donation is solicited for only the winning state).

Given this reference lottery, we now evaluate possible deviations from the equilibrium with respect to the donation (pertaining to the high income, of course). The utility of donating any g conditional on winning with reference points corresponding to g_w^a is (given there is no choice in the low-income state):

$$(1-p)[u(l) + p\eta\lambda(u(l) - u(w-g_w^a) + \omega(0) - \omega(g_w^a)] + p(u(w-g) + \omega(g)) + p$$
$$\begin{cases} (1-p)\eta\lambda(u(w-g) - u(l) + \omega(g) - \omega(0)]], & \text{if } g \;=\; g_w^a \\ [p\gamma[u(w-g) - u(w-g_w^a) + \lambda(\omega(g) - \omega(g_w^a)) + (1-p)\gamma(u(w-g) - u(l) + \omega(g) - \omega(0)]], & \text{if } g \;\leq\; g_w^a \\ [p\gamma[\lambda(u(w-g) - u(w-g_w^a)) + \omega(g) - \omega(g_w^a) + (1-p)\gamma(u(w-g) - u(l) + \omega(g) - \omega(0)]], & \text{if } g \;\geq\; g_w^a \end{cases}$$

Interpretating the first line: If I get low income, in addition to normal utility I feel a loss relative to the high -state (with has probability p) in both dimensions.

Interpreting the second line: If I get high income I will feel a gain relative to low income in both dimensions (unless g_w^b exceeds the difference in incomes). If I donate more than planned, I will have a loss in own consumption (but a gain in warm-glow). If I donate less, I experience a gain in own-income and a loss in donation income.

Intuition: Since the loss will always weigh more strongly than an equivalent gain, and g is optimally chosen w.r.t normal utility, deviations are not profitable.

This means that at $g = g_w^a$ the right derivative (higher g) is proportional to $-\lambda u'(w - g_w^a) + \omega'(g_w^a) < 0$ while the left derivative (lower g) is proportional to $-u'(w - g_w^a) + \lambda\omega'(g_w^a) > 0$, so that indeed g_w^a remains optimal.

Additionally note that equilibrium condition of the PPE model entails that if a decision maker forms a plan (in the context of determining the reference lottery), he will follow it through even if she is given opportunity to revise it later.

Finally, we note that the weaker concept of *personal equilibrium* instead of the *preferred* personal equilibrium may yield a larger range of predictions.

B.2 Tangibility

If uncertain winnings are less "tangible" than the same winnings after the uncertainty has been resolved this might also explain the patterns we see in the field experiment. There is abundant evidence for different mental accounting over different types of earnings or wealth. Several economists have found that subjects who play with standard laboratory "endowments" make less self-interested choices than when they use money they have either "earned" through a laboratory task or brought from outside the lab (Cherry et al., 2002; Hoffman, Spitzer, 1985; Burrows et al., 1994).

Reinstein, Riener, 2012a note that "people may treat money they are promised (or are given in the form of tokens) differently than cash they physically hold—we call this the *tangibility* effect, and find significant evidence from a laboratory charitable giving experiment supporting this. Along similar lines, Oberholzer-Gee et al., 2004 argue that subjects do not fully consider the opportunity costs of the funds they give away in experiments. Breman, 2011, offers field experimental evidence that people are more generous in making commitments to charity with *future* income rather than *present* income. None of these experiments document commitments made with truly *uncertain* income, but to the extent to which all of these endowments are broadly less tangible, these make a similar case.

In our experimental context, the predictions of the Tangibility model are the same as under Loss Aversion with a status-quo reference point. Both of these predict more giving in the *Before Both* treatment from low income relative to the *Benchmark* with the same income; we do not observe this. Neither of these predict the response to the *Uncertainty* treatment that we observe for males.

B.3 Uncertainty aversion (ambiguity)

Risk aversion, as explained by diminishing returns to consumption, will not predict any difference in donations between our treatments. However, if we assume (i) people inherently value uncertain and unallocated income (i.e., income that can be used for later choices, including consumption or charitable giving), less than certain unallocated income; and (ii) value uncertain committed donations as much as certain committed donations, then this may predict a greater willingness to commit from a gain (ex-ante). By "value uncertain income less," we mean that, ex-ante, the marginal utility of an additional unit of unallocated income that occurs with probability p is valued at less than p times the utility of the additional unit of unallocated income.

Intuitively, contributing from a gain reduces expected personal consumption, but it also reduces the uncertainty over this consumption. Giving up income solely to reduce uncertainty might never be valued in itself (although this might be predicted by "direct risk aversion", Simonsohn, 2009) but it may induce greater donations where there is also some additional from committing to make a donation. If the uncertainty is also "Knightian", i.e., ambiguous, committing to contribute from the gain state will also reduce the magnitude of this ambiguity.

B.4 Adaptation, habituation, relativity with complementarity

A simple model of rapid unanticipated adaptation/habituation (Helson, 1964) to the prize money in the presence of *anticipated* complementarity of happiness (or perceived wealth) and generosity should predict a greater conditional commitment *before* winning than the donation if asked *after.* Intuitively, people believe that contributing yields a greater marginal utility when they are feeling happier or more wealthy. They also overestimate how happy or wealthy the prize will make them feel[53]—thus, they would commit to contribute generously. On the other hand, individuals who win the prize quickly adapt to having won the prize and their happiness may be extremely short lived. If asked to donate even a few moments after they learn that they have won, they see themselves as "of average happiness" or "moderately well off", and they donate less they committed in the conditional *Before* environment.

In our experimental context, the predictions of the Adaptation model are the same as under Loss Aversion with an expected income reference point, where the reference point does not adjust within the relevant time period. However, we do not expect this motive to play a strong role in our experiment, as the adaptation typically discussed occurs over a much longer interval. However, this may be relevant in some of the real-world contexts we discuss.

53. There is extensive evidence suggesting that people overestimate the effect of good and bad events on their happiness; see Loewenstein et al., 1999 for a survey.

C Appendix: Further experimental descriptions and details

C.1 Project chronology, explanation

Our first experiment was pooled with a Valentine's card site; this was done both to provide a context that distracted participants from considering the purpose of the experiment, and also to test hypotheses surrounding "fear of losing face" (Gall and Reinstein, 2017). As seen above, the results from this experiment was strongly significant; the before treatment led to substantially higher commitments. However, some questions remained about the sincerity of these commitments; as noted above, many participants did not fulfill their pledges.

We next sought to garner evidence in a context where the commitments were binding, and thus the sincerity could not reasonably be doubted. This led to our Employability experiment. The pairing with the employability arms provided a distraction, but it also provided an additional source of funds for this experiment (as this was a university priority), and an opportunity to test the impact of employability reminders on later career performance (we have not had the opportunity to follow up on the latter). Here we found no strong overall effect of the treatment itself, but we found a strong effect for males, and a strong gender difference. However, we were concerned that this reflected the inherent biases from multiple hypothesis testing, and we looked for further evidence explicitly testing for a gender difference.

Over nearly the same time period as Employability, we planned the laboratory experiments (in Germany). As noted, these were designed with subtler treatments to test for mechanisms behind an observed difference between Before and After treatments. We also planned these to provide confirmatory tests/replications of our gender differential. However, in these experiments we did not find a gender differential, but we did find a baseline difference which was marginally statistically significant ($p<0.10$), as reported above.

At this point we reviewed our evidence and realized that we did not have strong power overall to detect an effect of what might be deemed a reasonable size, considering the magnitude of effects reported in previous publications. Furthermore, we were concerned that without having drawn a line in the sand, our results might be seen as ex post "just so" stories. Finally, we wanted to gain evidence from a relevant nonstudent population. We decided to take a more systematic approach. We planned and designed an experiment which we registered on the AEA RCT registry. In this registration we carefully describe the experiment, its goals, and the targeted sample size. As requested on this site, we noted our power calculations (estimating a minimum detectable effect size). We also registered a pre-analysis plan, specifying the hypotheses we intended to test and the nature of the statistical tests we planned to use.

We worked with the EssexLab to purchase a database of local nonstudent residents, to help them recruit this group to the lab. We also work with them to build an omnibus survey, to be given to all members of the laboratory subject pool. Again, this provided us a context as well as a source of some of the funds used for our incentives/endowments.

However, as noted in the pre-analysis plan, we decided to prioritize using the non-student participants for different experiment, and as the nonstudent response was smaller than expected, we did not have large enough numbers to also test "give-if-you-win" on the sample. Thus the Omnibus responses we report are for students only. However, because of the limited response rate and lack of nonstudents we decided it would be consistent with the spirit of our earlier preregistered plan to run further "give-if-you-win" experiments on the Prolific nonstudent sample. Thus, we subsequently recorded this change in our preregistration, and before running the Prolific experiment, we added this to the preregistration and pre-analysis plan.

C.2 Prolific, additional screenshots

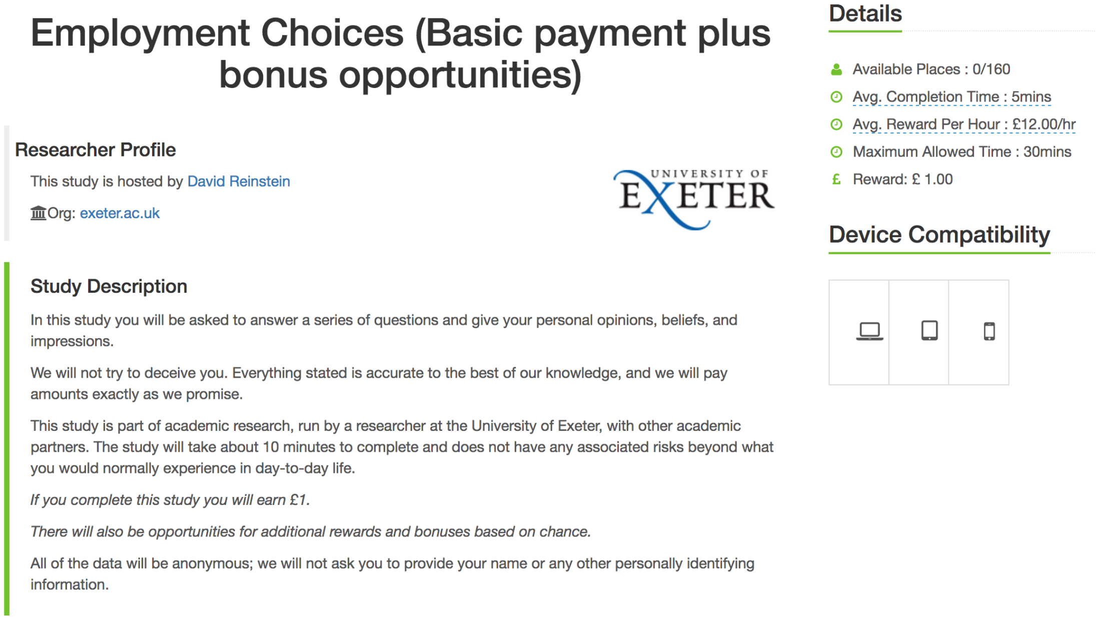


Figure C.0. Prolific entry page

C.3 Lab experiment

As noted above, the laboratory environment permits more control and a wider variety of treatments. The design is shown in Table C.1.[54] Subjects were seated at computer terminals and given a code number. They next performed a Raven's matrix task—a language-free multiple choice intelligence test—lasting about half an hour (Raven, 1936); this aimed to give the endowment the flavor of earned income, rather than a windfall or house money. Subjects were told they would be rewarded €7 for this (or €14 in more than half of the *Benchmark* and *Uncertain Collection* treatments) independently of their performance. Next, the *Benchmark* and *Uncertain* subjects were reminded of their earnings, and the rest were told "with a probability of 50 percent you will be rewarded a bonus of €7 on top of your already acquired income of €7."[55]

We took steps to demonstrate to the subjects that neither their performance nor their donation choices could affect their chances of winning. Each was given a printed code and pointed to a sealed envelope pinned to the inside of the laboratory door. They were told that the code would determine the "random" outcomes, and that they could check this against the sheet at the end of the experiment. This measure aimed both to rule out a direct material incentive to donate and to reduce the influence of magical thinking for subjects who believe that "karma" can influence future but not predetermined events.

Before Those in the *Before* treatment were given a chance to conditionally donate, before learning if they won the bonus, with the text (translated from German):

> In case of you winning the bonus of € 7, we now want to give you the opportunity to donate a part of the income you have earned in this experiment to a charitable organization. In doing so, you can choose between "Brot fur die Welt" (Bread for the World) and the "World Wildlife Fund (WWF)". [...] Please enter the amount of your donation in case of you winning the bonus (amount can be between € 0 and € 14). In case of you not winning the bonus, nothing will be deducted from your income and the organization will not receive a donation [...]

54. In the first wave of lab experiments we also included a *"Before Both"* treatment. Results are not sensitive to this: Table D.11 shows this treatment had similar effects as *Before*; further results are available upon request.
55. For those whose income was deterministic (*Benchmark* and *Uncertain*) and never expressed as a probability, we assigned more than half to the higher income. This allowed us greater power to distinguish between treatments from donation commitments from €14.

Table C.1. Laboratory - Experimental design

		Certain treatments		*Probabilistic treatments*
	Treatment	*Benchmark*	*After*	*Before*
	Income	Known	Bonus lottery	Bonus lottery
	Donation	Certain	Certain	Probabilistic
	Stage			
0	*Real-effort*	Task	Task	Task
1	*Learn income*	*€7 or €14*	*€7*	*€7*
2	*Bonus info*	No info	Possible €7 bonus	Possible €7 bonus
3	*Message 1*	Reminded flat-rate income	Learn bonus outcome (w or l)	*None*
4	*Giving decision*	Giving decision	Giving decision	Giving decision
5	*Message 2*	*None*	*None*	Learn bonus outcome (w or l)
6		*Belief elicitation and questionnaire*		

Before Both *Before Both* was identical to the *Before* treatment, except that subjects were asked, on the same screen, to choose a donation both for the case of their winning the bonus, and for the case of their not winning the bonus.

After *After* treatment subjects were first informed whether they received the bonus and then were given the opportunity to donate to the above organizations, with similar language as above.

Benchmark *Benchmark* subjects were also asked to donate from their (known) income, with virtually identical language as for the *After* subjects.

Uncertain Collection *Uncertain* subjects were told that with a probability of 50% they would have the opportunity to donate, and asked to enter a donation "in case of you being able to donate."

All subjects were told "we will increase your donation by an additional 25 percent taken out of our own budget." Following the donation decision, we asked the subjects to make a series of incentivized and hypothetical predictions, followed by survey questions. eWe first asked them to predict for what others donated; subjects were informed that they would be given €0.50 per answer that was within €1 of the correct answer. First, they were asked guess the average overall donation. We next told them the two possible earnings asked them to guess the average contribution from each level of earnings. Finally, we asked them a hypothetical question: what would their own donations have been had they earned the other income/bonus amount? Details of this part of the design, the incentives, and the results, are available by request. Finally, we revealed net earnings, and revealed to the *Uncertain Collection* subjects whether their donations would be collected. We opened the sealed envelope to demonstrate that the random draws had indeed been pre-determined. Payments were made and donations passed to the charities, with a subject monitor, as promised.

These experiments were run in Düsseldorf and Mannheim on a standard experimental subject pool (recruitment was conducted via ORSEE; Greiner, 2004), using virtually identical protocols and zTree code at each lab (Fischbacher, 2007). We ran nine sessions over five days in January–February 2013 and November 2014, and

24 sessions over 7 days in March–April 2016. A complete set of relevant screen shots and translations are available in our online appendix.

To protect anonymity, we were careful to ensure that lab subjects never learned *each other's* earnings or contributions, and we never connected an individual's identity to her treatment or her choices. Still, as noted in the introduction, we cannot rule out a signalling motive. In making payments, we (the experimenters) could infer how much each subject earned, which would have allowed us to make a probabilistic inference about her likely contribution. Subjects may have anticipated this, implying a possible "signaling to the experimenter" motivation. Furthermore, subjects may want to discuss their lab experience with others afterward. If it is common-knowledge that lying brings a strong internal moral cost, *reported* choices may hold a similar signaling power as actual verifiable choices. Finally, previous work suggests that subjects often bring real-world norms and heuristics into the laboratory (e.g. Burns, 1985; Hoffman, McCabe, et al., 1996).

C.4 Screen shots and further material

Further screen shots for all experiments available by request; viewable copy of Qualtrics survey instruments hosted online (see footnotes in main text). Additional screenshots, translations, and recruitment material is given in the external links and files described in section E.

Before

After

Figure C.1. Employability experiment screenshots

C.5 Valentine's experiment

In 2012 we ran an experiment tied to a St. Valentine's Day E-card web site accessible at three UK universities (Bristol, Essex, and Warwick). This was also advertised as a fundraiser for Right-to-Play, a popular international charity which had been endorsed by the Essex Student Union. Students and staff who completed a survey were randomly selected to win restaurant vouchers and were given the opportunity to donate from this voucher. The site was promoted through extensive flyering, postering, email lists to members of university organizations, and online media, including Facebook and Twitter. We offered a lottery for restaurant vouchers (worth roughly £20-30) as a participation incentive; participants were told the total number of restaurant vouchers to be given away, but not the exact chances of winning.[56]

We implemented two treatments, which we will refer to as *Before* and *After*, which were assigned orthogonally to other subtle design variations in earlier parts of the Valentine's promotion.[57] Individuals in both treatments were directed to a website informing them that the draw had taken place (so they had already won or lost, though they did not know which).

56. We advertised 75, 25, and 10 restaurant vouchers worth £20, £20, or £30 each at Essex, Bristol, and Warwick, respectively. The actual probabilities (ignoring the few who did not log in to check their prizes) were approximately 82% (75/92) at Essex, 32% (25/77) at Bristol, and 28% (10/36) at Warwick. The precise language (at Essex): "Here you have the opportunity to send Valentine's day E-cards to anyone with an Essex email. Just by logging in and completing the survey, you will get a coupon for 20% off at Naka Thai, and be entered into the draw to win one of 75 vouchers worth £20 for dinner at Naka Thai (on East Hill in Colchester). You do not have to send any E-cards." Again, a set of relevant screenshots is available in the online appendix.

57. In the earlier part, we varied whether a student's identity and her donation (from a prior donation request) would be revealed to the Valentine's e-card recipient(s).

Before In the Before treatment, participants were provided with information about the charity Right-to-Play and asked whether they were willing to donate £1 or more. After making their decision, they proceeded to a page letting them know if they had won. For winners this read "Congratulations you have WON the free dinner for two at [Restaurant name] (value [£30/£20/£20 at Warwick/Bristol/Essex respectively]). Please continue to learn how to claim your prize. For losers this read "Sorry, you have not won". Proceeding to the next page, those in the Before treatment received further instructions about how to claim their prize and how to fulfill their pledge.

After In the After treatment, participants were first directed to the page where they were informed whether they won. They were then asked to pledge to donate before learning *how* to claim their prize.

In this experiment pledges to donate were *not* binding; a donor had to follow through on her pledge by donating online, at the Student Union office, or at the restaurant itself. However, many did not fulfill their pledges; while 20 students owed a donation, our most inclusive measure suggests that only 12 of these made any sort of donation. In light of this, there are reasonable interpretations of these results, including:

(I.1) Students in both treatments may have pledged sincerely, and forgotten (and not seen our reminder emails) or found it too effortful to fulfill their small contributions.

(I.2) The *Before* treatment led to additional sincere pledges. However, at the point they were asked to *fulfill* their contribution, their income was certain and tangible, resembling the *After* treatment. This may have discouraged students from donating in spite of the disutility of cognitive inconsistency.

(I.3) Many *Before* pledgers never intended to fulfill their commitments; pledges may have been driven by magical thinking, a desire to please the experimenters, or simply carelessness.

Under I.1, *Before* would raise more than *After* if fulfillment were made easier. Similarly, *Before* would raise more under I.1 or I.2 with automatic deduction from prizes. However, under I.3, *Before* pledges (hence donations) under automatic deduction would fall to the level of *After* pledges.

Our evidence from the Prolific experiment—0/30 participants who could have lowered their contributions did so—suggests that I.3 is unlikely. However, as these contexts differed, the choices might have involved a different calculus. Still, our pooled results *excluding* the Valentine's experiment are similar (see previous version of working-paper).[58]

C.6 Employability

Our "Employability" field experiment was run in 2013/14 in the context of a promotion funded and announced by the University of Essex Faculty of Social Sciences.[59] Participants could win either a £20 Amazon or a £20 dinner voucher. Participants *knew* the (25%) probability of winning. Participants were informed "[…] your

58. Additional features specific to the Valentine's experiment: 1. All were asked whether they were willing to donate £1 or more. This aimed to increase the baseline rate of positive donation and enhance the experiment's power by legitimizing low-value contributions (see Weyant, 1978). 2. Losers were also asked if they wanted to donate (but 0/47 did so). 3. Losers at Essex were still given a small "prize", a 20% discount at Naka Thai. 4. We did not offer a matching contribution. 5. We provided a link for participants to verify the *total* contribution made. 6. There was only a single charity involved.

59. This faculty included Economics, Government, Sociology, Language and Linguistics departments, the Center for Psychoanalytic Studies, and later the Essex Business School. The first run was 4 June 2013 – 21 January 2014. Eligible students were in their first and second years through October 2013, and in their second and third (final) years in following academic year. The next run was 14 May–25 July 2014. We began with this same set of departments (excluding participants in the previous run), and ultimately expanded eligibility to all undergraduate students at the University of Essex, in order to use all of the prizes before our institutional deadline. The online appendix provides further practical details and a timeline.

donation will be automatically deducted from your prize and passed on to the charity of your choice, plus an additional 10% from our own funds […].”

Eligible undergraduate students were sent a series of emails, mainly from the departmental administrators, encouraging them to participate, with text such as the following.

> **Subject**: Employability promotion—a 1 in 4 chance of winning a £20 prize for doing a short survey.
> **Text:** Please go to [SITE]—we have 80 free dinners for two in Colchester to give away, worth £20 each and at least 40 £20 Amazon vouchers!! If you log on, you will have a one in four overall chance of winning one of these prizes!

This was also promoted through extensive flyering, postering, university web sites, and social media. We obtained 352 valid responses that involved a donation choice. No students were allowed to participate more than once. Participants first signed in with their email, department, and study year. Half were then asked to sign up for a job site (JobsOnline) and enter two jobs of interest to them.[60] Next, they were informed of which prize they had a 25% chance of winning (the prize selection was orthogonal to other treatments). After this they were presented our *Before* or *After* donation treatment (screenshots in appendix Figure C.1).

We offered a 10% matching contribution for all donations, and donations were publicly made and recorded on JustGiving, either anonymously or with a message, as the participant wished. We took these steps to offer an incentive to donate *within* the experiment and increase the baseline level of donations, and to reflect typical fundraising campaigns, which often involve matches and social incentives.[61]

60. This was one of two additional treatments administered orthogonally to the donation treatments, each for half the subjects. i. This “employability” treatment required half of participants to sign up for a jobs site and enter two jobs of interest. ii. A question and answer treatment asked about rates of employment and salary. The latter “information” treatment occurred *after* the donation treatment. We do not expect that the former treatment would have any effect on donation behavior, and our donation results do not differ substantially by this treatment. More details on these treatments and their assignment ordering are in the online appendix.

61. A copy of the experimental instrument can be tested at https://goo.gl/qSvhi1; this will cycle through each of the treatments.

D Appendix: Supplementary results

D.1 Randomization checks and summary statistics

Table D.2. Internet based experiments

Prolific	(1) Before	(2) After	(3) p-value	(4) N
Personal Income (GBP)	19500.00	20652.17	0.36	209
	(659.54)	(1176.61)		
Female	0.73	0.65	0.24	230
	(0.04)	(0.06)		
Age	28.03	25.93	0.01	230
	(0.53)	(0.39)		
Non-giver (self reported)	0.22	0.24	0.64	227
	(0.03)	(0.05)		
Gave up to 50 (self reported)	0.54	0.51	0.75	227
	(0.04)	(0.06)		
Gave over 50 (self reported)	0.25	0.24	0.93	227
	(0.04)	(0.05)		
N	162	78		
Omnibus	(1)	(2)	(3)	(4)
Female	0.63	0.62	0.84	460
	(0.03)	(0.04)		
Age	28.52	27.74	0.52	460
	(0.72)	(0.92)		
FirstLangEng	0.62	0.61	0.84	460
	(0.03)	(0.04)		
BirthNationUK	0.54	0.56	0.66	460
	(0.03)	(0.04)		
Siblings	1.84	1.80	0.80	460
	(0.08)	(0.13)		
DecisionModeImp	1.78	1.77	0.94	460
	(0.02)	(0.03)		
N	304	156		
Employability	(1)	(2)	(3)	(4)
Jobs Treatment	0.25	0.20	0.34	375
	(0.03)	(0.05)		
Female	0.53	0.59	0.37	323
	(0.03)	(0.06)		
Business	0.19	0.21	0.60	375
	(0.02)	(0.05)		
N	300	75		
Valentine's	(1)	(2)	(3)	(4)
Anonymity treatment	2.01	1.96	0.60	205
	(0.07)	(0.06)		
Donor treatment	2.47	2.46	0.91	205
	(0.05)	(0.05)		
Cards sent in Val St.	0.63	0.70	0.64	205
	(0.09)	(0.14)		
Year of study	3.29	3.23	0.83	204
	(0.19)	(0.19)		
Female	0.76	0.69	0.26	204
	(0.04)	(0.05)		
No religion	0.56	0.43	0.06	205
	(0.05)	(0.05)		
University	1.85	1.74	0.29	205
	(0.07)	(0.07)		
Previous volunteer	0.34	0.36	0.76	205
	(0.05)	(0.05)		
Previous donor	0.65	0.71	0.36	205
	(0.05)	(0.05)		
N	107	98		

Table D.3. Laboratory

	(1) Income Certain	(2) Before	(3) Before both	(4) After	(5) Uncertain	(6) p-value from joint orthogonality test of treatment arms	(7) N from orthogonality test
High Income	0.56	0.54	0.49	0.48	0.63	0.29	430
	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)		
Female	0.52	0.51	0.52	0.48	0.42	0.68	419
	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)		
Age	23.90	24.07	23.47	23.21	24.58	0.50	419
	(0.65)	(0.55)	(0.51)	(0.47)	(0.82)		
No religion	0.54	0.46	0.43	0.51	0.51	0.62	430
	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)		
Previous donor	0.33	0.35	0.37	0.30	0.27	0.66	430
	(0.05)	(0.06)	(0.05)	(0.04)	(0.05)		
N	79	74	79	115	83		

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.4. Further summary statistics: Prolific sample

	Mean	Std. dev.	N
Personal Income (GBP)	19880	(8498)	209
Female	0.704	(0.46)	230
Age	27.3	(5.80)	230
Non-giver (self reported)	0.225	(0.42)	227
Caucasian	0.829	(0.38)	240
Secondary school degree	0.350	(0.48)	240
Undergrad degree	0.267	(0.44)	240
Postgraduate degree	0.100	(0.30)	240
Labour party affil.	0.446	(0.50)	240
Has children	0.471	(0.50)	240
Identifies as monocultural	0.450	(0.50)	240

Notes: Key summary statistics from Prolific Academic sample, from previously collected survey 'screener' questions. Income imputed as mean of range-coding.

D.2 Results: Happiness, donations

Table D.5. Happiness, winning, and donations

	(1) Happiness at end, normalized	(2) Happiness at end, normalized	(3) Donation share (Prolific)
Won prize	0.75***	0.31	
	[0.38,1.13]	[-0.085,0.70]	
Omnibus expt	-0.39***	-0.63***	
	[-0.50,-0.28]	[-0.84,-0.42]	
Won prize, Omnibus expt		0.64***	
		[0.24,1.05]	
Happiness at start, normalized			-0.0021
			[-0.029,0.025]
Observations	460	460	240

Notes: This table reports coefficients, 95% confidence intervals, standard errors, and t-test p-values from OLS regressions. Data from Omnibus and Prolific experiments (columns 1-2), and Prolific only (column 3). Happiness variables de-meaned and divided by standard deviation, derived from self reported rating scales. Results of t-tests indicated at following significance levels * p<0.1, ** p<.05, *** p<.001..

D.3 Power calculations

Table D.6. Power calculations

	Min	d=0.2	d=0.5	d=0.8
Prolific	0.463	0.227	0.856	0.998
Omnibus	0.274	0.533	0.999	1.000
Employability	0.388	0.303	0.950	1.000
Laboratory	0.494	0.205	0.809	0.995
Valentine's	0.688	0.129	0.530	0.902
Pooled	0.166	0.922	1.000	1.000
Pooled preregistered	0.225	0.700	1.000	1.000

Notes: This table reports results from power tests for pairwise t-tests using the actual means of the proportion of the endowment donated in the After treatment, standard deviations (for each treatment separately), and observations for each experiment. Column 1 (Min) shows the standardized minimum detectable effect size between the Before and After treatments. Columns 2 to 4 report the power to detect a Cohen's d of 0.2, 0.5, and 0.8, respectively.

D.4 Heterogeneity and nonlinearity

Table D.7. OLS on Donations: Age, gender and religiosity, pooled over experiments

	(1) Proportion	(2) Level	(3) Incidence	(4) Proportion	(5) Level	(6) Incidence
Before	0.033***	0.49***	0.089***	0.031**	0.48**	0.074***
	[0.008,0.059]	[0.146,0.828]	[0.039,0.138]	[0.002,0.061]	[0.105,0.861]	[0.021,0.128]
Female (centered)	0.065***	0.97***	0.13**	0.053**	0.79**	0.095*
	[0.016,0.114]	[0.300,1.631]	[0.030,0.230]	[0.005,0.101]	[0.131,1.443]	[-0.005,0.195]
20 to <30 years	-0.011	-0.22	-0.15*	0.090***	0.65**	0.14***
	[-0.105,0.084]	[-1.358,0.920]	[-0.317,0.013]	[0.044,0.135]	[0.035,1.271]	[0.047,0.240]
30 to <40 years	0.10	1.15	-0.12	0.23**	2.30**	0.18
	[-0.088,0.293]	[-1.137,3.436]	[-0.375,0.141]	[0.049,0.405]	[0.155,4.440]	[-0.043,0.406]
40-50 years	-0.075	-0.97	-0.18	0.088	0.63	0.11
	[-0.295,0.145]	[-3.614,1.668]	[-0.560,0.193]	[-0.116,0.293]	[-1.834,3.100]	[-0.233,0.456]
Non-religious	0.019	0.18	0.0090	0.020	0.23	0.018
	[-0.019,0.057]	[-0.299,0.652]	[-0.125,0.143]	[-0.023,0.064]	[-0.328,0.793]	[-0.124,0.160]
Before ×	-0.045	-0.73	-0.059	-0.036	-0.68	-0.026
Female (centered)	[-0.106,0.017]	[-1.617,0.149]	[-0.175,0.058]	[-0.096,0.024]	[-1.568,0.207]	[-0.139,0.087]
Before × 20	-0.017	-0.12	0.084	0.028	-0.041	0.087
to <30 years	[-0.143,0.110]	[-1.646,1.398]	[-0.123,0.292]	[-0.028,0.084]	[-0.779,0.697]	[-0.023,0.196]
Before × 30	-0.15	-1.69	0.014	-0.12	-1.77	-0.0024
to <40 years	[-0.370,0.076]	[-4.350,0.966]	[-0.298,0.326]	[-0.315,0.081]	[-4.143,0.608]	[-0.268,0.263]
Before ×	0.12	1.40	0.15	0.12	0.90	0.12
40-50 years	[-0.162,0.399]	[-1.923,4.729]	[-0.291,0.598]	[-0.140,0.373]	[-2.154,3.952]	[-0.278,0.527]
Before ×	-0.032	-0.21	0.016	-0.044	-0.42	-0.014
Non-religious	[-0.088,0.025]	[-0.907,0.483]	[-0.148,0.180]	[-0.105,0.016]	[-1.195,0.355]	[-0.179,0.150]
Constant	0.14***	1.91***	0.43***	0.14***	1.90***	0.41***
	[0.117,0.160]	[1.608,2.207]	[0.387,0.479]	[0.120,0.168]	[1.584,2.213]	[0.362,0.458]
DM_DExperiment	Yes	Yes	Yes	No	No	No
Observations	1296	1296	1296	1296	1296	1296

Notes: OLS regressions on donation shares of endowment, levels and incidence for Before versus After treatments; interacted with de-meaned gender, age, non-religious", and risk attitude categorical variables, excluding donations from the lower income level. Missing values of these variables are linearly imputed from other variables in this regression. All regressions include dummies for each experiment (hidden). Columns 4-6 also include interactions of the Before treatment with de-meaned experiment dummies (not shown). Dependent variables: (a) shares donated from endowment, (b) actual donation levels in Euros, (c) donation incidence. Cluster-robust standard at the session (date) levels for the lab (web-based) experiments account for potential correlated errors at these levels. T-tests at * p<0.1, ** p<.05, *** p<.01.

Below, we report maximum likelihood estimates of models of the form

$$Y_i^\alpha = \beta_0 + \beta_1 X1_i + \beta_2 X2_i + ...\beta_j Xj_i + \epsilon_i,$$

where $\epsilon_i \sim N(0, \sigma^2)$.

Table D.8. Power model (nonlinear, ML) of Donation shares (from higher income), pooled over experiments

Number of obs =1353
Wald chi2(8) = 231.47
Log likelihood = 85.77, Prob >chi2 = 0.000

Donation Proportion	**Coef.**	**Std. Err.**	**z**	**P>|z|**	**[95% Conf.**	**Interval]**
Before	0.056	0.009	6.450	0.000	0.039	0.073
Dummy: Age<20	-0.137	0.034	-4.060	0.000	-0.203	-0.071
Dummy: Age 20-30	-0.053	0.020	-2.630	0.008	-0.092	-0.013
α	3.832	0.498	7.690	0.000	2.856	4.808
σ	0.227	0.004	52.020	0.000	0.219	0.236

Notes: Maximum likelihood estimates of power models (dependent variable raised to the power α) of donation shares as a function of treatment, and demeaned experiment and imputed age categories. Experiment/lab dummy coefficients hidden.

Table D.9. Power model (nonlinear, ML) of Donation shares (from higher income), pooled over experiments

Number of obs =1296
Wald chi2(17) = 184.69
Log likelihood = 67.41, Prob >chi2 = 0.000

Donation Proportion	**Coef.**	**Std. Err.**	**z**	**P>|z|**	**[95% Conf.**	**Interval]**
Before	0.060	0.011	5.57	0.000	0.039	0.081
Before × Dummy: Age<20	0.151	0.072	0.21	0.833	-0.125	0.155
Before × Dummy: Age 20-30	0.009	0.042	0.20	0.840	-0.074	0.091
Age<20	-0.147	0.056	-2.63	0.009	-0.256	-0.037
Age 20-30	-0.062	0.031	-1.97	0.049	-0.123	-0.000
Before × Female	-0.032	0.028	-1.13	0.258	-0.088	0.024
Female	0.041	0.022	1.82	0.068	-0.003	0.084
α	3.722	0.503	7.39	0.000	2.735	4.709
σ	0.230	0.005	50.91	0.000	0.221	0.239

Notes: Maximum likelihood estimates of power models (dependent variable raised to the power α) of donation shares as a function of treatment, and demeaned experiment, age category, and gender baseline, all interacted with treatment. Experiment/lab dummy coefficients and interactions hidden.

D.5 Pooled results: Robustness to modeling choices, stopping rules, and alternative approaches

In table D.10 below, we demonstrate robustness to the specification and modeling choices used in table 4. We report on all combinations of the following reasonable specifications; the original choices are given in bold.

- **For overall**, for preregistered
- **Outcome measures: amount, share of endowment, binary**
- Clustering/Standard Errors: Huber-white, cluster on date (where available), **clustering by date/field-of study**
- Specifications: **linear**, logit (for extensive margin), Negative binomial, Tobit

- Control variables: **Experiment dummy de-meaned and interacted with treatment**, experiment dummy without interacted term, ridge regression results with regularization over control variables only (for Prolific study only; as it contained the largest set of controls), Intuitive controls

Table D.10. Robustness to researcher degrees of freedom: outcome, specification, clustering, controls

Note: Notes here

(Above table in preparation (needs formatting), as requested by JPubE referee.)

Negative binomial For the continuous outcomes the true data generating process must be nonlinear, as giving can never be negative. For robustness to functional form mis-specification, we estimated a negative binomial model. This method is more robust to heterogeneity than the Tobit specification (Gourieroux C. et al., 1984; Greene, 1994).

Evidentiary value, P-curve approach

The p-curve approach Simonsohn et al., 2014 is meant to be applied to a set of results that may exhibit publication bias, to distinguish *evidential value* from spurious findings, in a way that is also (arguably) immune to p-hacking.

In our paper, we report results from a series of five experiments, of which, individually only two surpass the standard $p < 0.05$ threshold in our headline analysis. Using only two (or five) p-values can not yield a very strong statistical test for whether this constitutes "evidentiary value."

Nonetheless, in the spirit of p-curve analysis, we report the results from http://www.p-curve.com/app4/pcurve4.php of whether our (significant) coefficients provide evidentiary value or are suspiciously close to the threshold. This output, provided in the online appendix, reveals that considering the results of the "combination test, introduced in Simonsohn, Simmons and Nelson (2015) ... here both conditions are met, indicating evidential value."

As a more basic mode of meta-analysis, we also consider the traditional Combined Probability test (Fisher, 1925). This tests whether the observed p-values from several independent tests are likely to have arisen if the shared null hypothesis was true in all cases. This test statistic involves the summed logs of each p-value; calculating from the "levels" results of table 4:

$$X^2_{2k} := -2\sum_{i=1}^{k} ln(p_i) = -2(ln(0.0004) + ln(0.099) + ln(0.330) + ln(0.568) + ln(0.0004)) \approx 39.27.$$

Under independence, this is distributed $\chi^2(2k)$, where k is the number of tests (here, 5). This yields a "combined p-value" of below 0.0001.

Evidentiary value and data augmentation

There is renewed interest in the issue of statistical inference in the presence of sequential data collection with or without a pre-announced stopping rule (see especially Sagarin et al., 2014, Lakens, 2014, and Simmons et al., 2011). Conservative best practice is (arguably) to pre-register and commit to a definite sample size and commit not to seek further data. A more efficient approach is pre-registering a "sequential analysis" plan. Here, one commits to a process of continuing/stopping under prespecified conditions, and using a more stringent "critical p-value" in such a way that the overall process leads to the desired rate of type-1 errors (e.g., 5%). The most naive/opportunistic approach to data augmentation (aka "preferential stopping" Sanborn et al., 2014), labeled as a "questionable research practice" by John et al., 2012, is to run a series of experiments, stopping as soon as the aggregate data yields a $p < 0.05$, and continuing otherwise (or as long as the results look "promising", e.g., as long as $p < p_{max}$.) Such an approach will lead to an inflated type-1 error rate, slightly or substantially above the nominal $p < 0.05$. If researchers do this, and editors are willing to publish non-preregistered results that attain $p < 0.05$ significance, then our body of evidence will have a greater than 5% type-1 error rate.

This naive, opportunistic approach does not characterise our process, which is describe in the appendix C.1. E.g., our first experiment was strongly significant; with a naive opportunistic stopping rule we would have proceeded no further. We also see these pre-registered final experiments as an isolated "clean" dataset yielding strongly significant results.

However, even if we *had* used a naive opportunistic stopping rule, our results still would suggest a strong effect. Our aggregated p-values (with or without the non-preregistered; ranging from 0.004 to below 0.000 in table 4) are well below the standard critical values. It seems highly unlikely that, under the null hypothesis, such a result would arise for *any* reasonable stopping rule.

Suppose that researchers use the standard null hypothesis significance testing (NHST; frequentist) framework, and stop (and publish) whenever $p < 0.05$ and continue adding data otherwise. Such researchers will publish spurious findings greater than 5% of the time. Any such stopping rule yields, in aggregate, a type-1 error rate above 5%. Thus, adjusting the p-value for this augmentation process always leads to a p-value of the *process* above the nominal 5%.

However, these spurious findings will only *rarely* be accompanied by p values as low as 0.004.

We report this adjustment below, the so-called $p_{augmented}$.

As noted by Sagarin et al. (2014):

> $p_{augmented}$ will always exceed [the nominal critical value]. Given this, we recommend that reviewers, editors, and readers offer some flexibility toward researchers voluntarily disclosing post-hoc dataset augmentation, accepting, for example, the above final p value of .02 and $p_{augmented}$ range of .055 to .057 as providing sufficient evidence for a confident interpretation.

Our reported $p_{augmented}$ values of course exceed $p = 0.05$, as must arise mechanically, while our final p values are well below the just-mentioned .02.

The authors presumably make this argument for "flexibility" in light of a Bayesian understanding.

A process involving stopping "whenever the nominal p.0.5" and gathering more data otherwise (even rarely) must yield a type-1 error rate above 5%. Even if the subsequent data suggested a "one in a million chance of arising under the null" the overall process yields a 5%+ error rate. The NHST frequentist framework can not adjust ex-post to consider the "likelihood of the null hypothesis" given the observed data, in light of the shocking one-in-a-million result. While Bayesian approaches can address this, we are not highly familiar with these methods; however, we are willing to pursue this if you feel it is appropriate.

We could also justify this in a frequentist context. Considering the calculations in Sagarin et al., 2014, it is clear that $p_{augmented}$ should *overstate* the type-1 error of the process if there is a positive probability that after an initial experiment attains p< 0.05, more data is collected. A headline $p < 0.05$ does *not* imply that this result will enter the published record. Referees may be skeptical of other parts of the design or framework or motivation. They may also choose to reject the paper specifically because of this issue; they believe the author would have continued collecting data had the result yielded $p > 0.05$, thus they think it is better to demand more evidence or a more stringent critical value. Prompted by the referee, the author may collect more data even though $p < 0.05$. Or, she may decide to collect more data even without a referee report/rejection demanding it, for various reasons (as we did after our Valentine's experiment). Thus, we might imagine that there is some probability that after (e.g.) an initial experiment attaining p<0.05, more data is collected, implying that $p_{augmented}$ as calculated above overstates the type I error rate that would arise from these practices. As referees and editors, we should be concerned about the status of knowledge as accepted by the profession, i.e., in published papers. If we recognize the possibility of data augmentation after any paper is rejected, it might be a better practice to require a significance standard substantially below $p = 0.05$, in order to attaina type-1 error rate of 5% or less in our published corpus. (In the terms of Sagarin et al, this practice may amount to *enforcing* a valid "adjusted p_{crit}" in light of prevailing incentives to augment data.)

D.6 Additional results by experiment

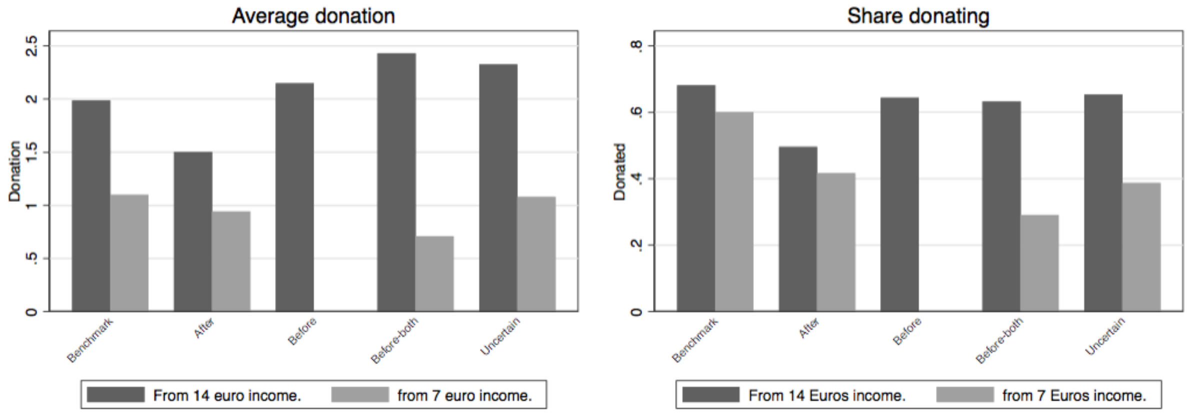


Figure D.10. Mean share committed by experiment, by Before vs. After

Table D.11. Laboratory: Donation amounts and incidence (OLS)

	Levels		Incidence	
	(1) High income	(2) Low income	(3) High income	(4) Low income
Treatment	ref.	ref.	ref.	ref.
– Income Certain	-0.27	0.27	-0.0047	0.22
	[-1.31,0.78]	[-0.50,1.05]	[-0.23,0.22]	[-0.082,0.53]
– Before	0.49		0.097	
	[-0.39,1.36]		[-0.066,0.26]	
– Before-both	0.18	-0.29	0.040	-0.17
	[-0.83,1.18]	[-0.92,0.34]	[-0.16,0.24]	[-0.39,0.046]
– Uncertain	0.072	0.031	0.0050	-0.065
	[-1.02,1.16]	[-0.89,0.95]	[-0.18,0.19]	[-0.38,0.25]
Constant	2.25***	0.90***	0.65***	0.43***
	[1.40,3.11]	[0.46,1.34]	[0.50,0.80]	[0.28,0.58]
Observations	304	205	304	205

Notes: This table reports coefficients, 95% confidence intervals, standard errors, and t-test p-values from ordinary least squares regressions on donations by Treatment for the lab experiment. The Benchmark treatment (no income or donation uncertainty) is the base group. As dependent variables we use (a) the levels of donations in Euros (Columns 1-4) and (b) donation incidence (Columns 5-8). In the Before-both treatment each subject made two choices – donation commitments from high income (if you win) and from low income are reported in the corresponding columns. We account for potential session-specific correlated errors by cluster-robust standard errors. Results of t-tests indicated at following significance levels * p<0.1, ** p<.05, *** p<.001.

D.6.1 Additional laboratory results. For the Before-both treatment, where the subjects are asked to make pair of conditional choices, one for the each income state, we also find a lower level and incidence of donation from the *lower* level of income relative to the benchmark (as well as relative to the Before-both choice for the winning state). As noted above, there are multiple interpretations of Before-both choices, so we do not highlight this result. Finally, we find that subjects with the lower income are less likely to donate in the *Uncertain-collection* treatment relative to the benchmark, where their income faced no uncertainty. This loosely suggests that the signaling model is not driving responses; however alternative explanations are possible, this is significant only at p<0.10, and we did not try to replicate this in other experimental environments.

D.7 Robustness checks, by experiment

Table D.12. Negative Binomial Regressions: Donations by experiment

Panel A: Levels	Prolific	Lab	Employability	Omnibus	Valentines
Donation from hi income					
Before	0.66***	0.39*	0.27	0.076	1.30***
	[0.26,1.05]	[-0.048,0.82]	[-0.31,0.86]	[-0.19,0.34]	[0.42,2.18]
Constant	-0.026	0.62***	0.46*	1.09***	-1.06**
	[-0.37,0.32]	[0.30,0.94]	[-0.069,0.98]	[0.87,1.31]	[-1.88,-0.25]
lnalpha	0.35**	-0.25	1.95***	0.96***	1.54***
	[0.039,0.66]	[-0.74,0.23]	[1.73,2.18]	[0.78,1.14]	[1.13,1.95]
Observations	240	129	375	460	159

Notes: This table reports reports coefficients and 95% confidence intervals from on donations for the Before treatment versus the After treatment in each experiment, excluding donations from the lower income level.

D.8 Related experiments

Table D.13. Comparing charitable giving experiments: Summary statistics and power

Experiment	N	Endowment	Share donating	Share donated	Mean donation	SD	SD/Mean	Effect Size %
This study								
- Valentine's	159	£20	28%	4.1%	£0.83	1.63	197%	103%
- Employability	375	£20	31%	8.3%	£1.97	4.59	233%	25%
- Lab (all treatments)	430	€7,€14	61%	16.3%	£1.99	2.5395	128%	24%
- Omnibus	460	£20	54%	26.2%	£3.13	4.15	133%	8%
- Prolific	240	£10	53%	15.8%	£1.58	2.1515	136%	57%
Authors' working papers								
Reinstein: Berkeley Pilot (Wave 1)	49	$10	74%	21.0%	$2.10	2.06	98%	
Reinstein (2010) ... Wave 2	48	$20	65%	23.0%	$4.60	4.94	107%	18%
Published studies								
Eckel et al. (1996)	48	$10	73%	30.1%	$3.01	3.19	106%	
Karlan et al. (2007)	50,083		2.0%		$0.90	0.05	6%	19%
Huck et al. (2011)	25,000	N/A	4.1%			10		44%
Reinstein, Riener (2012a)	190	€5,7.5,10	57%	18.3%	€1.23	1.75	142%	52%
Reinstein, Riener, 2012b	192	8 €	77%	25.0%	€1.80	1.81	101%	72%
Jones et al. (2014)	150	$10	73%	37.3%	$3.73	3.49	94%	21%
Tonin et al. (2014)	196	£10	81%	47.9%	£4.79	3.31	69%	21%

Notes: £: UK pounds, $: US dollars, €: Euros, no inflation adjustments. In experiments with multiple donations, results reported for first Ask only. **Endowment**: Amount(s) paid to participants which could be donated; vouchers for Valentines, Employability, Omnibus. **N**: Observations with a giving decision (Valentines: excludes non-winners). **Effect size %**: divides first reported (regression) result by mean donation
Kellner et al - Lab: Donation from higher income reported for "Before Both" treatment

Also include:

E Additional material; linked/files

E.1 Additional materials and screens: Valentine's experiment

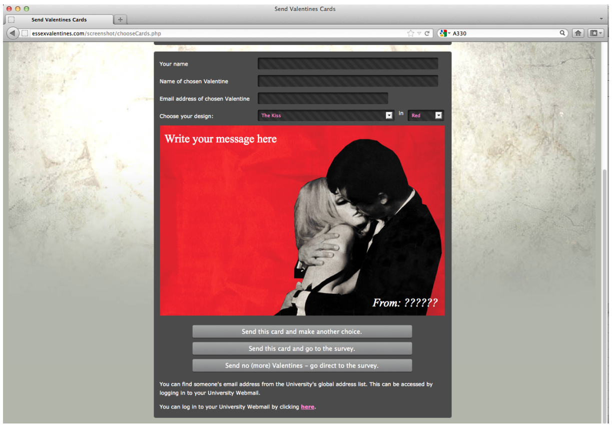


Figure E.13. Example Valentine's card

Participants in the Valentine's E-card website at any of the three universities were eligible to claim a prize. Above, a sample card is depicted.

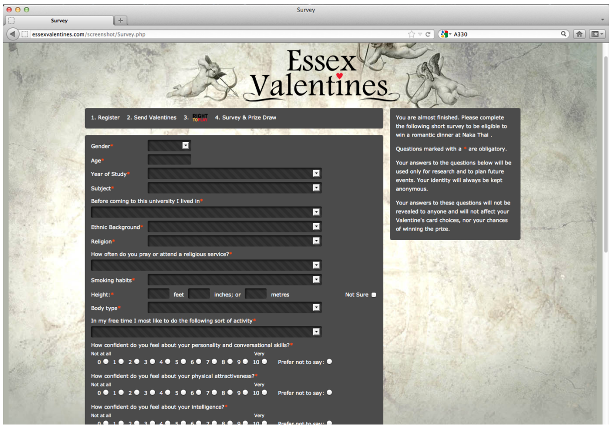


Figure E.13. Survey to be eligible for prize

To be eligible for the prize, they also needed to complete the survey, part of which is depicted above.

Above we display the charitable ask screens, for those in the "Before" treatment, and those in the "After" treatment who won and failed to win the prize, respectively. The University of Essex variants are displayed only (screens for Bristol and Warwick were appropriately adjusted).

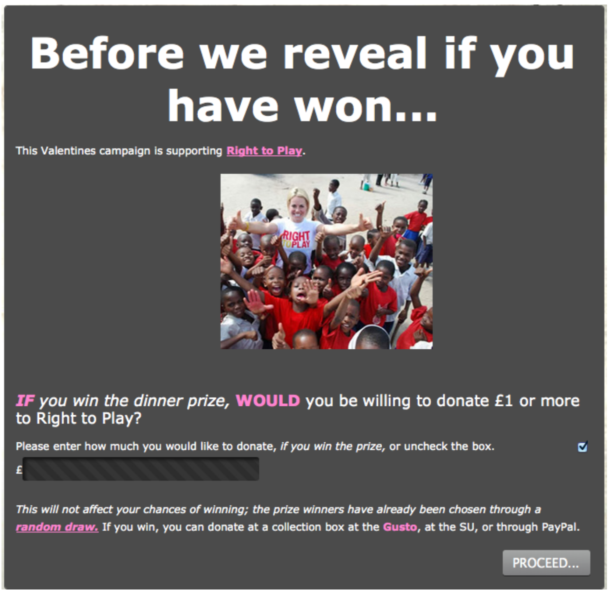


Figure E.13. Valentine's Before ask

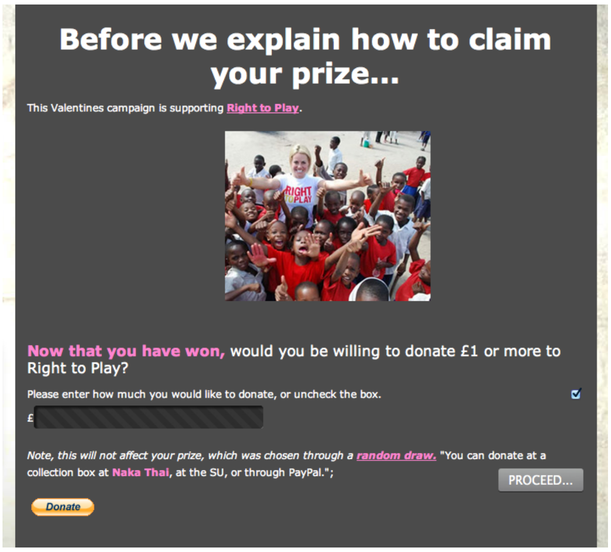


Figure E.13. Valentine's After (winners) ask

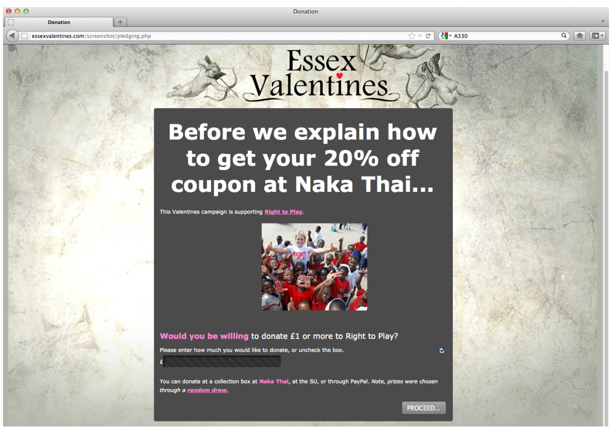


Figure E.13. Valentine's After (winners) ask

E.2 Additional materials and details: Employability experiment

Employability promotion (University of Essex). 12,240 total Essex students in 2013-14: 8,891 undergraduates and 3,349 postgraduates.

Key dates and actions

First Run (4 June 2013 – 21 January 2014)

- Prize switched (from dinner to Amazon voucher) after every approximately 40 winners.
- Next academic year (Autumn 2014)
- 25 September: changed site to have it say "second or third years," rather than "first or second years" , asked emails to be sent out again.
- 1 November: Adapted site to include Essex Business School (EBS), which joined the Faculty of Social Sciences in 2014
- 9 December – All relevant departments asked sent out emails (sent by 13/12 or 16/12) noting site would close by 31 December.
- Extended deadline until 20 January
- 21 January: closed survey, final download

Second Run (14 May – 25 July 2014)

Note: No first run participants were allowed to use the site sin the second run. Entry was strictly screened by a filtered Essex Email white list.

- 22 May, 27 May: Sent reminder emails (individually) to those who quit in the middle, giving them an opportunity to continue from where they left off (mainly at the stage of having to sign up for JobsOnline), with the same treatments.
- 23 June: Expanded to allow students in all University of Essex Colchester Campus departments, advertised to the largest departments via email from departmental administrators.

Advertising (both runs)

Advertisements:

- University and Student Union Societies Facebook and LinkedIn pages
- Careers center web site noted EEP.
- Twitter from employability coordinator.
- Posters around the university campus.

We repeatedly contacted the departmental administrators from each department, who sent a series of emails to students in participating departments. These were forwarded by administrators as coming from the Faculty of Social Sciences.

Examples of promotional material

Second run, after 23 June: Email sent by Departmental Administrators to undergraduate students in largest departments, on behalf of Essex Employability Prize

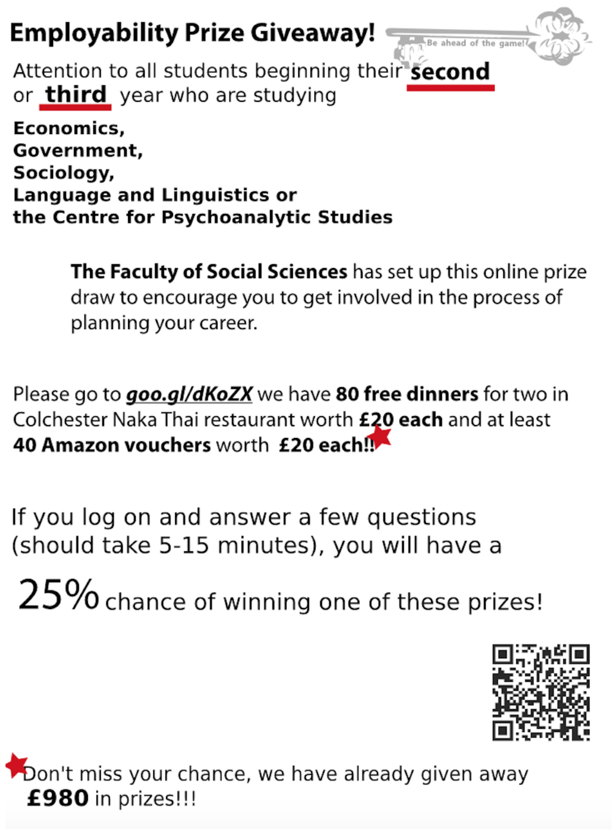


Figure E.13. Poster, first run, Autumn 2014

Subject: The Essex Employability Prize: 1 in 4 will win a £20 Amazon voucher

Body: We are writing to tell you about the Essex Employability Prize, designed to promote career development and awareness. Although this is sponsored by the Faculty of Social Sciences, they have just expanded eligibility to undergraduate students in all departments. They are giving away over £2000 in Amazon vouchers, and 1 in 4 who complete the survey will win a £20 Amazon voucher!

All undergraduate students at the University of Essex may now participate. You must complete the survey to the end to be eligible to win, and you can only enter the survey once. If you participated in a previous EEP you are not eligible, sorry.

The site will only be up through end of July, so please do not delay. Go to http://goo.gl/5DuppL and complete the short survey (5-15 minutes) for a 1 in 4 chance of winning a dinner an Amazon gift certificate for £20. (If clicking on the link does not work, please paste [full web address] into your browser)

Prizes. As advertised, all participants had a 25% chance of winning either a voucher £20 to be used at a local Thai Restaurant, or a £20 gift certificate from Amazon.co.uk. (As noted below, at the end of the survey, 1 in 4 were given an additional 1 in 10 chance of winning a £10 Amazon voucher). In the first run, both prizes were offered and advertised, and participants did not know which prize they were eligible for until the screen just before the one that revealed if they won (and in the *Before* treatment, asked if they wanted to donate). In the second run, only the £20 Amazon voucher was used and advertised. As mentioned above, all donations were deducted directly from these prizes, a 10% match was added, and publicly made on a JustGiving page set up specifically as fundraising by the Essex Employability Prize, along with any message chosen by the winner. For those who did not indicate a message, the donations were posted anonymously, with amounts shown.

All participants were sent an email to claim their prize within roughl ytwo weeks of winning the prize. Those who did not claim the prize soon were sent further reminders. Before issuing prizes, we checked all entries against the public university record and our data, to ensure they came from an eligible student email, on his or her first valid entry to the EEP site. In the second run checking was automatic, via an email white list.

Emails to participants. All emails to participants and winners, including responses to participant inquiries, came from "empprize@essex" email and, where signed, were indicated as "from the Essex Employability Prize." Emails to winners followed a standard form; we give some example cases below.

Naka winners, standard (no donation):

> *Congratulations, you have won a £20 voucher at Naka Thai in Colchester. We have passed your name to them so you can now claim your prize; please bring ID. Thanks for participating! Please encourage other eligible students to enter the Employability Prize Giveaway.*

Naka winners, donated some but not all of prize:

> *Congratulations, you have won a £20 voucher at Naka Thai in Colchester, and donated £X of this to [Charity= "Oxfam" or "The World Wildlife Foundation"]. Thank you for your donation. We have added 10% and donated £X*1.10 via our JustGiving page [link], and left your message if you gave one. We have passed your name to Naka Thai so you can now claim your net prize (£20-X voucher). Please bring ID. Thanks for participating! Please encourage other eligible students to enter the Employability Prize Giveaway.*

We checked prize winners at least once every two weeks while the EEP was active. Dinner winners were emailed that the restaurant had been notified and they can claim their prize. Amazon winners were told to check their inboxes.

People who began site but did not make it to the prize screen were emailed and directed to continue from where they left off (same treatments):

> *Hi, sorry to bother you. We noticed you began the Essex Employability Prize Giveaway but did not finish it. You need to complete this to the end to be eligible to win a prize. If you did this on your own computer or phone, just go back to this computer or device and click here [link] and you can continue where you left off. Don't worry, it shouldn't take a long time to finish!*

Treatment arms. As noted, there were two additional treatment arms, administered orthogonally to each other, and to the charity treatments. The flow of treatment assignments is depicted below.

Charity treatments (before/after and win/lose) were assigned by Qualtrics using random sampling without replacement until 8 treatments in the urn were assigned (before-win, after-win, before-lose, before-lose, before-lose, after-lose, after-lose, after-lose), and then the urn was reset. At the point of being asked to donate, the probability of winning conditional on all treatments the student observed was always 1 in 4. Note that losers were never asked to donate.

More specifically, the site:

i. Asked a student's department (course)

ii. Went to the JobsOnline treatment (50%) or "thank you for registering," balanced in every pair of observations within each department (sampled without replacement within each department)

iii. Within Prospects/No Prospects it balanced across all of the six below Information-win/Donation timing combinations, with the given probabilities (ignoring department):

Treatment combination shares:

12.5% before – win (info) [1 in 8]

25% before – lose (no – info) [2 in 8]

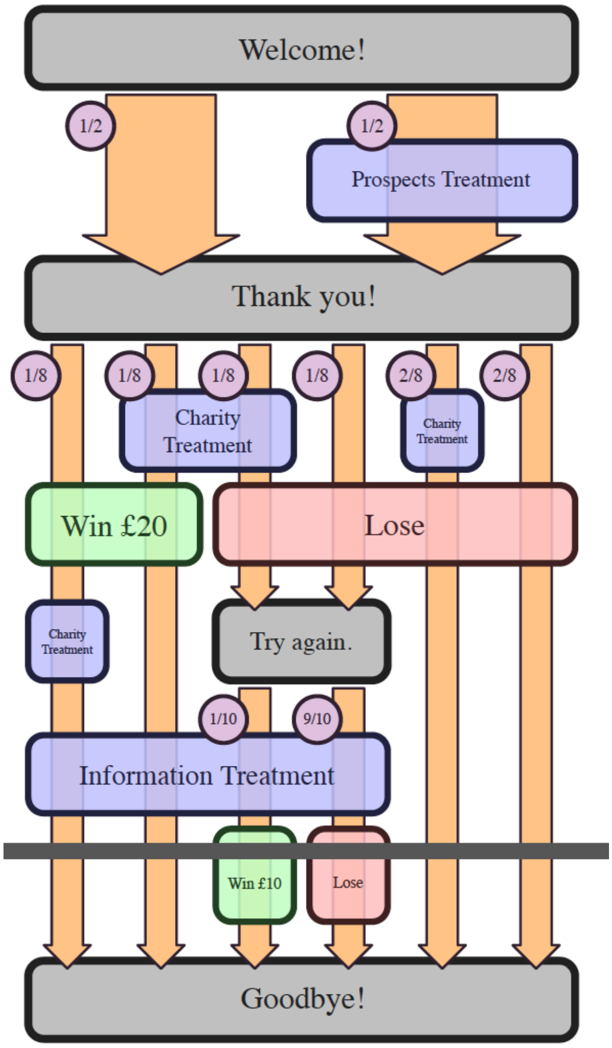


Figure E.13. Employability experiment treatments flow chart

12.5% before – lose (info, second chance) [1 in 8]

12.5% after – win (info) [1 in 8] 25% after (no ask) – lose (no – info) [2 in 8]

12.5% after (no ask) – lose (info, second chance) [1 in 8]

Overall shares:

Total prob info = 50%

Prob info|win = 100%

Prob info|lose = 33.33% (2 in 6)

Total prob win = 25%

Prob win|before = Total prob win|After/no ask =25%

Jobs Online (Prospects)

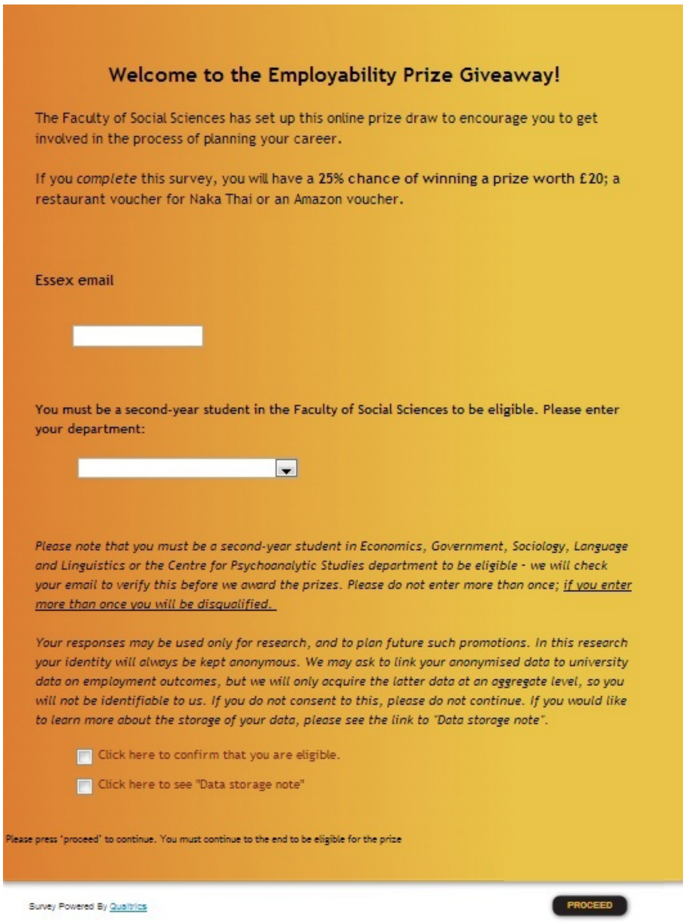
Third year students are typically encouraged by the Essex University Employability office to sign up as a JobSeeker on University of Essex Jobs Online to learn about career opportunities. Half of the students who entered the EEP web site were required to sign up for JobsOnline and enter two jobs they might be interested in, with details verified, before they were allowed to continue. Over half of those who logged on who were assigned this treatment quit at this point. To begin the EEP, students needed to give valid Essex emails; students who quit were either not allowed to enter the site again, or later on, were given the opportunity to enter again continuing from the same treatment (however, few students took this up).

Information treatment

A second experimental arm, again given to only half of participants, involved an informational intervention related to employment statistics. Students (based on their degree scheme) were asked to guess the share of a relevant peer group who were employed in a graduate level job within six months of graduation, and to guess average starting salaries. They were then told the statistics and required to enter these to continue. This treatment was given only to those who had won a prize (before they could learn how to claim it) and to others entered in a second draw with a 1 in 10 chance of winning a £10 Amazon voucher. This treatment always occured *after* the charitable ask treatments, so these could not impact the giving decisions.

Further Screen Shots. The screens below (depending on treatment) preceded the charitable ask and prize revelation screens shown in the main text. This version of the screens are from the earliest part of the first run, before the site was opened to students in other years of study and other faculties (later versions were identical except for small adaptations for these inclusions).

A copy of the experimental instrument can be tested at https://goo.gl/qSvhi1; this will cycle through each of the treatments.



Note: In later months eligibility was expanded. In the second run only the Amazon prize was used.

Figure E.13. Welcome screens

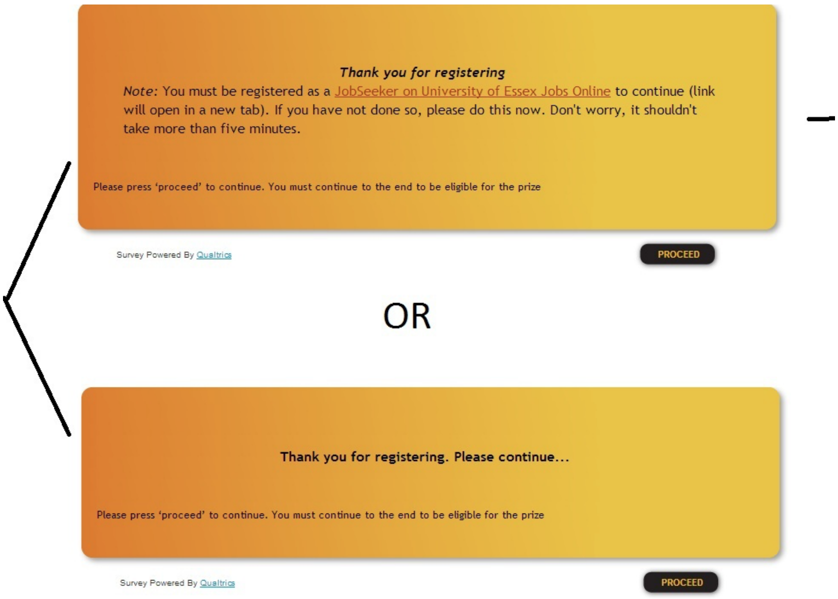
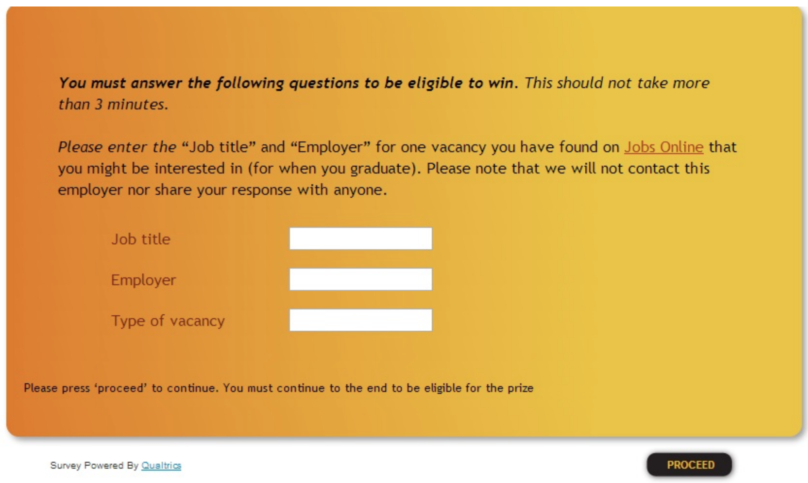
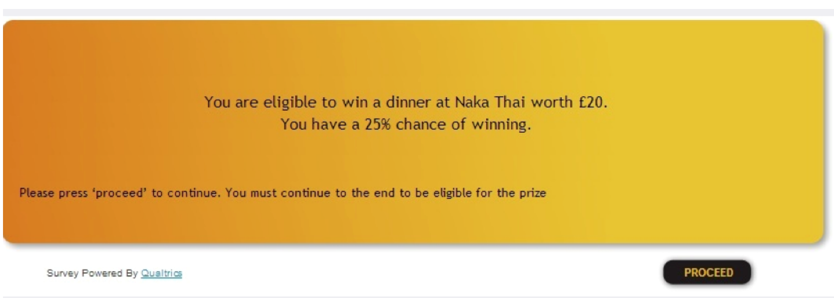


Figure E.13. Jobs Online Treatment or Continuation Screen



Note: only half of all participants were assigned this treatment. They had to complete this screen, and one additional one (about a distinct category of job), and we verified legitimate responses based on the cell "type of vacancy".

Figure E.13. JobsOnline treatment



Note: alternate screen displayed "You are eligible to win an Amazon gift certificate worth £20.".

Figure E.13. Prize Eligibility

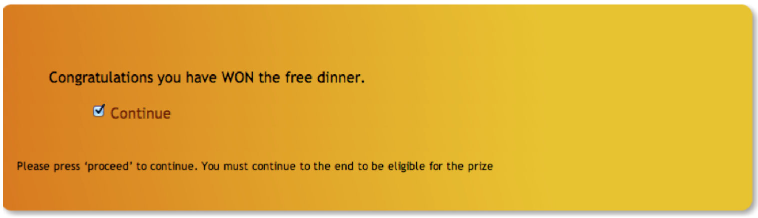


Figure E.13. Win announcement

Additional material to be provided and linked

1. Output of http://www.p-curve.com/app4/ here
2. Web-based experiment details and key screens
   - Omnibus: recruitment emails
   - Recruitment screen from Prolific
3. Lab: A complete set of relevant screenshots and translations
4. Annotated extract of pre-registration and pre-analysis plan HERE (filename info_foraearegistry_nonotes_GandPonly.pdf).