Analysis Plan (David Reinstein)

Trial title: Charitable giving trials linked to ESSExLab recruiting and omnibus: i. Crowding out (does one contribution/appeal come at the expense of another?) ii. Do people commit more before a win (conditionally on winning) or after a win? iii. Does the opportunity to donate boost reported happiness (and vice/versa)

All studies

We will perform standard parametric and nonparametric statistical analyses on full relevant samples, as well as testing for differences between subsets mentioned below. As treatments are administered randomly, and by design orthogonally to one another, we will first report statistical test without controls, particularly Fisher's exact test (for categorical outcomes and treatments) and Wilcoxon rank-sum test for continuous outcome variables.

Even under randomization, there will typically be an inexact balance in predetermined characteristics across treatments. Thus, regressions controls for predetermined observable characteristics (such as gender or income) can sometimes make estimates more efficient. Thus we report regressions with controls. We will perform model-fitting (e.g., stepwise regression) to determine the most efficient set of predetermined controls. We will run linear and Poisson-exponential regressions for charitable giving outcomes (as these are bounded variables), and also do robustness checks for other popular specifications used for bounded dependent variables, such as Tobit models. Similarly we will also report linear probability models and logit models for binary dependent variables, especially the extensive margin (donated a positive amount to a particular charity). We are not likely report so-called "conditional on positive" effects, as these are difficult to identify without an exogenous instrument for the extensive margin decision (we will only do so if such an instrument/shifter becomes apparent to us in a very obvious way after collecting the data.)

If the data suggests strong departures from the linear or exponential models, e.g., bi-modality or sensitivity to outliers, we may also report specifications more appropriate to these functional forms, and will make an argument for the suitability of these, noting this was not central to our initial plan.

Substitution

As noted above, we will first report simple uncontrolled tests for differences in treatments.

We will test for differences in the level and incidence of contributions at the *second intervention* (for the nonstudents this is via the Omnibus; for the Students, this is via the employability survey), between:

1. 'No ask' vs 'some previous ask'
2. 'Previous ask for similar charity' vs 'Previous ask for distinct charity'
3. 'Previous ask for similar charity in poverty domain' vs 'Previous ask for similar charity in health domain'
4. Recent previous ask (based on time email was sent to participant) versus less recent ask (we will fit a model of 'impact of delay on later contribution')

For 2-4, we will also test for differences in differences, e.g., we will test whether those who are asked to donate to a similar charity increase or decrease their donations (between the 2 time intervals) more than those asked to donate to a distinct charity.

We will also attempt to measure whether the impact of the first ask on later donations is *mediated* by the donation response to the first ask (e.g., those who *donate* in the first ask may donate less in the second ask, while those who do *not* donate in the first ask may donate more in the second ask) . However, mediation analysis is difficult, and we will state our results cautiously, following the guidance and techniques suggested in Heckman and Pinto (2014).

Differentiation of estimated effects (heterogeneity, interactions)

Previous work (Reinstein, 2011; Karlan and Wood 2017, indirectly) suggests a greater degree of substitution (crowding-out) among those who are large and regular givers. We do not have direct measures of this, but we will attempt to differentiate by likely correlates. (We may have to collect these in a follow-up study.)

Donors who have a personal optimization strategy and target, rather than donors who respond to emotional cues and powerful appeals, are also arguably more likely to exhibit substitution. Because of this, we will differentiate our estimates by personality attributes (measure in the omnibus) that previous literature find are associated with analytical versus emotional decision-making.

We will estimate all of the above pooling Students and Non-students as our primary object of interest; we will estimate these separately as a secondary descriptive result.

Because giving behavior has been found in many cases to differ by gender and by religious background, we will also bifurcate our estimates by these categories (gender, indicated religious affiliation vs. agnostic/atheist). We have no ex-ante hypothesis for differences in the substitution effects between these groups.

Direction of hypothesized effects

We believe it is most likely that we will find crowding out, and a greater crowding out for more similar charities, and from more recent asks. This is suggested by previous results (Reinstein; Donkers; Ek) and by standard predictions from economic theory, considering diminishing returns to the individual's benefit from each cause donated to, and considering the benefit as a perishable good.

This suggest negative coefficients on 'some previous ask' (1 above), on previous ask for similar charity' (2) and on 'more recent previous ask' (4). We have no prediction on whether health or poverty charities will be more in competition with one another (3).

However, there are also reasonable arguments that 'charitable giving is a learned behavior,' and that donations may have properties of experience goods, and learning may be positive on average. Furthermore, the effect (e.g., a feeling of social-pressure or a donation reference point) of an initial ask may carry over to a later ask. This may be particularly true among those who do not respond to the first ask (but differentiating this will be difficult because of selection issues associated with such mediation analysis). Thus, it is possible that being asked at an earlier date may increase giving at a later date. We will conduct two-sided hypothesis tests in all cases.

Giving and Probability

Our analysis will largely follow our earlier working paper (Kellner et al, 2016); see especially "Table 2: Donations/commitments by experiment: Before versus After". As noted above, we will also present controlled regressions.

We will test for differences in the level and incidence of contributions between:

1. (1/2 chance context): 'Before' Donation commitment when asked before learning if won vs. 'After' Donation made when asked after winning (same context)
2. (Ambiguous chance context): 'Before-50-a' Commitment when asked to donate if win (1 of 6) £50 prize vs. 'Before-10-a' when asked to donate if win (1 of 30) £10 prizes
3. (Ambiguous chance context): Before-50-a vs. 'Before-10-50-a' commitment from £50
4. (Ambiguous chance context): Before-10-a vs. 'Before-10-50-a' commitment from £10
5. (Ambiguous chance context): *Expected value* of donation from Before-50-a vs. from Before-10-50-a
6. (Ambiguous chance context): *Expected value* of donation from Before-10-a vs. from Before-10-50-a
7. (Both contexts): Donation from £10 only in ambiguous context vs. in 1/2 context
8. (1/2 chance context and certain £10 prizes): Before vs. Certain
9. (1/2 chance context and certain £10 prizes): Before-10-a vs. Certain
10. (1/2 chance context and certain £10 prizes): After vs. Certain

For 1, we will estimate this separately for this experiment, as well as pooling with our previous field experiments, and with all our previous experiments (comparing 'before' and 'after win' treatments only).

For 7 and 9, identification will be less clean. These groups may systematically differ in composition, as the ambiguous treatments are presented to later-responders; we will attempt to test the similarity along observable lines and control for differences in these populations.

For 8-10 the comparisons are also difficult, as we are asking for donations to either of 2 charities in the Before and After case (to maximize donations relative to variance) but only one charity in the Certain cases. Applying standard consumer theory would imply (weakly) more donations where more choices are presented. This comparison will only be easily interpreted if participants almost always favor one of the charities, suggesting we can ignore the other opportunity, and if we assume no framing effect from the inclusion of the rarely-chosen charity. However, under these assumptions, if we find more donations from the treatment with *fewer* options (e.g., from Before relative to Certain), this would suggest a lower bound, i.e., the true effect of (e.g.) the Before treatment is

For 8-10, we will estimate this separately for this experiment, as well as pooling with our previous lab experiments to compare After and Certain and to compare Before and Certain treatments. However, as the contexts differ, the pooled analysis will be considered secondary.

We will run controlled regressions following the control strategies and functional form considerations outlined above.

(As a secondary test, we will measure which of the above treatments had a greater impact on reported happiness.)

Differentiation of estimated effects (heterogeneity, interactions)

Because giving behavior has been found in many cases to differ by gender and by religious background, we will also bifurcate our estimates by these categories (gender, indicated religious affiliation vs. agnostic/atheist).

Some of our previous field evidence has found that male donations respond more positively to the 'Before' environment (relative to the after environment) than do females' donation. However, this result has not persisted strongly in all trials.

There is also a reasonable argument that religious people will respond more positively to the Before environment, believing that a positive donation will be divinely rewarded with a winning outcome. On the other hand, we might imagine religious people will donate in response to a win, to demonstrate their gratitude.

We will differentiate our estimates by stated risk-aversion; a 'Before' commitment increases the variance of the donation, while reducing the variance of the residual win.

We will conduct two-sided hypothesis tests in all cases.

Direction of hypothesized effects

As noted in our paper, there are several theoretical reasons to believe that people will be more generous (or at least commit to donate larger *conditional* amounts) in making commitments from uncertain income than when making commitments after a known win. Our previous evidence points in this direction. However, there are also theoretical justifications for the opposite effect. We will conduct two-sided hypothesis tests in all cases.

Happiness

We will test whether:

- being asked about happiness increases or decreases one's donation incidence/amount

  - reporting greater happiness is correlated to subsequent donation incidence/amount (non-causal interpretation)

- being asked to donate to charity leads to greater subsequent reported happiness

- being asked to donate to charity leads to greater reported happiness in a subsequent phase (i.e., much later)

- Mediation: this may depend on *whether* one donates; it may increase for donors and decrease for non-donors (tentative hypothesis)

As we observe happiness both before and after donation asks and choices (depending on treatment), we can use a difference-in-difference strategy. Essentially this will allow us to control for a correlation between propensity to donate and happiness, and consider the aforementioned mediation question.g

Proposed starting salary

We will test whether there is a systematic gender difference in the starting salaries requested in the hypothetical scenario. In this context we cannot guarantee that male and female respondents are coming from the same distribution of attributes, nor from a representative sample. This research is exploratory and serves as a pilot for future research. We will compare the gender differences in requested starting salaries, controlling for set of observable characteristics of the respondents deemed likely (consulting previous literature) to affect this outcome variable.