I have now read the proposal for the replication and extension of Baron & Szymanska (2011). My view is that these are worthwhile studies to replicate since they test some essential claims in the literature on charitable giving. (I have a strong prior that most of these studies will successfully replicate, though I have been wrong in the past about this). I have a few main comments:

1. I question the usefulness of the overhead funding extension. First, my understanding is that the manipulation the authors provide suggest that it is actually more cost effective to donate to the charity with the overhead that has been already paid. That is, the donor can save $5 to donate when the overhead has already been paid. This, from a utilitarian perspective, is more efficient, and therefore presents a confound with overhead. The authors could change it such that for B, a contribution of $100 buys the package—and another donor has already covered the $5 overhead costs—but the problem here is that B now has a waste confound which is already covered previously in the proposal. There may be a way around these confounds that the authors may want to think about. Second, my understanding of the initial results is that they found an overhead effect in a between-subjects design and this is a within-subjects design. As a result, this makes the overhead differences more salient which likely will cause individuals to alter how they behave relative to the real world. Although such an investigation is informative, it essentially is testing a boundary condition of the previously found effect.
2. Regarding sample size. I would suggest the authors consider adopting Simonsohn (2015) as a guide to determine a sample size for replication, which suggests simply multiplying the initial sample size by 2.5. This way, if the authors find a null effect they can say that the replication indicates that the initial studies did not have sufficient power to detect an effect. If the authors’ current proposal is above that number, then I suggest they note that their sample size is above the number set out by Simonsohn (2015)
3. Related to the above, one big caveat of all these studies is that they are all within-subjects designs. People may therefore be reacting in a way they wouldn’t do so if this were between-subjects. This ought to be mentioned as a main factor affecting the interpretation of these results.
4. Unlike the other studies, Hypotheses 3 (Diversification effect) uses an allocation of 0 as a standard, such that to act in accordance with utilitarianism, individuals should give all their money to one rather than more than one charity. However, it seems that even a few inattentive participants who just indicate randomly could show such an effect. This seems potentially problematic with the original study and problematic here too. Either a different test or a control condition would be appropriate to include.

Additional comments and references:

Page 11. There is a typo involving the apostrophe in Qualtrics

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological science*, *26*(5), 559-569.