

Main comments:

- I am not sure about one sentence in hypothesis 4 in Table 1. You say first that « expecting emotional rewards serve as a positive signal of moral character but expecting reputational and/or material rewards serve as a negative signal of moral character. » For me, there are three groups here: (i) people who don't expect any reward, (ii) people who expect an emotional reward, and (iii) people who expect a material reward. How do these groups sort? In the Design table, you say that people who expect to receive a material reward are perceived as less moral « yet less so when they expected to receive emotional benefits ». So, people from (ii) are perceived as less or more moral than people from (i)? (I understand that people from (iii) are the lowest, but I am not sure about your prediction about group (ii) relative to group (i).)
- I do not understand why there are 4 hypotheses in Table 1 and only 2 in the Design Table.
- I am a bit puzzled by the pre-test participants. Of course, I completely understand that it is important to test the design, the code, the duration, etc. However, we can agree that this generates some degree of freedom for the researchers that is not observable by the reviewers. I think that one of the main ideas of RRs is that it is always (or at least very often) easy to find ex-post a rationale to justify a decision (like excluding some observations). The objective of RRs is to explicitly say which rule we are going to use once we start data collection. So, in my view, I would either recommend to (a) drop these 30 observations whatever happens, (b) to test the survey, duration, etc before the IPA, or (c) to explicitly say which rule you will apply to decide whether you want to keep these observations (« case of serious technical issues » is, in my view, too vague).
- I am a bit surprised by the gender distribution in the US Prolific population. More than half of your participants report « other/did not disclose ». When I check on Prolific (US), there are 17,352 men (cis+trans), 27,802 women (cis+trans), 1,416 non-binary, and 271 « rather not say ». So, in prolific, there are less than 4% of Non-binary/rather not say, but you expect more than half of your sample. Why is there such a difference?
- I am also puzzled by the treatment without donation (Study 3). I understand the original concern that we could not say much about a person's behavior when we do not know what she did. On the Qualtrics survey, in the treatment where donation is not said, there are two questions that I find intriguing:

-To what extent is the person's behavior reflective of their intentions?

-To what extent did the person donate to make him/herself feel better? [The person did not donate, no?]

Regarding the first item, I do not have a strong opinion but I feel that « behavior » might refer to different things in different treatments (i.e., the physical response in case participants don't know the donation, and the donation in case they know it).

Suggestions:

- I would personally support the exclusion of the participants who fail the attention check because when they are forced to respond, they might then only look closely at the question which prevents them from going further, without paying attention to the rest of the questionnaire. In my view, accepting all participants could lead to an increase in the noise in the data and, thus, decrease the probability of detecting an effect (statistical power). I have no strong opinion here but I just wanted to mention this point.

- I would personally like to see on page 26 a footnote saying the minimally acceptable Cronbach's alpha for you to run the study. It is, in my view, important to set up criteria for the internal validity of the estimators.

- I have tried the Qualtrics survey. It seems to work well (I've been randomized to several treatment variations and the order of the two studies was also well randomized). I would only suggest integrating the Prolific ID in the URL such that participants do not have to type it.

Misc:

Page 11: Typo: « Studies 3 Study 6 »

Page 12: Number 4 in the hypothesis number in Table 1 missing.

Page 19, Table 3: Are we sure that the age in Prolific starts at 0?