

**Referee report on the manuscript entitled “Reference points and decision-making:
Impact of status quo, defaults, and past behavior in a conceptual replication and
extensions Registered Report of Dinner et al. (2011)” submitted as a stage one
Registered Report at PCI-RR
Sylvain Chabé-Ferret**

In this stage one registered report, the authors conduct a conceptual replication of the famous Dinner et al. (2011) study of default effects. In the original study, the authors asked a sample extracted from a panel of households in the US to answer a hypothetical question: whether to stay with the default installed lightbulbs after a remodeling at their home or to switch to a different set. The randomized manipulation was the default set of lightbulbs, which came with a different price tag: either incandescent lightbulbs (INC) priced at 0.5 USD per lightbulb for a total cost of 9 USD or Compact Fluorescent Lightbulbs (CFL) priced at 3 USD per lightbulb for a total cost of 54 USD. The scenario included a thorough description of the characteristics of the light bulbs (type of light emitted, reaction to being switched off and on repeatedly, ease of use and disposability, cost in electricity and expected lifetime). The scenario also made clear that there were no extra costs (such as labor costs) at switching from the default lightbulbs, apart from their different list prices. The original paper also tried to test for the reasons that might explain the existence of a default effect, focusing on reference dependence as explained by Query Theory. Reference dependence refers to the theory that agents will examine differently the default option that they are presented with and the alternative. Example of reference dependence includes endowment effects, where agents express preference for what they already possess. Query Theory theorizes that endowment effects emerge because agents reason differently about objects they possess versus objects they do not possess. For example, agents owning an object will think first about reasons to keep it before thinking about reasons to sell it, and, as a consequence, will come up with more reasons for keeping it than selling it, therefore impacting the likelihood of selling the object. The original paper contains three experiments intended to test for the ability of Query Theory to explain default effects. In experiment 1, the agents record their thoughts (aspects of the choice) while making their decision. In experiment 2, the agents record their thoughts after making their decision (in order to measure decision time accurately). In experiment 3, the authors use a 2×2 design manipulating the default and the listing order, listing aspects in favor of the default before aspects in favor of the alternative first or second. The original authors also collected data to test two other potential explanation of default (or endowment) effects: costs of decision and implied endorsement. Cost of decision refers to the fact that agents stick to defaults because of the costs of thinking through the alternative (or of going through with the alternative). Implied endorsement refers to the fact that agents might infer that the default option has been chosen by the contractor because it is their preferred option. The original authors measured cost of decision and implied endorsement using self-reporting questions. All original experiments find evidence of large default effects, with the proportion choosing INC doubling from 20% to 40% when it is made the default option. The authors also find supportive evidence in favor of Query Theory. Indeed, experiment 3 shows that, when agents are made to think first about the qualities of the non-default option, they exhibit no statistically significant default effect.

The authors of the stage 1 registered report first discuss the interpretation of the original study. They contend that the original study was not about default effects but about status quo effects. Indeed, in a related paper (Feldman et al., 2020), the authors made the distinction between default effects (the tendency to choose the default option in a choice set) and status quo effect (the tendency to stick with the already implemented option). The authors thus propose a conceptual replication of Dinner et al. (2011). They first propose to test the existence of a status quo effect (IV1), using the pre-setting of INC vs CFL lightbulbs as in the original experiment. This is the direct replication part of experiments 1 and 2 in the original study. The authors also plan to directly elicit participants’ opinion about why they made their choice in order to measure some of the possible explanations of the status quo effect (perceived effort and direct and indirect implied endorsement). In extensions, the authors plan to test for default effects and for the effect of past behavior. To test for default effects (IV2), the authors will randomly manipulate the default option on the choice set (INC, CFL or no default). To test for the effect of past behavior (IV3), the authors will randomize what they tell the agents about what they did in a previous similar situation: either that they typically chose to install INC light bulbs, or they

will say nothing to the agents about their past behavior. The overall design will be a $2 \times 3 \times 2$ design

This is an important replication since defaults are among the strongest of effects in choice architecture and nudges, and the original Dinner et al. (2011) paper has never been replicated. Here are my main comments on the stage one registered report:

1. At first, I was not convinced that the distinction between default effects and status quo effects was a very important one. It seemed mostly a semantic distinction to me. I now see that if the evidence for default effects comes from mislabelled status quo effects, but leads to the implementation of nudges based on default effects, and status quo effects are much larger than default effects, the actual impact of the nudges in the field might be disappointing. I think the authors should discuss that issue when they talk about the relevance of their study. That will make the distinction between status quo and default effects clearer for the reader, and of more interest. Indeed, we now see that if default effects are much smaller than status quo effects in the replication, we expect default nudges to be disappointing. Since the distinction between default and status quo effects is subtle, I think the authors could also discuss how they apply in actual nudges (default subscription to the list of organ donors, default subscription of electricity tariff). That would enlighten the reader as to the practical importance of the distinction.
2. The proposed replication does not test the main theoretical mechanism put forward to explain status quo effects in the original study: Query Theory. The proposed replication also does not replicate experiment 3 in the original paper, which is the main test of the Query Theory explanation for default effects. The authors also do not plan to collect data that were collected in the original experiment about the stream of thoughts that agents went through when making their decisions and that was the basis of the mediation analysis in the original study. Instead, the replication authors extend the original paper in two directions: the existence of default effects (as default in the choice set) (IV2) and of past behavior effects (IV3). I will talk about the way the authors of the replication plan to test for the effect of past behavior later. For now, I wonder whether it makes sense to exclude fully tests of Query Theory from the replication. All in all, the result of experiment 3, that changing the ordering of thoughts about the options nullifies the status quo effect is the nicest and probably scientifically the most important effect in the original study. It actually would be very interesting to test whether such an intervention nullifies default and/or status quo effects.
3. I do not understand the need for the extension on the effects of past behavior (IV3). It costs in terms of power for the interaction effects (and might alter the direct effects in the other two conditions). And I do not see how it is justified here. I agree that the agents' past choice might influence the current choice, but the reason why it does is unclear, and the reason why it is related to default effects is also unclear. Moreover, I find the way to test for this effect to be particularly weak, since the manipulation simply consists in telling the subjects that they made such and such a choice in a previous similar situation. How the agents will interpret that manipulation seems highly uncertain to me. My advice would be to dispense with this condition and to replace it with the manipulation in experiment 3 of the original study expressly telling agents to go through the list of good sides and bad sides of each option, starting either with the default one, or with the other one, or with the status quo one, or with the other one. In that case, the replication authors will be able to test directly for whether the manipulation cancels the default and/or status quo effects and will be an almost full replication of the original study. Experiment 3 is the most convincing evidence in favor of Query Theory in the original study. I understand that the authors do not want to spend too much research effort at collecting agents' thoughts as in experiments 1 and 2 in the original paper, and to conduct a mediation analysis. But at least replicating experiment 3 would be a very useful addition to the current experiment.
4. I found it hard to understand exactly the types of vignettes the agents will see during the experiment based on the stage 1 RR alone. I've had to go back to the vignette in the original study to make it clear. I suggest the authors of the replication add a vignette similar to the one in Appendix A in the original study.

Let me now list my answers to the more precise questions asked for a PCI RR of stage 1:

- (a) *Does the research question make sense in light of the theory or applications? Is it clearly defined? Where the proposal includes hypotheses, are the hypotheses capable of answering the research question? Yes, except for IV3, which I find less convincing and much harder to interpret.*
- (b) *Is the protocol sufficiently detailed to enable replication by an expert in the field, and to close off sources of undisclosed procedural or analytic flexibility? Yes.*
- (c) *Is there an exact mapping between the theory, hypotheses, sampling plan (e.g. power analysis, where applicable), preregistered statistical tests, and possible interpretations given different outcomes? Yes.*
- (d) *For proposals that test hypotheses, have the authors explained precisely which outcomes will confirm or disconfirm their predictions? Yes.*
- (e) *Is the sample size sufficient to provide informative results? Yes.*
- (f) *Where the proposal involves statistical hypothesis testing, does the sampling plan for each hypothesis propose a realistic and well justified estimate of the effect size? Yes.*
- (g) *Have the authors avoided the common pitfall of relying on conventional null hypothesis significance testing to conclude evidence of absence from null results? Yes.*
- (h) *Where the authors intend to interpret a negative result as evidence that an effect is absent, have authors proposed an inferential method that is capable of drawing such a conclusion, such as Bayesian hypothesis testing or frequentist equivalence testing? No.*
- (i) *Have the authors minimised all discussion of post hoc exploratory analyses, apart from those that must be explained to justify specific design features? Yes.*
- (j) *Have the authors clearly distinguished work that has already been done (e.g. preliminary studies and data analyses) from work yet to be done? Yes.*
- (k) *Have the authors prespecified positive controls, manipulation checks or other data quality checks? If not, have they justified why such tests are either infeasible or unnecessary? Is the design sufficiently well controlled in all other respects? Yes.*
- (l) *When proposing positive controls or other data quality checks that rely on inferential testing, have the authors included a statistical sampling plan that is sufficient in terms of statistical power or evidential strength? Yes.*
- (m) *Does the proposed research fall within established ethical norms for its field? Regardless of whether the study has received ethical approval, have the authors adequately considered any ethical risks of the research? Yes.*