Reply to PCIRR decision letter #393:
Dinner et al. (2011) replication and extension

[Important note: We are very grateful for the immensely constructive and positive feedback from the reviewers, which has helped us catch and address oversights. Yet, that also meant our revision round took longer than we expected.

We therefore feel it necessary to note that this submission is part of a MSc thesis project with the thesis submission date currently set to the end of June. We therefore note that we will have to proceed to pre-registration and data collection by June 21st the latest, regardless of whether we receive an in-principle acceptance from PCIRR or not, in order to ensure timely thesis submission. We do hope to be able to proceed to data collection with the community's endorsement, however we want to align expectations in case that is not possible.

In case we do not receive the community's in-principle acceptance in time and we proceed to pre-register and collect the data, then based on our previous discussion and correspondence with the recommender, this would mean an adjustment of the PCIRR control level towards "RRs involving existing data" from Level 6 to a lower level (per "Guide for authors" on the PCIRR website).]

We would like to thank the editor and the reviewers for their useful suggestions and below we provide a detailed response as well as a tally of all the changes that were made in the manuscript. For an easier overview of all the changes made, we also provide a summary of changes.

Please note that the editor's and reviewers' comments are in bold with our reply underneath in normal script.

A track-changes comparison of the previous submission and the revised submission can be found on: https://draftable.com/compare/tFYwnKlxQcDg

A track-changes manuscript is provided with the file: PCIRR-RNR-Dinner-etal-2011-eplication-extension-mainmanuscript-trackchanges.docx (https://osf.io/xtpd5)

Summary of changes

Below we provide a table with a summary of the main changes to the manuscript and our response to the editor and reviewers:

Section	Actions taken in the current manuscript
General	Removed the past behavior manipulation, added a control condition for status quo manipulation, changed the INC bulbs vs CFL bulbs comparison to CFL bulbs and LED bulbs comparison, improved writing quality, added clarity to definition.
Introduction	Removed background related to past behavior manipulation, included more on nudges and status quo, added implications in differentiating the defaults and status quo.
Methods	Removed the past behavior manipulation; Added vignettes for all manipulations. Increased the clarity in 1) default manipulation, 2) deviation on comprehension checkings, 3) word choices in the status quo scenario paragraph. Minor correction on word choices; added a control condition for the status quo manipulation. Added reporting of sociodemographic descriptives. Updated data analysis strategy (e.g., address non-parametric tests.)
Discussion	Added many planned discussions (e.g., potential influence of habits, past behaviors and preference on reference points)

Reply to Editor: Prof. Chris Chambers

I have now received three very detailed and constructive reviews of your Stage 1 submission. Broadly, the reviews are very encouraging and I believe the submission is a promising candidate for eventual Stage 1 in-principle acceptance. Major points to address include clarification of theoretical concepts and study rationale, increased details concerning the contingent analysis plans, and doubts about the validity and fidelity of the past behavior manipulation (IV3; this concern was flagged by two of the reviewers). The evaluations offer many helpful suggestions for resolving these concerns, including some potentially substantial design changes. I look forward to seeing your response to the reviews and receiving your revised manuscript in due course.

Thank you for the reviews obtained, your feedback, and the invitation to revise and resubmit.

To us, this peer review is an exemplary case of how important peer review on a pre-registration is and the potential in anticipating issues in advance to help authors tackle important oversights and improve.

Given the comments, we made the following adjustments: We decided to remove the past behavior manipulation, addressing Reviewers 1 and 3.

Another extremely valuable comment by Reviewer 2 helped us rethink the target's status quo manipulation and realize the value in implementing a neutral control condition that was not included in the original design.

Also, a side-comment about the lightbulb regulations in Europe has led us to look deeper into the situation for our target sample in the US and a realization that contrasting the INC and CFL light bulbs is no longer relevant. We therefore had to adapt our manipulations to contrasting CFL and LED bulbs instead. We also needed to add planned discussions about the implications of adapting replication materials to a changing political and environmental landscape.

Therefore, combining the three very important insights, our final design shifted from a 3 (default: INC/CFL/none) x 2 (status quo: INC/CFL) x 2 (past behavior: INC/none) to a 3 (status quo: CFL/LED/none) by 3 (default: CFL/LED/none) design.

Furthermore, many of the comments helped us improve on many aspects of both the survey implementation and our planned analyses. We are very grateful.

Reply to Reviewer #1: Dr./Prof. Laurens van Gestel

> **This Stage 1 manuscript proposes very valuable work and an important replication of impactful research. I believe that the current proposal largely meets the criteria that are important for Stage 1 acceptance, such as having a solid research question in light of theory or applications and having a sufficiently detailed protocol that enables replication. Moreover the sample size seems adequate, the proposed analyses are fit to test the hypotheses, and the questionnaire itself contains manipulation and quality checks.**

Thank you for the positive encouraging opening note.

> **Yet, I have some concerns regarding the operationalization of the default and other methodological decisions.**
>
> **1) The current study makes a conceptual distinction between status quo and defaults. Although this distinction makes sense conceptually, the operationalization of the default seems odd. In the study, status quo is manipulated by manipulating in the scenario whether the new fixtures have been outfitted with INC light bulbs of SFL light bulbs. Defaults are subsequently manipulated by placing the reference point of the slider either fully in favor of INC-bulbs or CFL-bulbs, or in the middle. However, preference is not identical to a choice or decision, so it is odd to place a default in that slider for measuring preferences. What is the rationale for this, and wouldn't it be better to place the default in the choice itself by for example preselecting one of the two options for the question "In this situation, what will you do?"?**
> **In Table 1, Hypothesis 3 is expressed as 'People are more likely to choose the default option', but in the current proposed study, it seems that this is not what is being measured. Rather, it seems that the effect of a default on expressed preferences is measured. I thus believe the conceptual distinction that the authors make between status quo and defaults deserves further refinement in the operationalization of the study.**

Thank you. This is valuable feedback.

We did indeed implement the default choice manipulation, and therefore believe that there was a misunderstanding regarding what we implemented. Your feedback has helped us realize that we should improve on the way we communicated our design. We tried to understand why this was not clear in our previous submission and then came to an important realization that the defaults might not have been clear enough with the chosen style of the Qualtrics survey in our previous

submission in which the whole question was highlighted in blue rather than having a clear circle checkbox next to each option. We therefore, as a first step, switched to a clearer Qualtrics survey style that explicitly displays a circle next to each option that allows a more intuitive interface in which the default is displayed as a circle with a dot in it. This is very valuable and important feedback and so we are very grateful.

The default was indeed also manipulated in the extension preference slider, yet - more importantly - it was first manipulated in the dichotomous choice between the two types of bulbs. The slider manipulation was just meant to match that in our extensions preference question, to align the default in the dichotomous choice with the slider choice.

To make things clearer, in our revision we now included visual examples of the two types of questions (for both dichotomous choice and preference slider) under the Method section in "Procedure" and "Manipulation" in the main manuscript.

Below is a sample screenshot that shows the dual manipulation of defaults for each scenario, first is the replication in providing the default clearly in the provided choice-set and below is the extension manipulating the default in the preference slider question:

We added this screenshot example to our manipulation.

You are also welcome to preview the Qualtrics with the new style template to assess it directly: https://hku.au1.qualtrics.com/jfe/preview/previewId/27899a48-ba75-4ee0-af66-5a2d3970d017/SV_bPBZkp8knWeyfQi?Q_CHL=preview&Q_SurveyVersionID=current

> **2)   Moreover, why do the authors decide to manipulate past behavior. How trustworthy do the authors believe that this manipulation will be? Wouldn't it be better to measure past behavior by letting participants indicate this themselves? Or if not, would it be possible to check for trustworthiness of this manipulation?**

This is very valuable feedback, and we appreciate that very much. It is sometimes the case that we aim to try and do too much, so feedback about simplifying things and focusing on the core is terrific.

Following on your and others' feedback, we decided to remove the 3rd independent variable of a past behavior manipulation in our experimental design. We added planned discussion regarding the reference points and past behavior as one potential reference point in the Discussion section for the Stage 2 manuscript.

> **Minor:**

> **1)   The introduction currently contains some repetition.**

Thank you. It is sometimes difficult for us when writing the introduction to notice such repetition and so would have appreciated specific feedback. We tried our best to review our introduction again and optimize it to remove repetition.

> **2)   The part about default effects seems unnecessarily critical, perhaps to make the point of distinguishing between default effects and status quo. However, the findings about organ donation defaults are clear regarding consent rates. Mixed results arise when speaking about actual organ donations, which is more of a downstream effect, but the default as such can be effective in stimulating consent rates.**

We are unsure what seemed critical to you in what we wrote, and why you thought that what we wrote was unnecessary. In the last three years several books came out regarding nudge and default effects by some of the most prominent figures in the field with them noting that the findings are not clear and often misinterpreted. This is an example from the recent "Nudge The Final Edition" by Thaler and Sunstein (2021) chapter 13:

> "But, surprise, that is not the conclusion we reached. After doing research on the topic, we advocated a different policy, one we call prompted choice. To our dismay, in the years

since, several countries, including Wales, England, and Germany, have considered or made the switch to presumed consent. After such laws are passed, we often get notes of congratulations via Twitter! It is frustrating that on the single topic on which the selection of default rules has gotten the most attention, countries are adopting what we think is the wrong system. Cue the sound of our teeth grinding. [...] To preview what is to come, we think that there has been some confusion about the desired goals of this policy. The goal is not just to choose a system that gets the bars on that chart to be as high as possible. A central goal is to save lives, by making more organs available for use. But that is not the only goal; it is also important to consider potentially competing interests, preferences, and rights. [...] We discuss this issue in detail not only because it is important in itself, but because it offers some general lessons on the way default options work. [...] We favor the policy of prompted choice because there is no evidence that a viable alternative system would save more lives (and hence is superior in terms of the interests of Patients), and because we think it does the best job of respecting the rights and interests of Potential Donors and Families."

The very people who are perceived to support and endorse default effects and opt-out are the ones arguing that they have been misunderstood, and are the ones critical of the way people think of and implement default effects. We cited a talk by Thaler which presents the following figure:

Presumed Consent *Reduces* Organs Transplanted
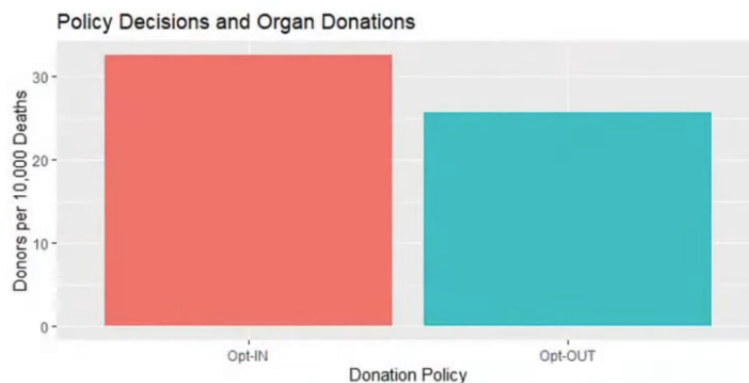
in which Thaler notes the following:

"This is the real graph, it's not famous, but this tells us the right answer".

Contrast that graph with the graph that features prominently in the earlier editions of Nudge as well as by the recent books by Johnson (2021) "The elements of choice: Why the way we decide matters" that came out on the same year (Chapter 5 "decisions by default").

We tried to be very humble and careful in the way we articulated things. We think it is important that audiences reading about defaults and nudges understand the complexity of the default effect situation and the need to get data on downstream effects that are opposite to the common narrative and go counter to the intended effect.

We would gladly improve further and revise if given clear editorial guidance.

> **Similarly, as the authors point out themselves, defaults are among the most effective types of nudges (e.g., Jachimowicz, 2019; Hummel and Maedche, 2019). The Mertens et al. (2022) meta-analysis that has led to controversy is a meta-analysis of all types of nudges across all behavioral domains, so this criticism cannot directly be applied to defaults specifically per se. This section deserves a bit more nuance.**

We would have appreciated more clarifications and guidance on what nuances you are expecting and think are relevant. We have tried to be very cautious and humble in how we described this literature.

We used the meta-analyses covering nudge as an indication that traditionally people think of defaults as one of the strongest among the nudge interventions, and have clearly indicated that the entire nudge literature seems to be affected by publication bias, and ended up with the broad argument that:

> "The debate about the robustness of nudge and default effects is still ongoing, showing the need for careful follow-up empirical work (Mertens et al., 2022b), and addressing publication bias through tools like Registered Reports."

We believe that this is a fair and uncritical reflection of the current state of affairs in the default effect literature.

We would gladly improve further and revise if given clear editorial guidance.

> **3)   Regarding past behavior, I would suggest the authors refer to literature on habits**

Yes, thank you, that is a valuable suggestion.

Given feedback from the other reviewers, we decided to remove the past behavior manipulation altogether and will briefly mention this direction to the planned Discussion section in the Stage 2.

> **4)   Regarding preferences, it may also be interesting to look at how preferences impact status quo effects. See for example, de Ridder et al., 2022 on Nudgeability and work by Venema et al.**

Thank you, this is an interesting direction for future research. In our experimental design, we use the term preferences to reflect one's indicated outcome regarding the specific choice regarding the provided scenario, rather than stable individual differences that would affect choice. Yet, we think it is indeed interesting to see the interaction of both choice and preferences with traits and inclinations. We therefore added this to the planned discussion to be completed in Stage 2.

> **5)   I applaud the authors for their scientific rigor. Checking comprehension seems very important to do. Yet, this makes the manipulation very explicit and perhaps a little unnatural as well, and this may impact results (e.g., demand effects). As this is a deviation from the original article, this is something that deserves attention from the authors.**

We included the comprehension to reduce misunderstanding to the implemented scenario.

We noted this as a deviation under the Method section in "Procedure" section, and then again discussed in the "Comprehension checks" subsection. It then is also mentioned in Table 5 (Comparison of target article versus replication) under "Procedure".

To further address this point, we also added this as a planned discussion in Stage 2 in the Discussion section under "Adjustments to the target article: Comprehension checks":

> "[Planned discussion about limitations and implications of making adjustments to the original design such as adding comprehension checks, balancing the need to ensure attentiveness and comprehension with making things explicit with the risk of demand effects]"

> **6)   The scenario now reads "Tomorrow evening the head contractor comes by your home to discuss the last aspects of the addition…" I believe it would be more comprehensible if 'tomorrow evening' is replaced by 'The next evening…'.**

Yes, that is a good point. The repeating phrase "Tomorrow evening…" was adopted from the scenario text used in the target article originally for replication consistency. We completely agree with you that this sounds awkward and might confuse readers.

We therefore replaced the phrase "Tomorrow evening" to "The following evening…" , and the second "tomorrow" referring to installing to "the next day". So, as an example, one of the scenarios now reads (we highlighted the changed text):

> Imagine that you are undergoing a significant amount of remodeling on your home.
>
> On the last day of work the contractors clean up all leftover dust, dirt, and paint. Before leaving, one of the workers tells you that the head contractor will be back tomorrow for a final inspection of the house.
>
>
> **The following evening** the head contractor comes by your home to discuss the last aspects of the addition. After showing you one of the newly installed light fixtures he mentions that all 18 bulbs in the new fixtures have been outfitted with Light Emitting Diode (LED) light bulbs, which cost a total of $90. He then asks you if these bulbs are ok, or if you would prefer Compact fluorescent (CFL) light bulbs which will cost $36. If you prefer to switch, he will send over a contractor to switch the bulbs **the next day**. There will be no labor charge for switching the bulbs.

We now also include an example of the scenarios in the main manuscript with the full scenarios in the supplementary materials.

> **7)   Incandescent lightbulbs have been phased out in the EU. I am not knowleable about current legislation in the USA, where the study will be conducted, but it may be worthwhile to check legislation across states to check whether this scenario is still applicable (in case the authors haven't done so yet).**

This is a very important comment, good catch!

We are also not based in the US and so we were not up to date with the recent developments, but your comment has helped us realize that we need to check things more seriously.

Indeed, you are right, there are complications. After consultation and long consideration we decided that these issues require us to change our study. After some research, our understanding is that the US government adopted a new energy efficiency plan to phase out the use of Incandescent bulbs. US is also in plans of moving away from CFL to LED, which presents an opportunity to make an inefficient-efficient contrast similar to that in the target article to be about a contrast between CFL and LED bulbs.

To address this point, we decided to modify the scenarios by replacing the INC bulbs with LED-light bulbs. Hence, we shifted from the old choice between INC (less efficient but cheaper) and CFL (more efficient but costlier) to the more updated choice between CFL (less efficient but cheaper) versus LED (more efficient but costlier). Given additional feedback, we removed the past behavior manipulation and added a control condition, and therefore our new paradigm is a 3

status quo conditions (CFL versus LED versus control) by 3 default conditions (CFL versus LED versus control).

To us, this shows the importance of peer review over plans and Registered Reports and the need to carefully consider needed adjustments when revisiting classics. Without this change, it is very possible that our replication would have confused participants and resulted in weird findings.

We added a planned discussion following Stage 2 in our Discussion section.

Reply to Reviewer #2: Dr./Prof. Julia Nolte

Thank you for giving me the opportunity to review Stage 1 of "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)" by Yam and Feldman.

The work in question proposes a replication and extension of two of the studies reported in Dinner et al. (2011). The authors suggest that although Dinner et al. (2011) identified their work as pertaining to default effects, they actually had measured status quo bias. As a result, Yam and Feldman plan to replicate Dinner et al.'s (2011) status quo bias findings while also extending said research to the role of default effects.

To do so, the authors provide a detailed study plan with a large planned MTurk sample of N = 1500 US adults. The authors' basic premise for the need for a replication appears sound, as do the proposed changes to the original study design by Dinner et al. (2011). Below, I offer additional suggestions for further improving the quality of the manuscript anful to d of the planned research.

Thank you very much for the positive and constructive feedback.

Writing quality

The manuscript requires careful proofreading. Below, I list some of the examples of typos and errors I came across.
- **Page 5: Samuelson and Zeckhauser → Samuelson & Zeckhauser**
- **Page 5: Misplaced "?" and " → (tendency to follow what one has done in the past? as an additional reference point factor),**
- **Several pages: Proofread for consistency (e.g., periods vs. commas vs. no punctuation when listing hypotheses)**
- **Page 5: Missing period when listing "Dinner et al"**
- **Keep spelling consistent (e.g., fluorescent vs. Incandescent; status-quo vs. status quo, Status quo vs. status quo, Compact fluorescent vs. Compact Fluorescent)**
- **Page 9: Stay consistent and refer not to "status quo effect" but to "status quo bias," which is the more commonly used term.**
- **Page 12: The sentence "Following the recent…" is not a grammatically complete sentence**
- **Page 15: meta-analysis → meta-analyses**
- **Page 25: in the page → on the page**

For improved readability, I would also suggest splitting longer run-on sentences. For example, on page 11, the sentence "At the time of writing…" spans six lines.

Thank you for pointing out the above writing mistakes, we appreciate that very much. We tried our best to correct these and other mistakes we found while revising.

Definitions and Specifications

Given that the paper's premise is that Dinner et al. (2011) did not correctly define the studied construct, I encourage the authors to define core concepts such as "nudge."

Good suggestion. We now refer to nudge and provide a brief definition in the introduction before discussing status quo and default effects.

> **Page 10: Your example of status quo bias is "a person, who has been enrolled in a health insurance plan for several years, not acting to change to a new plan that guarantees better coverage and cheaper installments with less risks." I want to raise two points here.**

> **First, in your definition of status quo bias, you present the status quo as created through somebody else's decision, whereas your example identifies the decision as made by the decision maker of interest. Please specify that status quo bias can result both from choices made by others and choices made by the decision maker of interest. In your design, status quo is set by somebody else, although you also account for simulated past behavior of the decision. It seems relevant to highlight the distinction.**

Good point. Thank you.

Regarding to the example of status quo, we added the following to "Status quo bias" in the introduction:

> Status quo could be a result of an individual choice or an externally set option. An example of the status quo bias is a person, who has been enrolled in a health insurance plan for several years (either self-enrolled or enrolled automatically by their employer), not acting to change to a new plan that guarantees better coverage and cheaper installments with less risks.

> > **Second, please make sure to explain how your example differs from the sunk cost fallacy (as some might argue your example applies to said fallacy).**

We understand this concern. We considered a more in-depth discussion of how status-quo bias differs from other biases, yet finally decided against it, given that a full discussion of differences between status-quo bias and other action-inaction biases is beyond the scope of our manuscript and we already tackled that in Feldman et al., (2020).

> Feldman, G., Kutscher, L., & Yay, T. (2020). Omission and commission in judgment and decision making: Understanding and linking action-inaction effects using the concept of normality. *Social and Personality Psychology Compass*, 14(8), e12557. doi: 10.1111/spc3.12557

We therefore added the following:

> For an in-depth discussion of the differences between status-quo bias and other action-inaction biases see review by Feldman and colleagues (2010).

We initially thought of adding the following paragraph but finally decided against it given that it feels too specific against one bias, and does not fit well with the flow of the manuscript, yet we add it here to explain how we see some of the differences:

> A related yet separate phenomenon is escalation of commitment (or, sunk cost fallacy), in which one has already invested resources in a specific course of action and despite negative feedback maintains or escalates their commitment by investing further (Arkes & Blumer, 1985). Status quo does not necessarily involve sunk costs or negative feedback that things are going badly and does not require a decision on whether to withdraw, maintain, or escalate. Escalation of commitment may also involve a decision to change from the status quo in order to try and recover what has been invested so far.

> Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. Organizational behavior and human decision processes, 35(1), 124-140.

> **On page 3, the manuscript suggests author Yam took charge of pre-registration, whereas elsewhere it says no pre-registrations have occurred. If necessary, correct on page 3 as "planned pre-registration."**

Much of the manuscript of this Stage 1 is written as a simulation of what the manuscript would look like after data collection in Stage 2, and we added clear indications that this is Stage 1 of a Registered Report with warnings throughout the manuscript in each section that this is the case and that no pre-registration has taken place. We would rather keep this consistent through the manuscript and keep things in past tense to not have to change the entire manuscript again after IPA from future tense to past tense.

Statistics

Page 19: Specify the version of "G*Power " used. I am surprised by the authors' choice to apply a power of .95. Although .95 is the default setting in G*Power, the norm in Psychology research is to apply a power .80. Although I realize the authors are proposing a very large sample of N = 1500, for the sake of power analyses, I would still strongly encourage the authors to re-run their power analyses with power .80 and to update their planned sample size

We find this to be a surprising suggestion. We are unsure about how you've come to consider the norm in psychology to be a power of 0.8 and why given the aim for higher power you would encourage lower power.

In any case, in most of our replications, many of which are either published or with in-principle acceptance from PCIRR (with the same handling recommender) we based our power analyses or sensitivity calculations on .95 to try and be more conservative about the effects that we are likely to detect. Given that these are sensitivity analyses, and higher power aim is more conservative, we do not see this to be of major importance. We kept our calculations to be based on .95 power.

However, given the new study design and your comments, we redid the sensitivity analysis and adjusted to address possible exclusions of 10% and to adjust contrasts to two-tail instead of one-tail.

The authors' did not specify planned analyses should the obtained data violate the requirements of parametric tests. I encourage the authors to put in place planned tests to examine data distribution and to propose non-parametric alternatives for planned analyses.

Planned analyses were not included in the target article and consequently our analysis might different from the target. In case of replication of the status-quo fails, we aim to run additional exploratory analyses to examine the normality and run non-parametric tests if normality is violated, with a lower alpha to compensate for multiple comparisons (i.e. $p = .005$).

We added the following to the "Data analysis strategy":

In case we fail to find support for our replication hypotheses (status quo effects), we will conduct an additional set of analyses examining normality and the need for running exploratory non-parametric tests. We reserve this for case of failure, given that our main aim is to follow the target and the target did not implement such analyses. In such a case we will report both tests side by side with the non-parametric tests clearly marked as

exploratory, and to address issues of multiple testing we will use a stricter lower target alpha for the non-parametric tests set to .005.

Page 34: add 'between' or 'within' to factor concerning default effects.

Thank you. We went through the manuscript again to ensure that all references to both status quo and defaults indicate that this is a between-subjects design.

Design

As the authors point out, the presence of pre-installed light bulbs acts as a status quo for the decision maker. To properly study the distinction between status quo bias and default effects, it seems to me that replicating the Dinner et al. (2011) task design without pre-installed light bulbs, but with a default option pre-selected for the decision maker, would be a more direct and straightforward way of testing default effects in the context of Dinner et al. (2011). Rather than making the study's design more complex my including another condition, I would encourage the authors to consider including a second sample which takes this approach (and this approach can also examine past behavior as a reference point etc). If doing so falls outside of the authors' means or the scope of this manuscript, I would encourage the authors to justify the choice not to directly run such a test/ to discuss the limitations of their extention.

This is very good feedback, thank you for that. We agree, a control condition for status-quo is missing from the target and would provide important useful information in our conceptual replication.

Initially, we planned for a 2x3x2 (status quo vs. default vs. past behavior) experimental design, and now revised it to a 3x3 (status quo vs. default) between subject design, with a control condition for each of the independent variables.

We summarized the new design in Table 4, and updated the manuscript accordingly throughout.

> **For any study, I would always advocate for assessing information concerning participants' demographic background. Given the authors are testing a decision with financial consequences (e.g., spending more vs. less), I would expect an assessment of participants' income or sociodemographic background.**

In our previous submission we already included a self-report question regarding social class. These were meant as exploratory and we did not plan to go into reporting these in-depth.

To address your comment, we now plan to report the distribution of participants' self-reported social class and other collected key demographic information in the "Participants" subsection of the method section.

Reply to Reviewer #3: Dr./Prof. Sylvain Chabé-Ferret

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, NFL 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 ×3 ×2 design

This is an important replication since defaults are among the strongest of effects in choice archi- tecture 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:

Thank you for the positive and constructive feedback.

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.

Thank you, great suggestion.

We initially aimed to provide only a concise example and attempted to illustrate the two effects. We appreciate the opportunity to expand on this further, and therefore added the following under the subsection "Differentiating between default bias from status quo bias" in the introduction:

> The distinction between the two biases seems especially important given that the impact of the two biases may vary depending on the context of the decision-making situation. Default options seem to be better suited for form or survey decisions in making certain options more salient, guiding decision-makers towards a recommended option, or helping those with decision fatigue or reluctance to choose to default towards a choice that policy makers would deem as the one best suited for them. For example, defaults have been shown to help guide patients' end-of-life choices (Halpern et al., 2013). Status quo bias seems to convey previous personal or external decisions that can be used to guide or justify current decisions, signaling that the option has had merit and/or has been vetted and tested, such as in evaluating new technologies (Smiley & Fisher, 2022). Changing from the status quo typically requires much more effort and investment than does selecting the non-default option.

> Contrasting the two biases against one another could help better understand whether the two biases are complementary or supplementary to one another, and which one of the two is stronger. If, for example, status quo is found to be the stronger among the two, then

labeling status quo biases as default effects may lead to an overestimation of default effects and a disappointment when implementing default options in real life.

> **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.**

Thank you for this feedback.

We understand your concerns on excluding Query Theory from our replication project. This was not an easy decision, and we contemplated long and hard before deciding to do so. Our take on this is that given our aim to disentangle status-quo and default effects this is already a very complex replication, and we wanted to ensure that the underlying phenomenon holds and is clear before we embark on trying to understand why that phenomenon happens and how. This is one of the reasons why we also chose to categorize the project as a "conceptual" replication, as we shifted focus from a direct replication to disentangling default and status quo effects.

To address your comment we added this as a planned discussion in the Discussion section:

> [Planned discussion: In our replication we chose not to examine query theory from the target article. Our main aim for this project was to first establish the phenomena. In this section, based on our findings, we will discuss future directions for follow-up research to test query theory and other possible explanations and mechanisms for the two effects.]

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.

Thank you very much for this feedback, we appreciate that.

After considering your and the others' feedback, we decided to remove the past behavior manipulation and instead implement a control condition for status-quo.

Please see our reply on this point above.

There are many challenges to replicating the target's Study 3 and the measures employed to indirectly examine Query Theory. The decision to not include Study 3 was not about the effort, but more to do with complexity and clarity of the project, as the scope and scale of the current manuscript was already very ambitious. We also wanted to focus on the simplest methods that we thought had the biggest potential for replications, and the way the target article tried to tap into Query Theory seemed rather noisy and overly complex, and therefore more risky, hence best kept for after we verified the underlying core phenomena.

> **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.**

Thank you, this is valuable feedback.

In our previous submission in the OSF we included the Qualtrics survey export files and printout and a link to our Qualtrics preview to allow you to browse the survey, but we now realize that we should have done better - it should be simpler for readers to review and understand the methods without having to look for it in external files or links .

We therefore now updated the main manuscript with examples and the supplementary materials with the full vignettes and screenshots of the manipulations.