Editorial Comment: One additional revision for the authors to make is in Footnote 3:
"This manuscript is a Stage 2 Registered Report of this Stage 1 Registered Report https://osf.io/xw6hn"
Replace that with the following (noting the change of URL to the official PCI RR registration)
"This manuscript is a Stage 2 Registered Report of this Stage 1 Registered Report: https://osf.io/ztucr (date of in-principle acceptance: 23/04/2022)"

Response: We have now changed the footnote accordingly!

Reviewer 1

Comment 1: I didn't read the final Stage 1, so this is based on my reading of the Stage 2. It is clear and concise. However, the registered and non-registered analyses should be in separate sections to more clearly separate them.

Response 1: The revised manuscript now splits the discussion section into 'Pre-Registered Analyses' and 'Non-Pre-Registered Analyses' in response to this concern.

Comment 2: Further, their use of the obtained effect size from their previous study as a basis for the power calculation for the current study has led to a non-signfiicant result with an effect size V = .09, with the original significant effect size being V = 0.1. I realize already vast numbers of subjects were needed for the RR as it is; but still the upshot is, the study was not powered to detect all effects of interest (if V = 0.1 was of interest, then so is V = .09). Thus, the abstract and discussion need to conclude more along the lines of "reserve judgment" rather than "we failed to replicate". The abstract and discussion should point out that although the current study was non-significant, it was not powered to detect all effects of interest.

Response 2: We have changed the wording in the abstract according to your recommendation. We continue to say that we are treating this pattern as being evidence for a failed replication given our study specifications, but we then go on to insert your request to phrase it in terms of reservation of judgement. We do the former primarily because this is what we set out in the pre-registration. We hope that the adjusted wording in the abstract and the discussion section are now adequate.

Example 2 (p. 1, p. 17-18): *[…] In this registered report, we further investigate this mechanism by (i) attempting to directly replicate the previous result and (ii) analysing if the Bayesian Truth Serum's effect is distinct from the effects of its constituent parts (increase in expected earnings and addition of prediction tasks). We fail to find significant differences in response behaviour between participants who were simply paid for completing the study and participants who were incentivized with the BTS. Per our pre-registration, we regard this as evidence in favour of a null effect of up to V=.1 and a failure to replicate, but reserve judgment as to whether or not the BTS mechanism should be adopted in social science fields that rely heavily on Likert-scale items reporting subjective data, seeing*

that smaller effect sizes might still be of practical interest and results may differ for items different from the ones we studied.

[…] The data presented in this paper do not show any significant differences between the Bayesian Truth Serum condition and the No Incentive control condition. As pre-registered, we treat this pattern of data as being evidence in favour of a null effect of up to Cramer's V=.1 and as such a failure to replicate the results of Schoenegger (2021). However, the current study was not powered to detect all effects of potential interest. Accordingly, we reserve judgment as to whether the BTS mechanism should be adopted in social science fields that rely heavily on Likert-scale items reporting subjective data as we have studied in this context. This reservation of judgement then opens up the space for further research as our inability to recommend the Bayesian Truth Serum as an incentivisation mechanism that ought to be applied widely leaves open the central question of how to properly achieve this task and what the effect of the BTS is in different contexts, for instance, for items that are different in nature than the ones we considered here. It may be that the Bayesian Truth Serum's applicability is more restricted than we anticipated, that another mechanism is better suited for this context, or that the present study was simply not sufficiently powered to detect small but still relevant effects. This is why we argue that, going forward, issues of incentivisation ought to remain central in further (social) scientific reform efforts and we call for more research in this area.

Reviewer 2

General Comment: my review of the paper is complicated by the fact that I missed the second review moment (finalization of stage 1). I also did not see a way to access the comments and replies pertaining to that round.

I note that the authors didn't follow all of my suggestions (the use of regressions to look at treatment differences, and explicit hypotheses), but because I don't have (or am not able to find) access to all the materials, it may be that this issue was discussed and the suggestion discarded for good reasons. I also don't know what final analysis was agreed upon.

Response: We are sorry for the confusion, but we have addressed these points in our previous review letter, drawing on old data from the previous work to motivate our omission of regression analyses.

Comment 1: However, looking at the track changes document, it appears to me that the authors followed their proposed methodology, and clearly flag preregistered and non-preregistered analysis. Since they find no effect, some of their analysis (and my criticism of it) has become less relevant. In particular, the decomposition of the effect was the source of some criticism of mine in the first round (I thought they should use more directional tests in disaggregating any effects), but given the null result, the issue is largely moot.

Response 1: We are happy that you agree that the manuscript properly follows the proposed methodology and that you agree that, given the pattern of results present, the suggest analyses that we did not end up following would not have been of interest anyways.

Comment 2: I thus think the paper meets the criteria for the registered report. My apologies again for not going through all stages of the report, and hence this somewhat handicapped final evaluation.

Response 2: Thank you!