Thank you for the opportunity to review this Stage 1 RR manuscript. I have organised my review according to the five criteria for PCI Stage 1 RRS, with two extra sections at the end. Overall I thought this was an extremely impressive manuscript, which displays a strong knowledge of the relevant literature, with appropriate and deeply considered methodological choices. My review here is quite long, but that’s more of a reflection of the fact that I was very satisfied with the broad strokes of this manuscript, and focused on relatively specific details in my review. I am confident that the issues I’ve raised below are things the authors could feasibly address, so am happy to recommend R&R.

# 1A. The scientific validity of the research question(s)

I agree that the key research question (as specified in the manuscript title) is valid.

# 1B. The logic, rationale, and plausibility of the proposed hypotheses (where a submission proposes hypotheses)

The manuscript provides a coherent and sensible justification for the proposed hypothesis that “brooding about distressing societal issues increases conspiracy beliefs”. The hypothesis seems sufficiently plausible to justify testing it.

# 1C. The soundness and feasibility of the methodology and analysis pipeline (including statistical power analysis or alternative sampling plans where applicable)

In general I think that the planned methods and analyses are very sound, feasible, and rigorous. I was impressed especially by the effort that the authors put into planning the manipulations and the equivalence tests. I nevertheless have some minor/specific points to raise.

1. The paragraph at the top of page 19 discusses the fact that different participants will be ruminating about and answering conspiracy beliefs questions about different social topics. I agree that the use of randomisation means that this isn’t a plausible confounding variable. That said, have the authors considered applying an analysis method that involves a random intercept for social issue? It’s possible that accounting for variance due to issue might increase power, and this could also facilitate generalisations *across stimuli/issues* (see Judd et al., 2012; Yarkoni, 2022). This would obviously complicate the analysis plan somewhat so I consider this an *optional* suggestion, but I’d be interested to know the authors’ thoughts.
2. I can understand why the authors have used a no-distraction control condition, but it does strike me that this means that participants in the reflection and brooding conditions will spend quite a bit longer on the study than those in the control condition. Consequently, the study will implicitly manipulate payrate (per hour) along with rumination. The participants in the reflection and brooding conditions will also spend this extra time answering questions that are *open-ended* and *repetitive* – a task which Prolific participants tend not to enjoy (going by what I see them saying on the Prolific subreddit!) Consequently, the manipulation may affect irritation levels, which at least in theory could thus provide an alternative explanation for an apparent effect (e.g., maybe the “real” effect could be assignment to reflection or brooding conditions >- longer study -> lower payrate -> irritation -> belief in conspiracy theories). I don’t think this is an especially strong or plausible threat to the validity of the findings, but it may be worth considering as a minor limitation.
3. Relatedly, from an ethical perspective, you might also want to consider whether it would be appropriate to pay participants in the more time-consuming conditions a bonus to compensate them for the extra time, or pay everyone a reward commensurate with the time of the longest condition.
4. It might be useful to build in an exclusion condition based on Prolific ID to check for inadvertent duplicate responses (this can happen when a participant has a problem with the survey webpage or something).
5. Re. statistical power, the manuscript says that “The suggested design simultaneously controls the Type I and Type II error rates of both the original null hypothesis test (which determines whether there was a significant effect on conspiracy beliefs) and the equivalence test (which determines whether values above d = 0.20 can be rejected): In the equivalence test, the original null and alternative hypotheses become reversed, so that the Type II error from the original null hypothesis test becomes the Type I error from the equivalence test, and vice versa (Lakens et al., 2021).” This would make sense if the authors planned only a conventional “nil” NHST plus an equivalence test, but as far as I can tell it doesn’t take into account the planned minimum effect tests. I.e., if the true d is 0.20, the power of the minimum effect test (where d = 0.2 is the “null” hypothesis) would be equal to a = 0.05, not 90%. Therefore, a power analysis for the minimum effect tests would need to hypothesise a true effect size greater than the SESOI of 0.20. Depending on what hypothesised effect size is chosen this could have substantial implications for the required sample size.
6. Relatedly, do the planned minimum effect tests have any implications for the alpha spending function in the sequential analysis? (I’m not very familiar with sequential analyses myself!)
7. I think the repeated measures design is appropriate for increasing power, but one risk it presents is that it could make it a bit easier for participants to guess the hypothesis. I don’t think this is a crucial or unusual feature of the study, but perhaps this could be flagged as a limitation in the eventual discussion.

# 1D. Whether the clarity and degree of methodological detail is sufficient to closely replicate the proposed study procedures and analysis pipeline and to prevent undisclosed flexibility in the procedures and analyses

There is a great deal of well thought-through methodological detail in the manuscript already, and this criterion is close to being met. However, it could be useful to resolve the following potential ambiguities:

1. I take it that all the materials presented to participants will be in German?
2. Although I appreciate that the conspiracy belief items for the RR are shared on the OSF project, given that these constitute the key DV it’d be useful to list all three in the manuscript itself rather than just one example (bottom of p. 19).
3. P. 20, please ensure that all measures that will be administered are listed (“such as” implies that there are other unstated measures).
4. What payment per participant will be provided on Prolific? Will they receive separate payments for the baseline and main study surveys?
5. Will participants be paid if they meet exclusion criteria? (I’d suggest they should be, since the [criteria](https://researcher-help.prolific.co/hc/en-gb/articles/360009092394-Approvals-rejections-returns) Prolific have for rejecting submissions are very narrow, and don’t really cover any of the exclusion criteria listed in this study).
6. P. 21, brooding manipulation – if participants complete all 7 questions twice, will they then be presented a third time? And what happens after 5 minutes – does the page automatically progress even if the participant is still typing?
7. P. 23, re. the question “What were your thoughts in the 5 minutes before we asked you the questions about the conspiracies?” Why not just put this question before the conspiracy ones, and say “..in the last 5 minutes”? Perhaps I’m missing something here.
8. P. 24 – “This means that we will consider our effect practically meaningful if the confidence interval of the effect size estimate is beyond *d* = 0.20, and practically negligible if it is below *d* = 0.20.” Does this mean that you will consider the effect practically meaningful if the lower limit of the 90% confidence interval for d is greater than +0.20? And practically negligible if the lower and upper limits of the 90% CI for d both fall within the range -0.2 to 0.2? In general, I’d encourage the authors to be a bit more explicit and rigid in how they state inferential criteria in the main text, though Table 5 is very helpful in this regard.
9. Could you please explain how you will apply the sequential sampling plan on Prolific? E.g., will you have just one study running, but pause it at particular points? Or have 3 separate study advertisements in a row, but exclude participants from part 2 if they participated in part 1, etc.? If the latter, be a bit wary with this – Prolific does have a prescreening function for excluding participants from prior studies you’ve run, but in my experience it doesn’t seem to work very predictably (e.g., when I apply it, the available sampling frame doesn’t seem to change). You might choose to use a “custom blocklist” instead or in addition.
10. Relatedly, how in specific will you run the baseline step? Does a 1/3 sequential block involve baseline plus main study survey, or will you first run a very large sample of baselines and then only the second part in blocks?
11. How will missing data be dealt with? Please consider both whether there might be a need to exclude participants based on quantity of missing data, and also how you might handle participants who are included in the final sample but have missing data on some items.
12. Could the authors confirm that all the t-tests mentioned in the analysis plan are Welch’s t-tests?
13. Could you please add a bit more clarity about the sequential analyses? I.e., what will the alpha level be in the first step? The second? The third? (Most readers won’t have used sequential analyses before, so try to step us through the details slowly!)

# 1E. Whether the authors have considered sufficient outcome-neutral conditions (e.g. absence of floor or ceiling effects; positive controls; other quality checks) for ensuring that the obtained results are able to test the stated hypotheses or answer the stated research question(s).

I think this criterion is largely met. The authors have been especially thorough in planning their manipulations check. A few minor comments:

1. P. 25, “So, we would consider an effect greater than d = 0.30 meaningful”. Does this mean that the lower limit of the 90% CI of the effect in the manipulation check has to be above 0.30 for you to consider the analysis to be a valid test of the hypothesis? If this criterion isn’t met, what will you do? Not report the main hypothesis tests at all? Report them but interpret them as being inadequate tests of the hypothesis? Something else? I do think there is a non-trivial chance that this manipulation check will “fail”, so it’s worth really thinking through this scenario. Relatedly, consider that for this manipulation check to have a good chance of “passing”, the true effect size needs to be quite a bit larger than 0.30…
2. P. 26, “That is, we will consider the brooding manipulation effective if (1) the brooding condition scores meaningfully higher (at least *d* = 0.30) on the brooding MC than the control group AND (2) the control group scores meaningfully higher on the ‘thinking as usual’ MC than the brooding condition.” I take it these criteria relate to the lower limits of a 90% confidence interval for d, not the sample estimates of d?
3. I appreciate the attention paid to exclusion criteria, but I’m a little sceptical of the exclusion criteria relating to “nonsense” responses. Consider that participants can *only* be excluded based on these criteria if they’re in the brooding or reflection conditions (people in the control condition won’t be answering these questions). This means this exclusion criterion is confounded with condition, and applying it could bias the resulting estimates of causal effects (e.g., by removing the least attentive participants in the sample from two conditions, but not the other). I’d tend to think that it might be safer to leave out this exclusion criterion in your main analyses, but consider applying it for supplementary robustness analyses.

# Comments on pilot studies

A relatively unusual feature of this Stage 1 RR is the quantity of existing empirical work presented (4 pilot studies). These studies were described very briefly in the main text, and then in more detail in the supplementary file. This was very impressive; I really appreciate the work the authros were willing to do by way of preparation for the RR.

It nevertheless took me a while to get my head around what took place in all these studies and how they relate to the RR, which was a bit of a challenge given the scheduled review format. One thing I puzzled over was the salience of the fact that three of these studies were preregistered, but not in a great deal of detail (the aspredicted template isn’t great for producing detailed registrations), and the reported analyses deviate from the preregistrations somewhat, without particularly detailed rationales. Usually, these would be things I’d grill authors about in a peer review! Yet what’s different in this case is that the authors seem *not* to be using the preregistrations to claim any special confirmatory status for the studies as tests of hypotheses, but are instead presenting them as tentative pilot studies that are used to refine the methods and main hypothesis. In that context, perhaps the deviations from the preregistrations are less important? I’d be interested in what the authors think about this topic, and also Chris’s take! Perhaps it’s something that could be touched on, albeit briefly, in the main text (currently I don’t think the preregistrations are mentioned there).

Beyond that, I have some more minor comments on the pilot studies, relating primarily to content in the supplement:

1. I assume all pilots were conducted in German?
2. In the supplement, please state all measures administered in the pilot for the sake of transparency, even if not all were used in the analyses presented here.
3. For replicability, please indicate which survey company or companies recruited participants.
4. For replicability, please indicate how participants were compensated.
5. In pilot study 2a, the significance test of the manipulation (i.e., effect on perseverative thinking) has p = .048 given a one-sided test, but the preregistration doesn’t say the test would be one-sided (and it wouldn’t be significant if 2-sided). Considering also that p values near 0.05 are only weak evidence against the null anyway (Benjamin et al., 2018) it might be worth noting that result provides only quite tentative evidence of an effect.
6. The prereg for Study 2a also says “We will check assumptions of normality and homogeneity of variance. If not violated, we will conduct student's t-tests to compare means in conspiracy beliefs between rumination and control conditions for each scenario separately. If assumptions are violated, we will conduct non-parametric tests of mean differences." That doesn’t link particularly closely with the Welch’s test reported, which is a reasonable choice but neither a Student’s t-test nor (strictly speaking) non-parametric. In general, I think it’d be useful in the supplement to signal and explain deviations from the preregistrations in a bit more detail, even if these deviations may be less crucial in the context of these being pilot studies than if they were being presented as confirmatory.
7. Pilot study 2b, “The main goal of this study was to replicate the finding from pilot experiment 1”. Does the latter refer to pilot Study 2a?
8. Pilot study 2b, “Given that the manipulation failed to exert a significant influence on state rumination, we neither expect an effect on conspiracy beliefs” – I understand what the authors mean, but it’s a bit of an odd phrasing given that the preregistered hypothesis said otherwise.
9. Am I right in inferring that pilot study 3 was not preregistered? (This seems fine, but perhaps mention this, since the others were)
10. I fully appreciate that the authors will be aiming to keep the descriptions of the pilot studies in the main text brief so as to meet word limits at PCI-friendly journals, but perhaps a tiny bit more info about fundamental issues like country of sample, language and recruitment method might be helpful to readers who look only at the main text.

# Comments on introduction

I thought the introduction was excellent, but just have a few very minor suggestions for clarifications.

1. Defining conspiracy beliefs subsection: “A conspiracy is a secret plot by a powerful group that aims to achieve a common goal. Importantly, the conspirators pursue their goal **without any regard for other people or consequences for society** as a whole.” This latter part of the definition is quite strong and a bit unusual relative to conventional definitions of “conspiracy”, which tend only to require that the conspirators are plotting to harmful or malevolent ends (one can have some regard for others while still doing something harmful). C.f. a [dictionary](https://www.dictionary.com/browse/conspiracy) definition. The authors aren’t “wrong” for using this definition (definitions are always socially constructed) but I’d be interested to know why they prefer it!
2. “Conspiracy beliefs have harmful consequences for individuals and societies” – I’d suggest softening this a bit to be more consistent with the rest of the paragraph, where you acknowledge that it “lies in the public interest to uncover true conspiracies”.
3. “Studies have shown that rumination increases depressed mood…” – Were these experimental studies (thus justifying this causal inference)? In general, you could perhaps give just a little more information about the methods of key studies you’re reviewing, so that the reader has some basic impression of the strength of the evidence in these prior studies.
4. “Considering that conspiracy beliefs tend to emerge when people experience negative affect…” The phrasing here could be read as implying that there is *empirical evidence* that conspiracy beliefs emerge when people experience negative affect, whereas I think you're just intending to say that this is what some *theories* assert. As you'll know, the empirical evidence that negative affect causes increased belief in CTs is somewhat tentative.

Lastly, my apologies for any typos or lack of clarity in my review; given the scheduled review format I thought it was more important to do this quickly than to carefully draft a very elegant review document! Please likewise forgive me if any of the points or questions I’ve raised above are already covered in places I’ve missed.