Referee Report: Nadeau et al. (2024) – esd-2023-23

1. General comments

In the revision process, the authors have made several important changes that have improved the quality of the manuscript. The introduction is now more concise, the methods section is more detailed and informative. The materials that are now provided make it easier to understand and reconstruct the methodology. However, now that the methodology is properly disclosed, some notable flaws have also become evident. Specifically, I have identified one major issue and multiple minor issues that primarily relate to the processing and interpretation of the collected survey data.

2. Specific comments

1) Section 4.1 & 4.2: Strictly speaking, the authors only used two questions to assess the "general state of knowledge of climate tipping points" in the Norwegian population: 1. "How familiar are you with the CTP concept?", and 2. "Can you give an example of a CTP? If yes: _____ ". This is not a comprehensive 'CTP knowledge test'. The authors did not specifically ask for a definition of the concept or a description of CTP characteristics. Nevertheless, the authors try to draw conclusions about laypeople's awareness of CTP characteristics from their answers to the second question, because it turned out that some participants (n = 161) submitted general comments on the CTP concept/ CTP characteristics. It is not entirely clear why that happened. It seems that many participants were unable to provide specific CTP examples, or they simply misunderstood the question. In any case, these general comments should not be used to draw statistical conclusions about laypeople's awareness of CTP characteristics. The comments are only the side product of an open-ended question that focused on a different issue. The authors should note that the frequencies displayed in Table 1 do not include the potential responses of the roughly 700 participants who chose not to provide general comments. Some of these participants might have been able to write something about CTP characteristics if they had been asked about this specifically. And these individuals might have had other characteristics in mind than the remaining participants. Thus, I would strongly advise the authors to disregard the general comments about CTP characteristics and focus only on the question of whether or not participants were able to recall a CTP example in response to this prompt question.

From a methodological perspective, this should then also preclude any broad conclusions about CTP knowledge in the general population of Norway. After all, the authors only asked for familiarity ratings and examples for CTPs. A categorization of participants into different 'CTP knowledge categories' seems inappropriate. Instead, the most important results of the survey can be summarized in three short paragraphs: A) one paragraph describing the familiarity ratings, B) one paragraph stating how many participants were able to provide a correct CTP example vs. how many were unable to do so/provided general comments, and C) a paragraph outlining which CTPs were mentioned most frequently. With this in mind, I would strongly encourage the authors to remove section 4.2, rewrite section 4.1, and reformulate their conclusions regarding the state of public knowledge about CTPs in Norway (see discussion).

Minor issues

2) Section 4.4: The authors still do not provide standardized effect sizes even though these are essential for the interpretation of the treatment effect. From the mean values and standard deviations, it can be derived that the effect that was observed here is really small (Cohen's d = 0.08 for the pre-post difference in Group A, with SD(y) in the denominator of d; note that in the social sciences, d = 0.2 constitutes a small effect, d = 0.5 constitutes a medium effect, and d = 0.8 constitutes a large effect; see Cohen, 1988). This should be explicitly acknowledged and discussed in the manuscript.

In this context, I would like to draw the authors' attention to the fact that the study by Formanski et al. (2022) found a difference of d = 0.04 between their experimental condition (non-linear climate change portrayal) and their reference condition (linear climate change portrayal). But that difference was nonsignificant, perhaps because the sample size used in that study (N = 360) was too small to detect such a tiny effect. Nonetheless, it can be seen that the effect size obtained in that previous study was not so different from the effect size observed in the present study, which used a large sample of N = 851. In large samples, even small effects become statistically significant. Given these considerations, I would not flatly conclude that the findings of the present study "contrast" with the results of Formanski et al. (see line 477) – that study concluded that such an effect might not exist – but if it exists, it is likely to be small. The present study now provides evidence that this effect could indeed exist, but that it is likely to be small (= the conclusions are not entirely contradictory).

3) Section 4.4: The present study also found that exposure to CTP information did not influence fatality ratings ("Is it too late to do anything about climate change?"), which is essentially congruent with what Formanski et al. (2022) found on their efficacy beliefs measure. This result is not mentioned anymore in the revised version of the manuscript, even though a) the item is mentioned in the methods section and b) the introduction raises the (very important) question of whether exposure to information on CTPs induces fatalism/reduces efficacy beliefs. I believe that this null finding is informative and that it should be described and discussed in the manuscript. I would just like to note that I am skeptical of the explanation the authors initially proposed for this finding. In the first draft, the finding was attributed to "the public's tendency to downplay the seriousness of these risks due to certain cognitive biases […]" – but that would have only made sense if the materials had, in some way, suggested that crossing climate tipping points is inevitable. A more plausible and straightforward explanation for this null finding could be that exposure to information about CTPs may not necessarily promote fatalism.

4) Section 3.1: The authors should clarify which items were used to measure climate change risk perceptions (CCRP). The keywords they list in line 273 do not match the questions q1r1-q1r3 in the appendix. In addition, the authors should follow common reporting conventions and provide a reliability estimate for their CCRP scale (e.g., McDonald's Omega).

5) Section 4.4: In the context of randomized controlled trials with a pre-post measurement, the ANCOVA technique is only used to compare the mean post-test scores across the experimental conditions (Group A vs. Group B), while including the pre-test scores (here: CCRP at t0) as a covariate in the model (e.g., Frison & Pocock, 1992). What this means here is that the ANCOVA should replace the independent sample t-test for the post-test scores (lines 419-421), because the ANCOVA is simply more informative as it also takes the pre-test scores into account. The paired-sample t-tests (lines 430-433) can still be presented as a follow-up analyses – these tests provide information that is not directly uncovered by the ANCOVA. The analyses that are currently presented in the first paragraph of section 4.4 are either irrelevant to the research question (--> effect of time averaged across the two conditions, see lines 409f), or already reported in the text (difference between the groups at t0, see line 412 vs. lines 427f), or not precisely described (line 411 – is this the result of the ANCOVA that is announced in the second sentence of the paragraph?).

6) Lines 105f: "Some studies have demonstrated that instruction, information, and knowledge about climate change increase climate risk perceptions (Aksit et al., 2018; Milfont, 2012; van der

Linden, 2015; Xie et al., 2019)" – Most of the studies cited here are only correlational studies, which is why the authors might want to rephrase this sentence; see also line 244 – same issue here

7) Lines 301-305: The main drawback of the sample is that it is not a probabilistic (random) sample. This means that the sample composition could differ from the composition of the general population in terms of relevant characteristics that were not considered in the quota plan (e.g., education, income, social status, personality traits…).

8) Lines 480f: "Our results were not independently verified by an unbiased and impartial third party, which is a limitation of our study." – Do the authors mean that the statistical analyses were not re-run by a third party? As far as I know, this is not common practice in social sciences studies. What is more common is that the data set obtained in a given study is made publicly accessible (in an anonymized form), so that everyone can re-run the analyses.

3. Technical corrections

1) Milfont (2012) is missing in the reference list

2) Line 167: "the likelihood of triggering climate tipping points is "dangerously close"…" This sentence should be rephrased.

3) Lines 170f: "e.g. shift in turbid and clear-water phase in lakes" – I would advise the authors to only cite examples of climate tipping elements here, to avoid confusion.

4) Lines 119-125: This paragraph could be removed to reduce the length of the article.

5) Lines 63-68: This part could be shortened – some of this information is repeated in section 2.4.

6) Line 407: "Effect of climate tipping points on Level of Concern for Climate Change", should be changed to "effect of information about…/ effect of exposure to information about…"

7) The results of F-tests and t-test should be reported in a consistent format throughout the manuscript (e.g., APA format).

8) In its current form, Appendix B lacks structure and contains several unclear phrases, e.g., "Your local environment - When do you think the climate crisis will start to affect the following?" – here, 'your local environment' should be placed after the question.