Author Response: 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.

Response: We would like to thank the reviewer for taking the time to provide thorough and valuable feedback on our revised manuscript. We appreciate the positive comments on the improvements. We acknowledge the thorough response with constructive suggestions. While we have adhered to most of these, there are cases where we also disagree, and below we have addressed the major and minor comments in detail.

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.

Response: The Reviewer is concerned with our interpretation of the survey results regarding knowledge, i.e., what kind of interpretations our data allows, and what constitutes a comprehensive knowledge test. We do not claim to have provided a comprehensive

knowledge test, but an assessment of awareness and understanding that both recognizes the complex nature of the variable we label ‘knowledge’ and provides meaningful insights about this variable for a national population, i.e., captures what is called ‘public understanding of science’.

Knowledge can be defined as cognitive processes that relate to awareness or familiarity with information that is believed to be factual. Different types of knowledge (e.g., scientific, elite or decision-maker, and public) and different degrees of understanding can be distinguished, e.g., abstract-conceptual knowledge to usable-practical knowledge. And there are various challenges related to the display/demonstration of knowledge, e.g., recall/memory vs. active generation of information. In our case, we are studying public knowledge of a scientific concept. In this context, the expectation is to gauge general familiarity with a phenomenon without necessarily expecting a study participant to provide a scientific definition of the concept.

In the case of climate tipping points, this methodological challenge - eliciting responses that demonstrate knowledge - is compounded by the fact that multiple scientific definitions exist, with some ongoing disagreement about the necessary conditions for a tipping process. In many ways, definitions depend on context and scientific discipline. Again, it would not be reasonable to expect members of the public to provide a definition, which is why our survey questions did not seek to elicit this form of knowledge demonstration. Instead, we opted for a familiarity rating on a scale, which is a common approach (Ladwig et al. 2012).

Ladwig, P., Dalrymple, K. E., Brossard, D., Scheufele, D. A., & Corley, E. A. (2012). Perceived familiarity or factual knowledge? Comparing operationalizations of scientific understanding. Science and Public Policy, 39(6), 761-774.

Further, examples are key for learning about a novel concept. Major cognitive theories (prototype and exemplar theories) (e.g., Park 2013) explain knowledge acquisition (i.e., learning a new concept) in terms of category development, which relies on examples. While the definition of tipping points remains unstable, there is broad agreement among scientists regarding examples of climate tipping points. All media reporting on the topic - the key source of public knowledge - also tends to focus on examples. For example:

- https://www.nytimes.com/2024/02/15/climate/tipping-points-for-the-planet.html
- https://www.nytimes.com/2023/07/25/climate/atlantic-ocean-tipping-point.html
- https://grist.org/climate-tipping-points-amazon-greenland-boreal-forest/
- https://www.carbonbrief.org/explainer-nine-tipping-points-that-could-be-triggered-by-climate-change/
- https://www.klimastiftelsen.no/publikasjoner/vippepunkter-i-klimasystemet
- https://www.forskning.no/klima-klimatiltak-miljopolitikk/forskere-advarer-mot-vippepunkter-i-en-ny-rapport/2300644

Hence, we developed a survey question that combined self-assessment of one’s ability to provide an example and actual ability to provide an example.

Park, J. J. (2013). Prototypes, exemplars, and theoretical & applied ethics. *Neuroethics*, *6*, 237-247.

Given this cognitive structure of knowledge, and the specific context of climate tipping points, we deliberately designed the survey to ask in multiple stages of increasing cognitive difficulty about participants' familiarity with climate tipping points, providing the opportunity to reveal how well they (believe they) know the phenomenon. The question sequence allows participants to first indicate their degree of self-assessed knowledge (familiarity) with a Likert scale without the need for active recall. For those participants who indicated some familiarity with the concept, i.e., could be expected to have some ability for active recall, this was followed by a question that gauged the self-assessed ability to recall an example. This was considered an easier cognitive task than providing a definition. If a person indicated that they were not able to provide an example, it is of course possible, but very unlikely, that they would have been able to offer a definition or characteristics of tipping point if specifically asked for those.

Participants who answered yes to this question (ability to provide an example) were invited to provide an example, and chose to respond in three different ways: (i) not to provide an example or any other information, (ii) to provide (what they thought was) an example, or (iii) to provide descriptions of the phenomenon instead of an example. While the reviewer suggests it is 'unclear why this happened', we do not think this is the case. Given the survey structure, these response types are expected. We argue that these responses can be interpreted regarding the respondents' state of knowledge:

- No response – The person experienced difficulty in recall and recognized that their self-assessment in the prior question had been too high. They might have some knowledge, but were not able to demonstrate it.
- Example – Demonstration of some knowledge, which can include instances of events that are not climate tipping points, i.e., misconceptions that are cases of incorrect knowledge;
- Broader, descriptive comments – The person has some knowledge but is not able to give an example (i.e., over assessed knowledge), and decided to provide a descriptive account or other conceptual associations instead of an example.

The descriptive comments could be coded based on common characteristics of tipping points. While we do not see this as redundant information or even flawed in the sense that it should be disregarded, we agree with the Reviewer that the data cannot be used for statistical analyses regarding all survey participants (n=851) with inferences for the population of Norway since we did not ask all participants to describe characteristics of tipping points. However, it is not unreasonable to use this information to describe the pattern of knowledge displayed by those respondents who saw and answered the open-ended question (n=153), i.e., the group with the most self-assessed knowledge. Hence, we clarified the circumstances and limitations of data collection and emphasised that our descriptive statistics do not necessarily apply to the whole population. Nevertheless, we have moved Table 1 into the Supplementary

material. We believe that providing this additional information with the necessary clarification about the data elicitation process is interesting and can inform future work, e.g., on the most and least understood features of tipping processes.

We also would like to emphasise that survey design, recruitment and statistical analysis were done in cooperation with a leading polling company in Norway. While we do agree that this does not provide "general state of knowledge of climate tipping points", and hence have revised our manuscript's terminology, we argue that it provides knowledge about the awareness of the concept, and also that the examples serve as a "control" to see if the knowledge self-assessment is justified by adequate examples.

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).

Response: We disagree with the Reviewer's argument that our data collection method does not allow any broad conclusions about CTP knowledge among the Norwegian public. As outlined above, we devised a scientifically grounded approach to knowledge elicitation among the Norwegian public. We have clarified our assumptions about public knowledge and justified the survey design that avoids asking for a definition or characteristics, but relies on self-assessment (perceptions) of ability with a 'check' for actual ability to recall an example. Other studies, including Bellamy 2023, also assess knowledge ('awareness') without asking participants to provide a definition.

We also argue that separating respondents into these four categories should be appropriate. These kinds of categorizations have some inherent subjectivity, but are a standard approach to characterise the population with distinct categories. We have however adjusted the way we categorise to maximise categorization in favour of knowledge and to remove a differentiation based on familiarity with characteristics (previous distinction between some and good knowledge). This revised categorization likely overestimates the state of knowledge among the public, but accounts for some of the Reviewer's concerns about the nature and interpretation of our data.

1) No knowledge - based on familiarity rating only (know very little, have never heard of climate tipping points) [52%]

2a) Some knowledge but no demonstrated ability to provide an example or unelicited description - based on familiarity rating (categories 1-3), answered NO to question about ability to provide an example OR answered YES to question about ability to provide an example, but did not respond to prompt to provide an example - [25% n=209]

2b) Some knowledge but only demonstrates incorrect knowledge - based on familiarity rating (categories 1-3), answered yes to question about ability to provide an example, but provided a wrong example or wrong unelicited comments [5% n=44]

3) Good knowledge - based on familiarity rating (categories 1-3), answered yes to question about ability to provide an example, provided at least one correct example or unelicited comments with correct characteristics of climate tipping points [13% n=109]

We have modified the text, retained section 4.2 with an expanded explanation of the nature of the data, and placed Table 1 in Supplementary Materials.

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.

Response: We thank the reviewer for pointing out the missing standardised effect sizes in our text. We ran the analysis again in R and found a standardised effect size observed to be 0.08 standard deviation units - which indeed shows that the effect observed in our data can be interpreted as very small. We have now included this in the manuscript (lines 499-502) and it is discussed (lines 569 - 571). We have also reflected this in the abstract (line 22) and conclusion (line 605).

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).

Response: This is true, it is well known e.g. from epidemiological studies with large samples that significance should be treated with caution, and statistical significance not by necessity implies relevant differences. We have now discussed this in the text, in order not to inflate our findings or the significance of our results (lines 569-571).

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.

Response: We no longer include the fatalism results in this manuscript because it is not central to our research questions, but thank the reviewer for pointing out the instances mentioned in the test. We have removed the text that refers to 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).

Response: We have clarified which questions were used to measure climate change risk perceptions in section 3.1. We have also provided a reliability estimate for our CCRP scale: "the calculated value for Cronbach's alpha was $\alpha = 0.897$ indicating strong internal consistency reliability among the questions measuring CCRP, suggesting that the questions are highly correlated and likely measure the same underlying construct effectively". (lines 279-285)

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?).

Response: We have rerun the statistical analyses and have rewritten our text to include the ANOVAs and ANCOVA to better communicate our step-by-step tests run in R. All the F, t and p values are shown in the text. We include one of the paired-sample t-tests clearly as follow-up analyses but have removed the specific t-test the reviewer points to as a repetition of the ANCOVA analysis (lines 482 - 517).

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

Response: We have reworded the sentence to reflect the limitations of statistical analyses more carefully. The sentence now reads "Some studies have shown clear correlations between instruction, information and knowledge on the one hand and climate risk perception on the other" (lines 107-109).

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…).

Response: We acknowledge the importance of utilising probabilistic (random) sampling methods for the generalisability of our findings to the broader population. As highlighted in your comment, non-probabilistic sampling has implications for the representativeness of the sample, particularly concerning characteristics excluded from the quota plan. However, quota sampling is a common and widely accepted approach in large-scale surveys - quota samples represent the population of interest in a real sense. The traits included in the sampling are key for a successful study. Our quota plan included major characteristics that reflect the demographic makeup of the Norwegian population. As stated above, to ensure the quality of our approach, we worked with one of Norway's best-known polling services with decades of experience. Nevertheless, we have duly noted the limitations associated with our sampling approach in the manuscript, especially when it comes to minorities in the population. We have added an extra sentence to emphasise this limitation (lines 328-330).

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.

Response: When rerunning our analysis, we also had our results verified by a third party.

3. Technical corrections

1) Milfont (2012) is missing in the reference list
Response: Milfornt (2012) is now included in the reference list.

2) Line 167: “the likelihood of triggering climate tipping points is “dangerously close”…” This sentence should be rephrased.
Response: We have rephrased this sentence to “the risk of triggering some climate tipping points may be “dangerously close”.

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.
Response: The intention of this example was to highlight the range of timescales relevant to tipping elements, however, we recognise that our paper focuses on climate tipping elements and therefore using this particular example may cause confusion. We have therefore removed this example from the manuscript and instead focused on timescales relevant to climate tipping elements.

4) Lines 119-125: This paragraph could be removed to reduce the length of the article.
Response: We have removed part of this paragraph; however, we have retained the first two sentences in order to highlight how the scholarship on climate tipping points differs from general climate change scholarship historically.

5) Lines 63-68: This part could be shortened – some of this information is repeated in section 2.4.
Response: This section has been shortened in order to limit repetition in the article (lines 68-71).

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…”
Response: Thanks to the reviewer for pointing this out, we have changed this to “Effect of Information about Climate Tipping Points on Level of Concern for Climate Change” to be in line with our article title.

7) The results of F-tests and t-test should be reported in a consistent format throughout the manuscript (e.g., APA format).
Response: We have changed the reporting of the results to match the APA formatting style.

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.

Response: We have restructured the supplementary material and cut back items that are not relevant to the article.