Environmental Psychology Open

Peer review process documentation

of the article

Characteristics of popular climate policies: A replication study after more than ten years of climate activism

by

Gundula Thiele, Ulrike Senftleben, Stefan Scherbaum, Caroline Surrey

published in

Environmental Psychology Open

DOI: https://doi.org/10.69805/epo.v30.a65

Handling editor: Florian Lange

Environmental Psychology Open

Review Round 1

Comments by the editor

1. Please make sure that your revised submission is in line with our manuscript template. I noticed some deviations with regard to, for example, margin size and reference formatting.
2. If you would like to include supplementary information, please upload them as a supplementary material file, that is, do not include an appendix in your manuscript.
3. Interactions like the one between coerciveness and behavioural cost could easily be produced by ceiling effects attenuating the effect of the other independent variable when costs are low. Please discuss whether this explanation is logically distinct from the one you prefer, whether it is likely, and whether it is problematic.
4. I find it difficult to wrap my head around the phrasing of Hypothesis 6. I think it is supposed to predict a three-way interaction (please check if this gets clear in the preregistration and results section). But the way it is phrased, it seems to describe an additive effect of two two-way interactions. If H4 is true and H5 is true, than H6 would be true by inference. Please check and sharpen the phrasing if applicable. Also, I missed a rationale on why there should be a three-way interaction.
5. I understand why your text focuses on the replication value of the study you chose to replicate and the quality of your study as a replication. However, I think should also include some critical thought on the general limitations of the chosen research approach, limitations that affect the validity of the original and your study alike. Do such scenario studies allow for the internally and externally valid study of the determinants of policy support? Could the design produce artifacts? Is the outcome validly assessed? Special attention might be given to the fact that your reasoning often relates to general categories (e.g., push measures), but then in the operationalization, you chose very few and very specific instantiations of these general categories. Does such operationalization adequately represent the category of interest? Please address such issues in the discussion section.

Comments by Reviewer C

In their manuscript, the authors present a direct replication of an experiment by de Groot and Schuitema (2012) which tested the effect of policy characteristics (pull vs. push, high vs. low costs) and social norms on policy acceptance. They could replicate most findings, although not all.

I appreciate the authors' efforts to replicate the experiment by de Groot and Schuitema (2012), especially since knowledge and attitudes about climate change and policies have changed since the time the original experiment was conducted. The replication was preregistered and the authors share their data. I think this replication is important and I appreciate the author's efforts in conducting this replication.

I do have some comments, mainly regarding the Introduction and Discussion:

Environmental Psychology Open

I like the motivation of this replication, especially since time has passed since the original experiment was conducted and public attitudes and knowledge about climate change and action as well as the political and environmental reality has changed at least to some degree since before 2012. However, also the scientific literature has changed since then: There have been experiments published on the effects of the studied factors since 2012. Although the authors mention this in the Discussion, I think this should be (also) part of the Introduction. It does make a difference if studies (or basically conceptual replications) since then have found consistent or diverging results. This does not affect the scientific value of the present replication, but mentioning research which could inform the question whether the findings replicate nowadays does seem important.

I would appreciate it if the authors could provide a bit more of a rational for the main and interaction effects. The main effects are relatively straightforward, whereas the interaction effects are not so obvious to me and, in my opinion, require some explanation. Also, the authors should explicitly state which hypothesis was tested in the original study and which was proposed by the authors of the manuscript. This can be inferred from the remaining text, but I think it should be explicitly noted in the list of hypotheses in the Introduction. I think especially the three-way interaction requires some explanation as it does not seem obvious.

I appreciate that the authors used a larger sample than the original study. However, I think the authors should acknowledge somewhere that their study was run with a student sample and the original study with a general-population sample. This difference alone could explain differences between the original study and the replication. For instance, it does not seem surprising that climate policies receive relatively high acceptance if they are tested with a young, educated, and mostly female sample. Therefore, I would also be hesitant to interpret the high acceptance score as a societal trend since 2012.

This is just a suggestion, but I think Figure 1 would be more easily to grasp if it showed. A 2 x 2 matrix instead of row with 4 plots. When reading the Results section (especially with regard to the interaction effects), it was challenging to determine which plots I needed to compare. I think a 2 x 2 matrix would make it much easier to read the figure.

In the Discussion, the authors write that their results are in line with research on policy packaging showing that packaging increases acceptance. I do find the research on policy packaging interesting and agree that packaging represents a promising approach to raise acceptance, but I am not sure how the results in this manuscript are in line with these findings. The authors always test only one policy at a time and simply vary policy characteristics (and social norms). I think the authors should either elaborate on that point or cut the argument.

Minor comments:

Page 5, “age: 18-18”: there seems to be a typo here. I guess the authors mean 28
Page 8, “all pairwise comparison”: Please define which pairwise the authors are referring to

Environmental Psychology Open

Page 8, “but were not independent of the behavioural costs”: I was not sure if I understood this sentence (but that might also relate to the comment above). Do I understand correctly that there was an effect of behavioural costs even for pull measures?
Page 8, “suggesting that all policies were perceived as somewhat acceptable”: I am not sure whether one can infer that *all* policies were perceived as somewhat acceptable when examining the overall mean. I would rather change the wording to, for instance, “were generally perceived as…”.

Comments by Reviewer D

Thank you for the opportunity to review this interesting replication study. It replicates a study from Groot and Schuitema from over ten years ago that aimed to experimentally manipulate policies’ target behaviors, their coerciveness and the related social norm. The study is straightforward, sound and timely. However, a few conceptual, methodological, and presentational issues should be addressed before publication. I first list major comments before turning to minor ones:

The study seems to use acceptance and acceptability interchangeably. This might warrant an explicit statement. In general, a short elaboration on this central outcome in the study would help inform the reader. In this context this is a helpful reference: https://doi.org/10.1080/14693062.2019.1611534

When presenting the literature on acceptance determinants, some important publications are omitted, that the authors should engage with - particularly in the discussion. https://doi.org/10.1038/s41558-022-01297-6; doi: 10.3390/su12062404; doi: 10.1515/roe-2021-0046
Particularly, fairness and effectiveness perceptions as important determinants need to be engaged with to convey a fuller picture of what the current literature has to say about acceptance determinants.
In line 78, it seems all mobility behaviors are high in monetary costs. However, the central mobility behavior change in the study is the switch from car use to public transport, which is arguably less monetarily expensive. The authors should sharpen the description of the general literature regarding high-cost behavior and how it applies to the specific context of the measures presented in the study.
This also leads to some inferences going beyond what the study can deliver in the impact presentation and the impact statement. If higher monetary expenses with regards to high-cost behavior are presented, the authors should make explicit how the study’s results regarding the switch to “cheaper” public transport warrant such conclusions.
In paragraph 3 the authors reason, why it would be pertinent to replicate the earlier findings, as some of the effects might now be different. From the 3 initial main effects, however, only 2 are addressed (norms and target behavior). Was there deliberately no reasoning regarding the third factor, i.e. coerciveness?
In line 106, the authors mention data exclusions but none are discussed later in the paper. Were there any exclusion criteria that weren’t met and therefore there were no exclusions?
I am sorry, if I misunderstood something, but why is the norm intervention a between-subjects design (line 126)? From how I read table 1, different norm levels were presented to a participant, which would also make it a within-subjects design. If this was not the case, please make it more explicit when presenting the design. For instance, also lines 141-143 are not clear in this regard.

To me, lines 166-167 would indicate a partial rejection of H4, as H4 specifically assumes that "pull measures are rated as acceptable regardless of the behavioral cost." Please discuss.
Similarly regarding H5 (lines 168-170): Isn't H5 only partially contradicted, since the acceptance of pull measures was expected not to differ depending on norm levels?
In my opinion, the moderation results would benefit greatly from figures that visually illustrate the direction of the interactions. (One easy solution could be the Excel files from https://www.jeremydawson.co.uk/)
For example, Figure 1 suggests that pull measures are even more popular when social norms are strong, even if the difference isn't statistically significant. Is this the case? Please discuss.
I am missing control variables, although very common ones are mentioned in lines 73-74. Why weren't they included? If there were no control variables included in the study, please provide a rationale and at least present this as a limitation in the discussion.
In some parts, the discussion could be tightened to be easier to follow, e.g. in lines 194-202. First, effectiveness as a concept is engaged with for the first time. Then the statement about the costliness of climate-friendly behaviors follows without guiding the reader, why these two are presented jointly. Then in line 197 the authors write "This could be mitigated" without clarifying what this refers to - effectiveness or costliness or both? If both, in what way? Which conclusions from the study's results exactly can help mitigate what?
At the end of that paragraph, the authors allude to how some policy aspect combinations can increase acceptance and effectiveness. But the one concrete combination they mention is with regards to low-cost behaviors, which in the literature have been found to be generally less effective in combatting climate change. Please discuss this tension.
This is also relevant for the statement in lines 216-218. There, the authors unfortunately obscure which aspects within one policy and which combination of policies with different features are the ones increasing policy support? It would be very interesting for the reader to see in what way the concrete suggestions relate to the important concepts of actual and perceived effectiveness.
In general, the discussion is missing engagement with the limitations that small student samples bring. Particularly when discussing the rejection of H5, the student sample might inspire explanations as to why social norm manipulation had no effect (students in Germany are maybe more progressive and tend to support climate measures regardless of general norm?)
In general, the practical implications beyond littering and mobility are somewhat hidden behind very general statements (e.g. lines 216-218). While acknowledging generalizability challenges, please guide the reader with one or two examples of other climate policies and how the results inform decision-makers in increasing support.
How do the authors reconcile the fact that acceptance is precisely a condition for policy implementation, which is in turn a prerequisite for effectiveness - as mentioned in the introduction - with the last statement (lines 218-219) that implies that effectiveness is more important than acceptance?

Further minor comments:

For the German reader, the translation of climate policies to Klimagesetze is not the best fit in my opinion. A better, more neutral translation would be "politische Klimaschutzmaßnahmen". To the German reader, Gesetze is too specific - especially with regards to the climate policy measures used in the experiment. (More broadly, the German abstract and impact statement could benefit from a careful review by a native speaker, as, to me, the text

Environmental Psychology Open

appears to follow the English source word for word, resulting in overly literal phrasing and an accumulation of subordinate clauses that read a bit unnaturally in German.)

In lines 64-66, complex and global solutions are called for and then these are exemplified with policies targeting individuals in a society and ones connecting these and local efforts with large-scale changes. My questions are: In what way are these examples global? In light of the arguably simple, national-level policy measures presented in the study, why is this reference to complex and global solutions relevant?

To make it even more comprehensible, the paragraph on design and procedure would benefit from using the terms policy scenario and policy measure consistently.

When describing the sample, there may have been a mistake in the age range. It probably wasn't 18-18. Also, the gender distribution in percent would be easier to assess.

Environmental Psychology Open

Author Response to the Reviewers Round 1

Dear Prof Lange,

Thank you and the reviewers for your thorough evaluation of our manuscript. We greatly value the feedback provided, which we believe has significantly enhanced our work.

Below, you will find our detailed responses, addressing each point in the order presented in the review. Additionally, we send you the revised manuscript and supplement with tracked changes. Please note that the affiliation of the first author has changed due to a new position.

Sincerely,

The authors

Editor comments

Please make sure that your revised submission is in line with our manuscript template. I noticed some deviations with regard to, for example, margin size and reference formatting.
If you would like to include supplementary information, please upload them as a supplementary material file, that is, do not include an appendix in your manuscript.

Interactions like the one between coerciveness and behavioural cost could easily be produced by ceiling effects attenuating the effect of the other independent variable when costs are low. Please discuss whether this explanation is logically distinct from the one you prefer, whether it is likely, and whether it is problematic.

> **Response:** As our post-hoc tests indicate, there is still a significant difference in acceptability for the coerciveness conditions within the condition low-cost behaviour. But as policies targeting low-cost behaviours are generally evaluated as very acceptable, this might amplify the interaction effect. We would argue that this explains part of the effect, but not fully. We added the following limitation in the discussion: „Note that the interaction effect in our data might have been amplified by the generally higher acceptability of low-costs behaviours; however, post hoc tests still revealed a significant difference between push and pull measures for low-cost behaviours.“
> Furthermore, a (partial) ceiling effect would be relatively unproblematic as the implication would be that our scale might not be able to represent the variance of responses and that policies targeting low-cost behaviours will be rated as more acceptable, even when they are more coercive (push). However, we refrain from adding this to the discussion to not inflate the practical implication of this ceiling effect.

I find it difficult to wrap my head around the phrasing of Hypothesis 6. I think it is supposed to predict a three-way interaction (please check if this gets clear in the preregistration and results section). But the way it is phrased, it seems to describe an additive effect of two two-way interactions. If H4 is true and H5 is true, than H6 would be

true by inference. Please check and sharpen the phrasing if applicable. Also, I missed a rationale on why there should be a three-way interaction.

> **Response:** We indeed hypothesized a three-way interaction and rephrased the hypothesis as follows: „Push & high-cost & weak norm: Social norms moderate the interaction effect of behavioural costs and coerciveness so that a strong social norm attenuates the reduced acceptability for policies that are push measures and target high-cost behaviours. A weak social norm on the other hand may amplify these negative effects (exploratory)".
> We also added more reasoning for the hypothesis at the end of our introduction, which reads as follows: „ Additionally, we wanted to explore the three-way interaction between the factors to see whether social norms had an influence on the interaction between coerciveness and behavioural cost. We reasoned that if there is an interaction between coerciveness and behavioural costs, then policies might be much less favourable when they are "negative" on both factors—that is, when they rely on coercion through punishment and entail high behavioural costs. The resulting decline in acceptability could potentially be mitigated by a strong social norm. We therefore aimed to examine this interaction more closely and added it as exploratory to the replication hypotheses: (...)"

I understand why your text focuses on the replication value of the study you chose to replicate and the quality of your study as a replication. However, I think should also include some critical thought on the general limitations of the chosen research approach, limitations that affect the validity of the original and your study alike. Do such scenario studies allow for the internally and externally valid study of the determinants of policy support? Could the design produce artifacts? Is the outcome validly assessed? Special attention might be given to the fact that your reasoning often relates to general categories (e.g., push measures), but then in the operationalization, you chose very few and very specific instantiations of these general categories. Does such operationalization adequately represent the category of interest? Please address such issues in the discussion section.

> **Response:** Thank you for this input. While we could not elaborate on each of your points in detail while keeping the discussion concise, we added a section on limitations and addressed the points that we assumed to be most relevant for future research. We discussed our sample, which consisted largely of individuals younger than 30 years who identified as women. We also addressed our strategy of assessing acceptability with only one item and our operationalization of each condition combination with only one scenario. This operationalization limits study validity, allows subjects to see through and react to our design, and reduces the generalizability.

Reviewer C

In their manuscript, the authors present a direct replication of an experiment by de Groot and Schuitema (2012) which tested the effect of policy characteristics (pull vs. push, high vs. low costs) and social norms on policy acceptance. They could replicate most findings, although not all.

I appreciate the authors' efforts to replicate the experiment by de Groot and Schuitema (2012), especially since knowledge and attitudes about climate change and policies have changed since the time the original experiment was

Environmental Psychology Open

conducted. The replication was preregistered and the authors share their data. I think this replication is important and I appreciate the author's efforts in conducting this replication.

I do have some comments, mainly regarding the Introduction and Discussion:

I like the motivation of this replication, especially since time has passed since the original experiment was conducted and public attitudes and knowledge about climate change and action as well as the political and environmental reality has changed at least to some degree since before 2012. However, also the scientific literature has changed since then: There have been experiments published on the effects of the studied factors since 2012. Although the authors mention this in the Discussion, I think this should be (also) part of the Introduction. It does make a difference if studies (or basically conceptual replications) since then have found consistent or diverging results. This does not affect the scientific value of the present replication, but mentioning research which could inform the question whether the findings replicate nowadays does seem important.

> **Response:** We added references to more recent publications to update the background of our study. However, to stay concise in the introduction, we decided not to discuss these newer studies extensively, since they add converging evidence.

I would appreciate it if the authors could provide a bit more of a rational for the main and interaction effects. The main effects are relatively straightforward, whereas the interaction effects are not so obvious to me and, in my opinion, require some explanation. Also, the authors should explicitly state which hypothesis was tested in the original study and which was proposed by the authors of the manuscript. This can be inferred from the remaining text, but I think it should be explicitly noted in the list of hypotheses in the Introduction. I think especially the three-way interaction requires some explanation as it does not seem obvious.

> **Response:** We provided more rationale for the interactions, especially the three-way interaction in hypothesis 6. We also made it more explicit which hypotheses are replication (1-5) and which one is explorative (6). The end of the introduction now provides more rationale for the hypotheses and reads as follows: „This included testing the main effects for each factor separately and then looking at interactions. As De Groot & Schuitema (2012) argued, push measures enforce greater behavioural restrictions, making high-cost behaviours feel even more constraining and therefore less acceptable than low-cost behaviours (Steg et al., 2005). Pull measures, in contrast, are generally acceptable regardless of the behavioural cost. But pull measures are not always effective enough for climate change mitigation (Gärling, 2007). But a strong social norm might signal the necessity and effectiveness of a policy. Since effectiveness is also a determinant of policy acceptability (Bergquist et al., 2022), a strong social norm might mitigate the detrimental effect of push measures on acceptability, while a weak norm might amplify it. Additionally, we wanted to explore the three-way interaction between the factors to see whether social norms had an influence on the interaction between coerciveness and behavioural cost. We reasoned that if there is an interaction between coerciveness and behavioural costs, then policies might be much less favourable when they are "negative" on both factors—that is, when they rely on coercion through punishment and entail high behavioural costs. The resulting decline in acceptability could potentially be mitigated by a strong

social norm. We therefore aimed to examine this interaction more closely and added it as exploratory to the replication hypotheses: (...)“

I appreciate that the authors used a larger sample than the original study. However, I think the authors should acknowledge somewhere that their study was run with a student sample and the original study with a general-population sample. This difference alone could explain differences between the original study and the replication. For instance, it does not seem surprising that climate policies receive relatively high acceptance if they are tested with a young, educated, and mostly female sample. Therefore, I would also be hesitant to interpret the high acceptance score as a societal trend since 2012.

Response: We thank the reviewer for raising this important concern. We agree that our sample is not representative of the general population of Germany and might therefore not show general societal trends. We added this as a limitation to our discussion, which reads as follows: „First, while our sample has a broad age range, about two thirds of participants were younger than 30 years. Additionally, two thirds of participants identified as female. Previous research indicates that women* and older individuals generally show higher acceptance of environmental policies (Lehrer et al., 2025). This might have influenced the outcomes for acceptability in our sample in both directions. Future studies should therefore focus on collecting data from representative samples or aggregating diverse samples from multiple studies.“

This is just a suggestion, but I think Figure 1 would be more easily to grasp if it showed. A 2 x 2 matrix instead of row with 4 plots. When reading the Results section (especially with regard to the interaction effects), it was challenging to determine which plots I needed to compare. I think a 2 x 2 matrix would make it much easier to read the figure.

Response: We edited figure 1 to show the combinations of each pair of factors in a separate plot for easier comparison. In our opinion, this makes it easier to grasp the relevant main and interaction effects. Please not that the new figure is a result of both your feedback and the feedback of another reviewer, which is why we did not directly implement your suggestion of a 2x2 matrix.

In the Discussion, the authors write that their results are in line with research on policy packaging showing that packaging increases acceptance. I do find the research on policy packaging interesting and agree that packaging represents a promising approach to raise acceptance, but I am not sure how the results in this manuscript are in line with these findings. The authors always test only one policy at a time and simply vary policy characteristics (and social norms). I think the authors should either elaborate on that point or cut the argument.

Response: This is correct, what we wanted to express is that combining favourable with unfavourable features of policies raises acceptability, but combining features of multiple policies is different from combining these features into one single policy. We changed the framing of our argument to present policy packaging as an alternative when combining different features into one policy is not feasible: “For example, we found that high-cost behaviours were rated more acceptable when targeted with pull in comparison to push measures. This suggests that acceptability for adaptation to high-cost behaviours

Environmental Psychology Open

might be higher if we create incentive structures and make the pro-environmental behaviour easier than its alternatives. Combining different features into one single policy might not always be feasible, but similar results were found for policy packaging, where combining different policy aspects might raise acceptability (Givoni, 2014; S. Wang et al., 2020)."

Minor comments:

Page 5, "age: 18-18": there seems to be a typo here. I guess the authors mean 28

Response: Thank you for pointing out this error. This seemed to be an issue with the Rmarkdown code. We corrected it, the age range is 18 to 82 years.

Page 8, "all pairwise comparison": Please define which pairwise the authors are referring to

Response: We now specify the pairwise comparisons as follows: „Post-hoc testing showed significant differences in acceptability in all pairwise comparisons, that is in pull vs. push measures separately for high- and low-cost behaviours, and in high- vs. low-cost behaviours separately for pull and push policies (all *p*s < .05, see Table A3 in the supplement)."

Page 8, "but were not independent of the behavioural costs": I was not sure if I understood this sentence (but that might also relate to the comment above). Do I understand correctly that there was an effect of behavioural costs even for pull measures?

Response: Yes, you understood this correctly and based on your and Reviewer D's feedback, we changed the parts of the results and discussion section that concern H4 to make this clearer.

Page 8, "suggesting that all policies were perceived as somewhat acceptable": I am not sure whether one can infer that *all* policies were perceived as somewhat acceptable when examining the overall mean. I would rather change the wording to, for instance, "were generally perceived as...".

Response: We see your point and changed the wording to "We found a mean acceptability rating of M = 5.51 (Md = 6, SD = 1.81) across all experimental conditions, suggesting that the presented policies were generally perceived as somewhat acceptable (see Figure 1 and Table A1)."

Reviewer D

Thank you for the opportunity to review this interesting replication study. It replicates a study from Groot and Schuitema from over ten years ago that aimed to experimentally manipulate policies' target behaviors, their coerciveness and the related social norm. The study is straightforward, sound and timely. However, a few conceptual, methodological, and presentational issues should be addressed before publication. I first list major comments before turning to minor ones:

Environmental Psychology Open

The study seems to use acceptance and acceptability interchangeably. This might warrant an explicit statement. In general, a short elaboration on this central outcome in the study would help inform the reader. In this context this is a helpful reference: https://doi.org/10.1080/14693062.2019.1611534

> **Response:** Thank you for pointing this out! We changed our wording so that we use acceptability in each instance and added a footnote in the introduction to make this clear and to guide interested readers to the publication you suggested.

When presenting the literature on acceptance determinants, some important publications are omitted, that the authors should engage with - particularly in the discussion. https://doi.org/10.1038/s41558-022-01297-6; doi: 10.3390/su12062404; doi: 10.1515/roe-2021-0046
Particularly, fairness and effectiveness perceptions as important determinants need to be engaged with to convey a fuller picture of what the current literature has to say about acceptance determinants.

> **Response:** We added your suggestions about fairness and effectiveness in the introduction: "Aside from individual factors, the specific aspects of policies also influence their acceptability, such as fairness and effectiveness of climate change mitigation (Bergquist et al., 2022)."

In line 78, it seems all mobility behaviors are high in monetary costs. However, the central mobility behavior change in the study is the switch from car use to public transport, which is arguably less monetarily expensive. The authors should sharpen the description of the general literature regarding high-cost behavior and how it applies to the specific context of the measures presented in the study.
This also leads to some inferences going beyond what the study can deliver in the impact presentation and the impact statement. If higher monetary expenses with regards to high-cost behavior are presented, the authors should make explicit how the study's results regarding the switch to "cheaper" public transport warrant such conclusions.

> **Response:** While we see the reviewer's point, in our study we focussed on behavioural costs for individuals. These include monetary costs, but also costs like time or effort that is associated with a certain behaviour. While taking public transportation might be cheaper in terms of monetary costs, it might be more effortful and time-consuming than taking the car for an individual, depending on the specific context. We therefore view it as a high-cost behaviour. We changed some wordings in the introduction to make this clearer: "They varied the behaviours the policies targeted in regard to the costs of behavioural adaptation, which includes monetary costs as well as effort. Policies either targeted mobility behaviours, which w considered high-cost with respect to changing them, or they targeted littering, which is considered low-cost (e.g., Rau et al,, 2024, Wang, Gao, & Li, 2024)."

In paragraph 3 the authors reason, why it would be pertinent to replicate the earlier findings, as some of the effects might now be different. From the 3 initial main effects, however, only 2 are addressed (norms and target behavior). Was there deliberately no reasoning regarding the third factor, i.e. coerciveness?

Environmental Psychology Open

Response: This was not a deliberate decision with regards to the content but more to the word limit for our initial submission. We added more reasoning for the coerciveness main effect and the interaction effects at the end of the introduction: “”

In line 106, the authors mention data exclusions but none are discussed later in the paper. Were there any exclusion criteria that weren’t met and therefore there were no exclusions?

Response: We did not exclude any of the data aside from incomplete datasets. We added this information in the participants section.

I am sorry, if I misunderstood something, but why is the norm intervention a between-subjects design (line 126)? From how I read table 1, different norm levels were presented to a participant, which would also make it a within-subjects design. If this was not the case, please make it more explicit when presenting the design. For instance, also lines 141-143 are not clear in this regard.

Response: We tried to make it clearer that based on their assigned group, one participant saw either the information for the weak norm or the strong norm condition in all four scenarios. We changed the passage in section 2.2 as follows: “Each policy was accompanied by a statement that either a minority (weak social norm) or a majority (strong social norm) of pilot participants supported the policy, depending on the group each participant was assigned to. Participants therefore saw either the minority or the majority information for all four scenarios.”

To me, lines 166-167 would indicate a partial rejection of H4, as H4 specifically assumes that “pull measures are rated as acceptable regardless of the behavioral cost.” Please discuss.
Similarly regarding H5 (lines 168-170): Isn’t H5 only partially contradicted, since the acceptance of pull measures was expected not to differ depending on norm levels?

Response: This is correct and we edited parts of the results and discussion section to make this clearer.

In my opinion, the moderation results would benefit greatly from figures that visually illustrate the direction of the interactions. (One easy solution could be the Excel files from https://www.jeremydawson.co.uk/)
For example, Figure 1 suggests that pull measures are even more popular when social norms are strong, even if the difference isn't statistically significant. Is this the case? Please discuss.

Response: We worked on figure 1 to better illustrate the differences in acceptability ratings, including line plots as in the solution you suggested. You are correct to point out that there is a visible difference between pull measures in the strong vs. the weak norm condition. The interaction of social norm and coerciveness was not significant ($p = .17$). The difference seen in the figure is caused by the main effect of social norm, suggesting a global effect that is independent of policy coerciveness.

I am missing control variables, although very common ones are mentioned in lines 73-74. Why weren't they included? If there were no control variables included in the study, please provide a rationale and at least present this as a limitation in the discussion.

> **Response:** Participants could voluntarily enter their age and gender identity. Based on your inquiry, we performed a mixed ANCOVA using age and gender as covariates. The results for our hypotheses did not change qualitatively. Gender had no significant main or interaction effect, age showed no significant main effect, but significant interactions with behavioural cost and coerciveness. However, it should be noted that to do this, we had to perform a parametric ANCOVA, since the MANOVA.RM package does not have an option to include covariates. For our analyses, we decided to use MANOVA.RM to calculate a nonparametric ANOVA since not all assumptions were met and our groups were unbalanced. Therefore, we consider the results of this explorative ANCOVA to be less reliable and decided to not include them in the paper. However, they are available in the public code, and the control variables are included in the open data for further use.

In some parts, the discussion could be tightened to be easier to follow, e.g. in lines 194-202. First, effectiveness as a concept is engaged with for the first time. Then the statement about the costliness of climate-friendly behaviors follows without guiding the reader, why these two are presented jointly. Then in line 197 the authors write "This could be mitigated" without clarifying what this refers to - effectiveness or costliness or both? If both, in what way? Which conclusions from the study's results exactly can help mitigate what?

> **Response:** We see that this paragraph changes topics a bit and might therefore be a bit disrupting. We discuss effectiveness here, because when mitigating climate change is our central goal, then the the effectiveness of a policy to contribute to that is the most important criterion for that policy. And while we said in the end of the previous paragraph that policies with low coerciveness, low behavioural costs and a strong supportive norm might be the most likely to be implemented in terms of their acceptability, we wanted to limit this statement here by contrasting acceptability with effectiveness and how those goals are not always aligned. For example, pull measures, which were rated as more acceptable in our study, are found to be less effective than pull measures in some cases (Gärling, 2007). To make it clearer that we are contrasting these two goals, we changed the sentence to "This conflict between acceptability and effectiveness could be mitigated by combining different attributes of policies, as our interaction effect suggests".

At the end of that paragraph, the authors allude to how some policy aspect combinations can increase acceptance and effectiveness. But the one concrete combination they mention is with regards to low-cost behaviors, which in the literature have been found to be generally less effective in combatting climate change. Please discuss this tension.

> **Response:** You are correct, this might not be the best example for our point here. We changed the example to "For example, we found that high-cost behaviours were rated more acceptable when targeted with pull in comparison to push measures. This suggests that acceptability for adaptation to high-cost

Environmental Psychology Open

behaviours might be higher if we create incentive structures and make the pro-environmental behaviour easier than its alternatives."

This is also relevant for the statement in lines 216-218. There, the authors unfortunately obscure which aspects within one policy and which combination of policies with different features are the ones increasing policy support? It would be very interesting for the reader to see in what way the concrete suggestions relate to the important concepts of actual and perceived effectiveness.

Response: Thank you for pointing this out. We added a sentence clarifying the specific combination of aspects. Since we did not assess actual and perceived effectiveness, we do not wish to make inferences on how our findings relate to those concepts, but we do think it would be interesting for future studies. The passage now reads as follows: "Combining different aspects within one policy or combining policies with different features might therefore alleviate some resistance towards environmental policy making. This could mean implementing a pull policy that targets a high-cost behaviour or lowering the behavioural costs for a push policy, according to our findings. But, while acceptability is important, the effectiveness of climate protection measures should be the central concern. Future research should therefore investigate how to raise acceptance for policies that effectively mitigate climate change."

In general, the discussion is missing engagement with the limitations that small student samples bring. Particularly when discussing the rejection of H5, the student sample might inspire explanations as to why social norm manipulation had no effect (students in Germany are maybe more progressive and tend to support climate measures regardless of general norm?)

Response: Based on your comment and the comment of Reviewer C, we added a section on the sample to our discussion. Our sample has a broad age range, but two thirds of the sample are younger than 30 years, which might have influenced acceptability ratings, although a study in the German context found that acceptability is generally higher in older groups (Lehrer et al., 2025). We cannot discuss whether this is a student sample and how this might influence the results, because we did not assess the occupations of our participants and therefore do not know whether this is a student sample. While our recruiting through the experimental server at the University makes this more likely, we do not wish to assume.

In general, the practical implications beyond littering and mobility are somewhat hidden behind very general statements (e.g. lines 216-218). While acknowledging generalizability challenges, please guide the reader with one or two examples of other climate policies and how the results inform decision-makers in increasing support.

Response: We wanted to refrain from giving strong recommendations for policy-making as this is only one study, but we see that there are more areas than mobility and littering where policies can shape behaviour to be more pro-environmental. We therefore added the following passage to our discussion: "Using only one scenario per combination of conditions also limits generalisability, while environmental policies are not limited to mobility and littering, for example the German Environmental Agency recommended reducing subsidies for meat pricing (German Environmental Agency, 2025), which would

Environmental Psychology Open

combine a high-cost behaviour with a pull approach on behavioural adaptation. Using more than one scenario per combination of conditions would therefore improve methodological limitations as well as generalizability."

How do the authors reconcile the fact that acceptance is precisely a condition for policy implementation, which is in turn a prerequisite for effectiveness - as mentioned in the introduction - with the last statement (lines 218-219) that implies that effectiveness is more important than acceptance?

Response: It is true that policies can only be practically effective if they are implemented and that implementation is influenced by policy acceptability. But the anticipated effectiveness of a policy (that is not yet implemented) is also a contributor for its implementation. With our statement in the discussion, we tried to express that anticipated effectiveness and acceptability are not always aligned, which might warrant a combination of different attributes to raise the acceptability of otherwise effective policies.

Further minor comments:

For the German reader, the translation of climate policies to Klimagesetze is not the best fit in my opinion. A better, more neutral translation would be "politische Klimaschutzmaßnahmen". To the German reader, Gesetze is too specific - especially with regards to the climate policy measures used in the experiment. (More broadly, the German abstract and impact statement could benefit from a careful review by a native speaker, as, to me, the text appears to follow the English source word for word, resulting in overly literal phrasing and an accumulation of subordinate clauses that read a bit unnaturally in German.)

Response: As you inferred, we tried to keep the German Abstract as close to the English one as possible, but we acknowledge that this makes the abstract wordy. We reviewed the German Abstact and changed the wording to „politische Klimaschutzmaßnahmen" in the first instance and to „Klimaschutzmaßnahmen" in later instances, to be in line with the abstract word limit. The updated Abstract is as follows:

„Politische Klimaschutzmaßnahmen stellen ein zentrales Instrument zur Begrenzung des Klimawandels dar. Die Implementierung solcher Klimaschutzmaßnahmen hängt jedoch auch von ihrer Akzeptanz in der Bevölkerung ab. De Groot & Schuitema (2012) zeigten, dass die Akzeptanz durch den ausgeübten Zwang, die Verhaltenskosten der Maßnahmen sowie soziale Normen beeinflusst wird. Hier replizieren wir diese Studie in einer deutschen Stichprobe. Jede:r Teilnehmende (N = 193) las vier Maßnahmenszenarios, die unterschiedliche Grade an Zwang (push- und pull-Maßnahmen) ausübten und auf Verhalten mit geringen oder hohen Verhaltenskosten (Vermüllung und Mobilität) fokussierten. Zusätzlich erhielten Teilnehmende die Information, dass das jeweilige Gesetz von einer schwachen oder starken sozialen Norm unterstützt werde (between-subjects). Wir konnten die Haupteffekte replizieren, dass Akzeptanz bei Pull-Maßnahmen, geringen Verhaltenskosten und starken unterstützenden sozialen Normen gesteigert war. Außerdem konnten wir die Interaktion von Zwang und Verhaltenskosten replizieren, wonach die Akzeptanz bei Push-Maßnahmen dann noch schwächer war, wenn sie Verhalten mit hohen Kosten zum

> Ziel hatten. Die Interaktion von sozialer Norm sowie die Dreifachinteraktion der Prädiktoren waren nicht signifikant. Unsere Ergebnisse deuten darauf hin, dass die untersuchten Faktoren die Akzeptanz von Maßnahmen beeinflussen und sich potentiell durch komplementäre Kombination Widerstand gegenüber einer Maßnahme verringern lassen könnte."

In lines 64-66, complex and global solutions are called for and then these are exemplified with policies targeting individuals in a society and ones connecting these and local efforts with large-scale changes. My questions are: In what way are these examples global? In light of the arguably simple, national-level policy measures presented in the study, why is this reference to complex and global solutions relevant?

> **Response:** With the word global, we wanted to open the dimension of individual behavioural change vs. change in many individuals' behaviour or societal ways of living. The presented policies target individual behaviour, but the policies themselves are global measures. While the behavioural change of an individual only concerns this individual (and potentially their social circle), policies coordinate the behaviour of a large number of individuals living in a society. Of course, there is another level of targeting more central entities like corporations. But since the relevance of individual behaviour change in the context of climate change is an ongoing debate in environmental psychology, we wanted to explicitly focus on measures that transcend individual actors in our study.

To make it even more comprehensible, the paragraph on design and procedure would benefit from using the terms policy scenario and policy measure consistently.

> **Response:** We edited the section so that we use the term policy scenario for instances where we refer to the section of the experiment and the term policy when referring to the policy described in the texts of the policy scenarios.

When describing the sample, there may have been a mistake in the age range. It probably wasn't 18-18. Also, the gender distribution in percent would be easier to assess.

> **Response:** Thank you, this seemed to be an issue with the Rmarkdown code. We corrected it, the age range is 18 to 82 years. As per your request, we changed the gender information for the sample to „66.3% female, 32.1% male, 1.6% neither".

Environmental Psychology Open

Review Round 2

Comments by Reviewer A (corresponds to Reviewer D in Round 1)

The manuscript has much improved since the first version, and nearly all of my comments were adequately addressed.
I really appreciate your update with the robustness and sensitivity analysis. I was really surprised that the point of dichotomisation had such little effect!
There are two comments left however, where I do not think that the authors answers are completely valid.
As the second one in its current form in the manuscript might leave a misleading impression about methods in the literature, I would suggest some clarifications in the manuscript here. But as both issues are nitpicky details that do not change the overall interpretation of the results, I would also be fine with the editor deciding to advance the publication process to not cause any further delays.

Regarding the EK x CO2e interaction:
I don't think your argument against testing the EK x CO2e interaction is completely valid.
The problem here is, that the direction of the effect of EK is potentially different depending on the differences in CO2e:
A person with high knowledge about impacts should not only know on which days choosing the sustainable choice matters.
They should also know on which days it does not matter. And knowing that the difference does not matter today (or that one of the 'sustainable options' is even worse than the omnivore option) should (in theory) decrease the likelihood of choosing sustainable. Hence, whether the effect of EK is positive or negative on the sustainable choice may depend on the day/difference.
When you only test for an EK-main effect on the meal choice, you only test for an average effect over all days - and this averaging over positive and negative effects can result in non existing overall effects. Therefore, not finding a main effect from EK on meal choice might be a consequence of an interaction effect (and not an hindrance to its interpretation).
But probably the reviewer comments of your article are not the best place for this kind of in-depth discussion and I should just finally get around to write an article/comment about this...

As you already tested the effect and reported an absence of it in the answer to my comment, you might consider mentioning it briefly somewhere in the results/discussion of your paper, as I think this is an empirical finding with some potential implications for theory (as argued above).

Regarding the random intercept for weekdays:
Your argument here confuses me, and I also do not think the added half sentence in the manuscript adds sufficient clarity here, yet. While all variance of the CO2-difference is explained by the weekday, not all variance in the weekday will be explained by the CO2-difference. Most obvious example: The ratio of available sustainable meals to omnivore meals differs between days, this probably explains some variance in the likelihood of sus. meals chosen.

Environmental Psychology Open

I thought you could use random effects as a 'catch all' for these other sources and that variance for fixed intercepts would be extracted before modeling of the random effects in LMMs (but this may depend on model specification?) I am no expert in (G)LMMs, so maybe I missed something very basic, but what confuses me with your argument here is, that it seems at odds with your own model:
You use a random effect for participant and still include fixed effects directly tied to the person in your model (EM, EK, socio-demographics...). So this does not seem to be a fundamental problem for the model. So is it a problem for model convergence (or model identification), when you include both? For power? Because you have way more degrees of freedom on the participants than on the days? And could including a random intercept for persons, but not weekdays, potentially bias variance allocation towards effects CO2e diff and/or the interaction?

From a practical standpoint regarding the interpretation of your results this probably makes no difference. But I think for our discipline in general it would be good to be more specific and clear here to avoid a potential spread of misconceptions about methods!

Comments by Reviewer B (corresponds to Reviewer C in Round 1)

Thank you very much for your thoughtful engagement with the comments. I believe all of them have been addressed well, like through the addition of the sensitivity analyses, the clear distinction between the current findings and directions for future research, and the revised wording around the knowledge findings. I also believe that changing implicit assumptions to requirements/conditions improves the clarity of the argument. It was great to read the manuscript again and I would be excited to see it published!

Environmental Psychology Open

Author Response to the Reviewers Round 2

Dear Prof Lange,

Thank you and the reviewers again for your feedback, we hope to have implemented everything to your satisfaction and corrected all formatting issues. Below, you find our detailed responses addressing each point raised by Reviewer A. We have also attached the revised manuscript and supplement with and without tracked changes.

Sincerely,

The authors

Editor comments

Both reviewers from the previous round kindly agreed to have another look at your manuscript and both commended your revisions. In line with their reviews, I think that your manuscript has been further improved and is close to acceptance. However, please have a close look at the comments of Reviewer A who identified several remaining issues. I also attach a version of your submission with minor formatting advice from my side (you do not need to respond to those comments of mine in your response letter).

Response: Thank you for your remarks on the formatting. We would like to respond to one specific point, concerning confidence intervals and effect sizes in the results. Unfortunately, we cannot report effect sizes or confidence intervals for the ANOVA-type statistic used in the parameter-free rmANOVA, as the MANOVA.RM package (Friedrich et al., 2018) does not calculate these. We therefore can only report means and confidence intervals for acceptability ratings in each condition, which are included in Supplement A.

Reviewer A

Thank you for carefully addressing my comments. I appreciate the effort that went into revising the manuscript. Overall, the paper has improved: several conceptual points have been clarified, the discussion is more balanced, and the additional rationale for some hypotheses strengthens the manuscript. I also appreciate that the authors revisited the analyses regarding control variables and tried to improve the visualization of the interaction effects.
Most of my previous concerns have been addressed satisfactorily. I nevertheless have a few remaining comments that I hope the authors will consider before publication. Lines refer to the manuscript with tracked changes.

Remaining comments

1. I commend the effort in revising Figure 1. The added line plots now make the direction of the interaction effects clearer. However, I find the revised figure somewhat confusing because it now contains only three graphs with two-way interactions each instead of the initial graph shown in the previous version that showed all two-way interactions with and integrated the third interaction level from social norm (H6). This had the advantage that it allowed readers to compare all three experimental factors (coerciveness, behavioral cost, and social norm) simultaneously, including the three-way interaction (H6). In contrast, the revised figure appears to focus exclusively on the first-order interactions: Panel A shows the interaction between behavioral cost and social norm, Panel B between coerciveness and social norm, and Panel C between behavioral cost and coerciveness.

1. If this interpretation is correct, it would be helpful to explicitly refer to the respective panels in the Results section (e.g., "...see Fig. 1C" when discussing the coerciveness × behavioral cost interaction). More importantly, I am unsure what purpose Panel A serves. If it is intended to illustrate the three-way interaction (H6), it only presents a first-order interaction and therefore does not seem to correspond to the reported analysis. If it is not intended to represent H6, it would be helpful to clarify why this particular interaction is included and where it is discussed in the text.
2. As a minor point related to the figure, I also found it slightly confusing that coerciveness appears as the x-axis in one panel but only as the legend in another. A more consistent presentation across panels would improve readability.
3. Related to this, I still think the suggestion made by the other reviewer to include a 2×2 overview figure is an excellent idea. I also want to clarify that this is not inconsistent with my previous recommendation regarding line plots. In my view, the two visualizations serve different purposes: a 2×2 overview facilitates comparisons across the experimental conditions, whereas line plots are particularly useful for illustrating the direction of interaction effects.

Response: Thank you for your feedback! We see that the revised figure is confusing and does not show the three-way interaction. We have therefore followed your suggestion to split the figure into a 2×2 overview that shows the distributions for each condition and an additional line plot to better illustrate all interactions, including the three-way interaction in Figure 1B.

2. I am still not entirely convinced by the reasoning in lines 153–157. I have difficulty following the proposed mechanism through which strong social norms would indicate or imply policy effectiveness. While I understand that the authors wish to motivate their hypotheses conceptually, it remains unclear why perceived effectiveness needs to be introduced at this point. If effectiveness plays a role in the argument, I think the underlying mechanism should be explained more explicitly.

Response: We introduced effectiveness at this point to illustrate that while pull measures might be more acceptable, they do not always achieve goals of climate protection effectively. Therefore, it might be interesting to researchers and policymakers to investigate how we could compensate the more negative evaluations of push measures. It is true that a social norm does not automatically imply effectiveness, but might communicate the necessity of implementing unfavourable policies which are more effective, namely push measures. We have changed the wording as follows to make this point clearer:

> *"Pull measures, in contrast, are generally acceptable regardless of the behavioural cost. But pull measures are not always effective enough for climate change mitigation (Gärling, 2007). A strong social norm might signal the necessity of effective policies and might therefore mitigate the detrimental effect of push measures on acceptability, while a weak norm might amplify it."*

3. The newly added discussion of multidimensional acceptance (around line 374) remains rather brief. It would strengthen the discussion if the authors elaborated on what the multidimensionality of acceptance could imply for interpreting the present findings. Instead of just pointing to the recommendations in the cited reference rather explain how they relate to the aims and findings of the present study.

Environmental Psychology Open

Response: Thank you for this feedback, we elaborated a bit more on this point, with a focus on the lack of a behavioural dimension in acceptability ratings compared to ratings of acceptance and support. The paragraph now reads as follows:

> *"Second, we assessed policy acceptability by using a single item. This reduces complexity and effort for participants but potentially forgoes information by operationalising acceptability as unidimensional. Additionally, ratings of acceptability are "passive evaluative response[s] to proposals" (Kyselá, Ščasný, & Zvěřinová, 2019, p. 886) which do not include a behavioural component (e.g. a readiness to act in favour of such a policy). This means that the ratings on this item express a tendency to passively accept such policies, if they were implemented, but do not allow for inferences about actual acceptance or active support of environmental policies. Future studies could implement more complex measures that not only include acceptability, but acceptance and support of policies, to see which attributes also influence policy evaluations that include a behavioural component."*

4. I found the proposed avenue for future research (around line 372) interesting. It would be valuable if the authors expanded this point by reflecting on how the suggested research direction could be expected to influence the effects investigated in the present study.

Response: Thank you for this remark, we have added a short reflection on what we would expect but did not want this to take up too much space, given the limited word count and that expected effects in such studies would greatly depend on specific design choices. We therefore added the following sentence to the end of the paragraph:

> *"While we expect the effects found in our study and in De Groot to be generalizable, it would be interesting to see if their strengths differ between domains of environmental policies (e.g., transportation, food, littering, etc.) and whether a greater variability of stimuli would be a more robust test."*

5. The discussion of limitations has improved. However, the limitation about awareness of comparisons in line 388 remains somewhat incomplete because, unlike other limitations, it is not accompanied by a discussion of its implications or possible ways to address it in future research. I encourage the authors to briefly explain the consequences of this limitation and how future studies could overcome it.

Response: We expanded on this briefly but did not want to give this point too much room, because we cannot make claims about how well the participants in our study were aware and reacting to the comparisons. The section now reads as follows:

> *"Additionally, participants might see through the logic of comparisons and react to the design by intentionally behaving conforming or contrary to our hypotheses, confounding the effects in question. This could either be countered by using a larger stimulus set or controlled by asking participants afterwards about whether they used a specific strategy to answer the questions or what they think the hypotheses are."*

Environmental Psychology Open

6. I also remain somewhat unconvinced by the argument in the concluding discussion that effectiveness should be the primary concern whereas implementation depends on acceptance. The impact statement itself emphasizes the need to investigate how acceptance and effectiveness should be balanced. I therefore think the concluding argument could be framed more carefully to better acknowledge this interdependence rather than presenting effectiveness as the primary consideration.

Response: Thank you for your feedback. You are correct that although the climate crisis is urgent, acceptability and effectiveness are interdependent. We have changed this sentence as follows: "Furthermore, as previous research has shown, effectiveness and acceptability are interlinked and should therefore be taken into account together when designing climate policies."

7. I appreciate that the authors examined the influence of the available control variables and explained why they decided not to include them in the main analyses. Nevertheless, I would still welcome a brief acknowledgement in the discussion that the study did not account for several variables that have consistently been shown to predict climate policy acceptance (e.g., climate change concern). Even if these variables were omitted for instance for parsimony, their absence should be recognized as a limitation.

Response: We have included this aspect within the mentioned limitation section. We now summarized both behaviour and other control variables as variables on the side of the participant who rates a policy's acceptability. The end of this paragraph now reads as follows:

> *"Additionally, future studies should collect data both on existing behaviours like transportation habits or littering, and on control variables that have previously been shown to predict policy acceptance (e.g. environmental concern or values, see Ejelöv & Nilsson, 2020). This would give a more holistic picture of the reasons for (un)acceptability ratings for specific policies, both on the side of the policy as well on the side of the individual giving the rating."*

8. Finally, around line 401, the term "aspects" remains somewhat vague. Are "aspects" and "features" intended to be interchangeable? The added example in line 403 appears to illustrate different aspects within a single policy, but it would also be helpful to provide an example of combining policies with different features.

Response: Thank you for this remark, we have streamlined the wording to "features". On the other hand we would refrain from making an example of combining multiple policies as our study focused on singular policies, not policy packaging. We have, however, cited Geels (2012), Givoni (2014) and Wang et al. (2020), where readers can delve more into policy packaging.

Minor comments

I would recommend another careful proofread for English language and grammar as I came across several optimizable sections while reviewing the revisions. Several sentences appear to have been inserted during revision without fully adapting the surrounding text. More generally, whenever new sentences are added, the neighboring text should be checked to ensure that transitions and references remain clear.

- For example, line 106 ("the policy' acceptability") appears grammatically incorrect.
- The transition "In addition" (around line 115) feels somewhat unnatural given the preceding revisions.

Environmental Psychology Open

- Around line 120, the pronoun "they" no longer has a clear antecedent following the inserted text (the reader does not know whether “they” is with regards to Schuitema et al or “other research”). "in regard" may not be the most idiomatic wording in this context.
- Line 122 contains an extra "w".
- The statement regarding globally stronger environmental protection norms (around line 138) would benefit from a supporting reference. I apologize for overlooking this in the previous review.
- Lines 152–153 begin two consecutive sentences with "But".
- Line 264: remove the unnecessary "as".
- In lines 282 and 290, "rated as acceptable" suggests a threshold between acceptable and unacceptable. I think "rated as more acceptable" would better reflect the continuous nature of the measure.
- Line 343: the author's initials should be removed.

Response: Thank you for those detailed remarks, we have integrated them and looked up the literature on the development of societal environmental norms. In the end, we changed the argument slidely, as we did not find reports on environmental norms, but environmental concern over time. Importantly, this argument is speculative and primarily the motivation to replicate the study by de Groot & Schuitema.

Thank you for your detailed feedback! We hope to have treated each point satisfactorily.