

Conditional Bidirectionality Between Climate Policy Acceptance and Perceived Effectiveness

Corresponding Author: Mr Jean-Henri Huttarsch

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The experimental design (FSE) and sophisticated statistical analysis (path analysis with cluster-robust standard errors, random effects regressions) are appropriate for the research questions. The core findings regarding conditional bidirectionality and the importance of fairness and personal costs as gateways to perceived effectiveness are significant and robust. However, the authors explicitly identify several limitations and future research avenues that need to be addressed, either by strengthening the discussion or clarifying methodological choices. The specific revisions requested below aim to address these inherent limitations and clarify the theoretical implications of the results.

1. The authors acknowledge that the national German sample and the specific focus on heating policy may limit the straightforward generalizability of the findings to other domains (like transport or agriculture) or across cultures. Even though the authors assert the results are likely generalizable to other WEIRD countries due to homogeneity in public support findings, the practical implications for policy design (Section 359 onward) are profound. The discussion should more explicitly tie the findings to the context of international sustainability, perhaps by framing the German heating sector as a microcosm of global decarbonization challenges that rely on public acceptance. The need for replication in different political contexts remains a key caveat.

2. The manuscript notes that the experimental setup was kept deliberately simple, which is "far from reality," where policies are often part of complex mixes. The authors suggest future work should embed these psychological mechanisms within the complexity of real-world policy packages and consider a wider array of instrument-specific beliefs (e.g., perceived personal consequences, revenue use). Given the journal's focus on sustainability, the authors should dedicate more discussion space to how their current, simpler findings might translate or change when applied to complex policy mixes, or how policymakers should proceed in such complex environments based on their core findings (e.g., prioritizing fairness cues in any policy mix communication).

3. The study largely confirmed the moderating impact of GCPS on the effects of fairness and effectiveness cues on acceptance. However, the authors note that their findings suggest a potential boundary condition regarding the assumption that general support dampens the negative effect of policy intrusiveness (regulatory depth). The authors themselves suggest future studies re-analyze this using a wider spectrum of policy intrusiveness (e.g., bans and mandates). This point is critical because it suggests a theoretical tension. The discussion should elaborate on why the specific design features tested (push/pull mechanisms) might not have revealed the expected moderation effect concerning policy intrusiveness, or how the current findings relate to the concept of regulatory depth, before relegating the issue entirely to future work.

4. The findings show that, for the entire sample, the alternative causal pathway model (SRMR = .01) is statistically more appropriate than the traditional pathway model (SRMR = .06). The authors later justify this by showing that only supporters exhibit the traditional pathway, while both groups show the alternative pathway. This result is fundamental. The Discussion (starting around Section 308) should explicitly synthesize this finding—that the alternative, acceptance-driven pathway is a "more general mechanism" while the traditional, effectiveness-driven pathway is conditional on prior support—to fully explain why the whole-sample model fit comparisons favored the alternative pathway.

5. Furthermore, the Methods section indicates that 678 individuals were excluded from the focal analytical sample because they were assigned vignettes focusing only on energy security rather than climate targets. It is noted that the analysis focuses only on effectiveness levels 1 and 2 (climate effectiveness). While this is understandable for an analysis focusing on climate policy attitudes, the exclusion of a large portion of the data based on the effectiveness dimension (climate vs. energy

security) should be made clearer earlier in the Results/Methods introduction to avoid confusion. Also, General Climate Policy Support (GCPS) was operationalized as the average of two items, which showed satisfactory internal consistency ($\alpha = .83$). Given the heavy reliance on this construct as a moderator and group differentiator, and the authors' own call for future research to assess GCPS more thoroughly, it would be beneficial to briefly mention the two-item operationalization as a study limitation in the discussion or methods, perhaps noting its parsimony while reaffirming the finding's robustness. 6. Finally, the manuscript effectively uses the Ratio of Indirect Effect to Direct Effect (RID) to gauge the importance of the indirect effects. For the alternative causal pathway involving instrument type, the RID is approximately 141.1 times the direct effect. This is a powerful statistic indicating a nearly full mediation. For the traditional causal pathway, the RID is 4.1. Ensuring that the discussion clearly highlights the magnitude difference between these RIDs, especially in the multi-group comparison, reinforces the asymmetry in information processing between the pathways.

Reviewer #2

(Remarks to the Author)

I have read the abovementioned study which attempts to establish a causal structure between climate policy support and perceived effectiveness. I find the study is important, it uses mostly suitable data and methods. However, I still see some room for improvement before the study can be published.

Below I list my observations in the order I made them

1. "technical effectiveness" sounds strange, just effectiveness of perceived effectiveness is likely meant

2. Authors refer in introduction to expert evaluation of distinct policies. There are two recent studies on international survey of experts on perceived policy effectiveness and support. I think authors could engage with these studies as they seem very relevant for further discussion

Savin I., S. Drews and J. van den Bergh, 2024, Carbon Pricing – Strengths, Weaknesses and Knowledge Gaps according to a Global Survey, *Environmental Research Letters*, 19 024014 <https://doi.org/10.1088/1748-9326/ad1c1c>

Drews S., Savin I. and J. van den Bergh, 2024, A Global Survey of Scientific Consensus and Controversy on Climate Policy, *Ecological Economics*, 218: 108098 <https://doi.org/10.1016/j.ecolecon.2023.108098>

3. Figure 2. I find one important element potentially omitted in the analysis, namely perceived peer support, or others call is second order beliefs. The fact I believe my peers support sth affects me. Many studies found a strong and significant link between policy support and SOBs. See Drews et al. 2022. I think authors could at least acknowledge as limitation and direction for future research

Drews S., Savin I. and J. van den Bergh, 2022, Biased perceptions of other people's attitudes to carbon taxation, *Energy Policy* 167:113051 <https://doi.org/10.1016/j.enpol.2022.113051>

4. line 143 Figure numbering is corrupt as you have Figure 1 again.

But more importantly I find the study would benefit if authors could quantify the share of variance in the dependent variable being explained by the focal explanatory variable. This is done by Shapley value decomposition of R^2 . I think study would benefit from clarifying this information.

See Savin et al. 2024 for an example

Savin, Ivan, Philipp Mundt, and Margherita Bellanca. The paradox of climate policy diffusion. No. 196. BERG Working Paper Series, 2024.

5. lines 230/240 Authors quote a lot of coefficients but give not so much intuition for the matter discussed. I would suggest to rewrite some text here to stress less figures and more the intuition

6. line 250 Authors split the analysis between policy supporters and policy opponents. I find that dichotomous split not good enough. A recent study by Smith et al. 2025 forthcoming in *Nature Climate Change* shows that the most interesting group of people are conditional middle as they are more likely to still change their opinion.

I think authors should try to split their sample in three groups eliciting undecided and show how robust their results will be

Smith, E. Keith, Žan Mlakar, Alessio Levis, Mary Sanford, Lea Stapper, Thijs Bouman, Johannes Emmerling et al. "Climate Policy Feasibility across Europe Relies on the Conditional Middle." (2025).

Reviewer #3

(Remarks to the Author)

Version 1:

Reviewer comments:

Reviewer #2

(Remarks to the Author)

I think authors for addressing the issues I raised.

I recommend publication of these results

Reviewer #4

(Remarks to the Author)

In general, the authors have done well in responding to the concerns of Reviewer 3 and considerably revised their language around bidirectionality. I believe the manuscript is now more careful not to overstate the reported results. While the manuscript has considerably improved, there are a couple more points that I believe would further improve its' quality.

First, regarding the comparison between the traditional and alternative causal pathway, I believe a comparison of symmetrical models could strengthen the claims for evidence consistent with bidirectionality. There are papers that report a causal link between policy design features such as push vs. pull measures with perceived effectiveness (see e.g. Swim & Geiger, 2021). Seeing as other literature also points towards correlations between fairness perceptions, perceptions of personal impact and effectiveness perceptions are correlated, I think it is fair to say this link is established in the literature. As such, I think a symmetrical comparison added to the supplements could alleviate concerns that the specific model comparisons chosen by the researchers lead to evidence suggesting that the alternative causal pathway is the more general mechanism at play here.

Relatedly, I think the authors could point towards other ways at getting at causality such as instrumental variable estimation using a 2-Stage Least Squares approach (see e.g. Bastardo et al., 2023; Sajons, 2020) or Directed Acyclic Graphs as avenues for future research to identify causal mechanisms in climate policy support research.

Second, I think the manuscript would benefit from a clearer definition of what the authors mean by policy acceptance and policy support. I can see the way the operationalized the concepts but I nevertheless think a somewhat more explicit discussion of what they are measuring and how policy acceptance differs from GCPS (and if what they call GCPS is actually support or also acceptance) would help the field as these terms are used relatively loosely throughout much of the literature. The authors do already cite a paper offering a framework/construct of how to differentiate between different constructs in climate policy support research (Kyselá et al., 2019), so it is somewhat surprising to see them not making use of the construct Kyselá and colleagues provide at all.

Finally, while I do think the findings on GCPS-opponents are highly relevant, they could be integrated more explicitly with research showing that highlighting the co-benefits of climate mitigation motivates climate action (see e.g. Bain et al., 2015). In my opinion, the results reported here and in papers such as the Bain et al. paper strongly suggest that we do not actually need to communicate about climate change or what a certain policy does in relation to climate change to convince people to support it. I think this point deserves somewhat more attention in the manuscript.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Reviewer 1

The experimental design (FSE) and sophisticated statistical analysis (path analysis with cluster-robust standard errors, random effects regressions) are appropriate for the research questions. The core findings regarding conditional bidirectionality and the importance of fairness and personal costs as gateways to perceived effectiveness are significant and robust. However, the authors explicitly identify several limitations and future research avenues that need to be addressed, either by strengthening the discussion or clarifying methodological choices. The specific revisions requested below aim to address these inherent limitations and clarify the theoretical implications of the results.	Note: Alle line numbers provided refer to the manuscript version that displays all tracked changes.
1.The authors acknowledge that the national German sample and the specific focus on heating policy may limit the straightforward generalizability of the findings to other domains (like transport or agriculture) or across cultures. Even though the authors assert the results are likely generalizable to other WEIRD countries due to homogeneity in public support findings, the practical implications for policy design (Section 359 onward) are profound. The discussion should more explicitly tie the findings to the context of international sustainability, perhaps by framing the German heating sector as a microcosm of global decarbonization challenges that rely on public acceptance. The need for replication in different political contexts remains a key caveat.	We thank the reviewer for the suggestions to further sharpen contextualization and thus implications for generalizability of our results. To this end, we improved the introduction of the heating sector (line 123) and reformulated lines 460-470.
2.The manuscript notes that the experimental setup was kept deliberately simple, which is "far from reality," where policies are often part of complex mixes. The authors suggest future work should embed these psychological mechanisms within the complexity of real-world policy packages and consider a wider array of instrument-specific beliefs (e.g., perceived personal consequences, revenue use). Given the journal's focus on sustainability, the authors should dedicate more discussion space to how their current, simpler findings might translate or change when applied to complex policy mixes, or how policymakers should proceed in such complex environments based on their core findings (e.g., prioritizing fairness cues in any policy mix communication).	We've taken into account the important suggestion to further elaborate the implications of our findings and extended our paragraph on practical implications – lines 519-526.

3. The study largely confirmed the moderating impact of GCPS on the effects of fairness and effectiveness cues on acceptance. However, the authors note that their findings suggest a potential boundary condition regarding the assumption that general support dampens the negative effect of policy intrusiveness (regulatory depth). The authors themselves suggest future studies re-analyze this using a wider spectrum of policy intrusiveness (e.g., bans and mandates). This point is critical because it suggests a theoretical tension. The discussion should elaborate on why the specific design features tested (push/pull mechanisms) might not have revealed the expected moderation effect concerning policy intrusiveness, or how the current findings relate to the concept of regulatory depth, before relegating the issue entirely to future work.	We have extended the part of the discussion dealing with the implications of our findings for the integrative acceptance framework in which the regulatory depth concept plays a major role. We hope lines 423-430 address this comment satisfactorily.
4. The findings show that, for the entire sample, the alternative causal pathway model (SRMR = .01) is statistically more appropriate than the traditional pathway model (SRMR = .06). The authors later justify this by showing that only supporters exhibit the traditional pathway, while both groups show the alternative pathway. This result is fundamental. The Discussion (starting around Section 308) should explicitly synthesize this finding—that the alternative, acceptance-driven pathway is a "more general mechanism" while the traditional, effectiveness-driven pathway is conditional on prior support—to fully explain why the whole-sample model fit comparisons favored the alternative pathway.	Thank you for this suggestion that improves the key take-aways from our investigation. We hope the clarifications in lines 393-399 better highlight this fundamental finding.
5. Furthermore, the Methods section indicates that 678 individuals were excluded from the focal analytical sample because they were assigned vignettes focusing only on energy security rather than climate targets. It is noted that the analysis focuses only on effectiveness levels 1 and 2 (climate effectiveness). While this is understandable for an analysis focusing on climate policy attitudes, the exclusion of a large portion of the data based on the effectiveness dimension (climate vs. energy security) should be made clearer earlier in the Results/Methods introduction to avoid confusion. Also, General Climate Policy Support (GCPS) was operationalized as the average of two items, which showed satisfactory internal consistency ($\alpha = .83$). Given the heavy reliance on this construct as a moderator and group differentiator, and the	Thank you for this helpful suggestion regarding the further experimental conditions. We have clarified earlier in the manuscript (right before the results and in the beginning of methods) that the focal analyses are restricted to a subset of effectiveness conditions, with additional conditions reported in the Supplementary Information, to avoid potential confusion (see lines 139-145 and 537-540). For each results section, we also added statements about how the results regarding the additional experimental conditions provide robustness to our focal analyses.

authors' own call for future research to assess GCPS more thoroughly, it would be beneficial to briefly mention the two-item operationalization as a study limitation in the discussion or methods, perhaps noting its parsimony while reaffirming the finding's robustness.	Regarding the GCPS operationalization, we acknowledge that our two-item measure represents a parsimonious choice that reflected the survey-length constraints common to large studies. We note three considerations that support the robustness of our findings despite this brevity. First, the two-item scale showed satisfactory internal consistency (Cronbach's $\alpha = .83$). Second, the moderation patterns observed for GCPS were consistent across both the climate-effectiveness and the energy-security-effectiveness experimental conditions (see Supplementary Information Section 2), indicating that the construct functioned reliably across distinct framing contexts. Third, the simple-slopes analyses (Tables S12 and S13) yielded a coherent and theoretically expected pattern across GCPS levels, further supporting the validity of the operationalization. We have nevertheless extended the discussion of GCPS-related limitations to acknowledge the value of more comprehensive assessments in future research (lines 486-490).
6. Finally, the manuscript effectively uses the Ratio of Indirect Effect to Direct Effect (RID) to gauge the importance of the indirect effects. For the alternative causal pathway involving instrument type, the RID is approximately 141.1 times the direct effect. This is a powerful statistic indicating a nearly full mediation. For the traditional causal pathway, the RID is 4.1. Ensuring that the discussion clearly highlights the magnitude difference between these RIDs, especially in the multi-group comparison, reinforces the asymmetry in information processing between the pathways.	Thank you for helping us improve the description of the RID. We agree that both RIDs are consistent with a pattern of (near) full mediation. At the same time, we would like to be cautious in interpreting the large difference in RID values between the two pathways. Because the RID is a ratio, it is highly sensitive to the magnitude of the direct effect in the denominator. In the case of the alternative pathway, the direct effect is very close to zero, which mechanically inflates the RID. By contrast, in the traditional pathway, the direct effect is somewhat larger (albeit still statistically indistinguishable from zero), resulting in a smaller ratio. For this reason, we believe that comparing the magnitude of the RIDs across pathways may be misleading, as the difference primarily reflects variation in the magnitude of the residual direct effects rather than substantive differences in mediation strength. In line with standard mediation interpretation, we therefore place greater emphasis on the size and significance of the indirect effects. To address this point, we have added a note in lines 212-214 to avoid overemphasizing cross-pathway comparisons of its magnitude.

Reviewer 2

I have read the abovementioned study which attempts to establish a causal structure between climate policy support and perceived effectiveness. I find the study is important, it uses mostly suitable data and methods. However, I still see some room for improvement before the study can be published. Below I list my observations in the order I made them	Thank you for the instructive comments. We hope we have addressed them in satisfactory manner. Note: Alle line numbers provided refer to the manuscript version that displays all tracked changes.
1. "technical effectiveness" sounds strange, just effectiveness of perceived effectiveness is likely meant	We have reformulated the phrase and hope it sounds more straightforward now.
2. Authors refer in introduction to expert evaluation of distinct policies. There are two recent studies on international survey of experts on perceived policy effectiveness and support. I think authors could engage with these studies as they seem very relevant for further discussion Savin I., S. Drews and J. van den Bergh, 2024, Carbon Pricing – Strengths, Weaknesses and Knowledge Gaps according to a Global Survey, Environmental Research Letters, 19 024014 https://doi.org/10.1088/1748-9326/ad1c1c Drews S., Savin I. and J. van den Bergh, 2024, A Global Survey of Scientific Consensus and Controversy on Climate Policy, Ecological Economics, 218: 108098 https://doi.org/10.1016/j.ecolecon.2023.108098	Thank you for these valuable references that back up the puzzle used in the introduction - we included them.
3. Figure 2. I find one important element potentially omitted in the analysis, namely perceived peer support, or others call it second order beliefs. The fact I believe my peers support sth affects me. Many studies found a strong and significant link between policy support and SOBs. See Drews et al. 2022. I think authors could at least acknowledge as limitation and direction for future research Drews S., Savin I. and J. van den Bergh, 2022, Biased perceptions of other people's attitudes to carbon taxation, Energy Policy 167:113051 https://doi.org/10.1016/j.enpol.2022.113051	We thank the reviewer for raising this important point. Second-order beliefs about peer attitudes are indeed a well-documented determinant of climate policy support. Our study deliberately draws on GEF and GEF-related aspects of the integrative policy acceptance framework, which focus on policy-specific evaluations rather than social-cognitive determinants. We have acknowledged this limitation in the respective section (lines 482-486).
4. line 143 Figure numbering is corrupt as you have Figure 1 again. But more importantly I find the study would benefit if authors could quantify the share of variable in the dependent variable being explained by the focal	Thank you for this helpful suggestion. Following your comment, we have revised our figures to report the equation-level R^2 for each endogenous

explanatory variable. This is done by Shapley value decomposition of R^2. I think study would benefit from clarifying this information. See Savin et al. 2024 for an example Savin, Ivan, Philipp Mundt, and Margherita Bellanca. The paradox of climate policy diffusion. No. 196. BERG Working Paper Series, 2024.	variable in the SEM, thereby making the model's explanatory power more transparent. With regard to the proposed use of Shapley value decomposition of R^2, we agree that this approach can be informative in standard multiple regression settings to attribute shares of explained variance to individual predictors. However, its application is less straightforward in the context of SEM and mediation models. Our analytical framework is based on a model that explicitly decomposes effects into direct and indirect (mediated) pathways. In contrast, Shapley decomposition attributes contributions to explained variance without distinguishing between these pathways. As a result, variance attributed to mediators absorbs, by construction, parts of the indirect effects of upstream variables, thereby distracting from the causal interpretation that is central to mediation analysis. Nevertheless, in response to the reviewer's suggestion, we have implemented a Shapley decomposition of R^2 and report the results as a descriptive complement in Tables S8 and S9.
5. lines 230/240 Authors quote a lot of coefficients but give not so much intuition for the matter discussed. I would suggest to rewrite some text here to stress less figures and more the intuition	We have rewritten some these parts (now lines 287-292) to focus less on coefficients and to be more intuitively understandable. We hope this in line with what the reviewer intended.
6.line 250 Authors split the analysis between policy supporters and policy opponents. I find that dichotomous split not good enough. A recent study by Smith et al. 2025 forthcoming in Nature Climate Change shows that the most interesting group of people are conditional middle as they are more likely to still change their opinion. I think authors should try to split their sample in three groups eliciting undecided and show how robust their results will be Smith, E. Keith, Žan Mlakar, Alessio Levis, Mary Sanford, Lea Stapper, Thijs Bouman, Johannes Emmerling et al. "Climate Policy Feasibility across Europe Relies on the Conditional Middle." (2025).	Thank you for pointing us to this valuable reference. Following this suggestion, we have updated our analyses to include a third group beyond supporters and opponents - the undecided - and extensively rewritten our results and the discussion accordingly. The results partly corroborate Smith et al.'s (2026) conclusion that beyond supporters, the middle segment can also be swayed. There is however an important methodological caveat that we've noted in lines 440-444. Smith et al.'s 'Conditional Middle' is identified through a two-dimensional classification combining mean support and response variability across multiple policies. Our single-domain design (heating sector vignettes) does not allow us to reproduce this classification fully. Our 'undecided' group

	thus likely combines what Smith et al. distinguish as 'Neutrals' (disengaged middle) and 'Conditionals' (engaged but persuadable middle).
--	---

Reviewer 3

Summary: The authors use a survey experiment with German respondents to test the causal relationship between climate policy acceptance and perceived policy effectiveness. Specifically, they argue that most models of policy acceptance assume that effectiveness drives acceptance, whereas they find greater evidence for the alternative pathway: that acceptance predicts effectiveness. Additionally, they find that these results are moderated by whether respondents generally support climate policy.	Note: Alle line numbers provided refer to manuscript version that displays all tracked changes.
Major Comments: The potential bidirectionality of policy acceptance and effectiveness perceptions is an intriguing and interesting possibility, especially considering recent literature showing that correcting misperceptions about the effectiveness of pro-environmental behaviors and climate policy does in fact increase respondent willingness to take such behaviors and policy support. I had some concerns and questions as I read through the paper that I have enumerated below — I hope these can be helpful.	
1. It would be helpful to see more motivation for why fairness and personal benefit are expected to work better than effectiveness information, and why these two factors over the many other determinants of policy support identified in the literature.	We thank the reviewer for this comment that highlights important points of clarification regarding the structure of our argument. The comment raises two related but distinct points. First, we do not hypothesize that fairness and personal benefit related instrument features are inherently more powerful drivers of acceptance than effectiveness. Rather, we argue they are more cognitively accessible to citizens. While effectiveness is often abstract and difficult to gauge <i>ex ante</i>—particularly for novel or unimplemented policies—fairness and personal costs/benefits are anchored in directly observable design features. Consequently, individuals likely rely on these more "tangible" cues when forming judgments, even when

	abstract effectiveness information is provided. We elaborate this reasoning in the revised introduction (lines 52-64). Second, our focus on these three variables is grounded in the Integrative Policy Acceptance Framework and the GEF framework, which identify effectiveness, fairness, and personal benefit as the three canonical pillars of policy-specific evaluation. Furthermore, meta-analytic evidence (Bergquist et al., 2022) suggests that policy-specific evaluations are the strongest predictors of support. While we acknowledge the importance of factors like trust, political identity, and social norms, these are dispositional or contextual variables rather than policy-design features. Since our experimental design specifically targets the manipulation of policy architecture, these dispositional factors fall outside our scope—a choice we now explicitly acknowledge as a limitation (lines 482-486).
2. Why is information about distributional impacts and personal benefits not also considered part of the traditional causal model, alongside effectiveness? (e.g. in the depiction of Figure 1, but also in the text) And why are they not also assumed to be impacted by policy acceptance in a similar bidirectional model to that of effectiveness? In general, the specific assumptions here seem somewhat adhoc and would benefit from greater motivation make the case for bidirectionality or the alternative pathway in the case of effectiveness and not the other inputs to policy support	Indeed, the distinct bidirectional role of primarily perceived effectiveness and not fairness and personal benefit is central to our RQs and needs to be clear. Our framework is intentionally not fully symmetric across all policy-specific beliefs. It builds on established distinctions in the literature regarding how different evaluative dimensions are formed and updated. In particular, perceived fairness and personal benefits are directly anchored in concrete, observable policy design features (e.g., subsidy levels, eligibility criteria) and immediate distributional consequences. Because these dimensions are more tangible, they provide less latitude for subjective re-interpretation. By contrast, to our knowledge there is only contesting literature regarding effectiveness that suggests that it is often less directly observable and therefore more susceptible to construction, inference, and post-hoc rationalization, especially in contexts of uncertainty or limited experience with policy outcomes. This motivates our focus on effectiveness as a candidate for bidirectional updating processes between policy evaluations and perceived effectiveness.

	We agree that, in principle, fairness and personal benefit perceptions could also exhibit some degree of endogeneity with policy support. However, the existing literature provides comparatively few arguments for such feedback processes in those domains. Our model therefore reflects a deliberate selection focusing on the construct where such dynamics are most strongly theorized and empirically suggested. We have clarified this asymmetry and its theoretical motivation in the revised manuscript (lines 60-64) and edited the Figure 1 description accordingly.
3. The moderation by general climate policy support (lines 81-83) was also lacking motivation — and in fact I could not quite see how it tied cleanly into the first question about effectiveness. I was also unclear about the predicted moderation directions in reading the paper. Was this study pre-registered? This would help mitigate concerns that some of these analyses seem a bit arbitrary or unmotivated. For example, in a caption I see that you assume that general policy support reduces the impact of type of instrument and increases responsiveness to distributional principles and effectiveness — however, why this is the case is not clearly laid out.	Thank you for pointing us towards these issues. Regarding the relationship between RQ1 and RQ2, we agree that the connection between the two research questions was not made sufficiently explicit, and we have revised the introduction (lines 92-96 and 121) so that it clarifies the underlying logic: both RQs address distinct but complementary gaps in the literature on climate policy acceptance. Regarding preregistration: the study was not. We apologize if any confusion arose. We distinguish analyses that were directly derived from our focal theory-driven research questions and hypotheses (RQ1 and RQ2) from those that weren't (exploratory analyses – see also reply to comment 7). We hope that the revisions in the manuscript with regards to the assumptions underlying RQ1 (comments 1, 2 and 15) and RQ2 (this comment) now clearly lay out our motivation. Regarding the direction of the moderating effects, we agree that they were not adequately motivated, and we edited lines 102-108 to more explicitly draw on the underlying acceptance framework and its reasoning.
4. My biggest concern is with the design and analyses used to identify evidence in favor of the alternative causal pathway over the traditional one. In the current study, the authors manipulate policy instrument, distributional	We appreciate this important methodological point and agree that our narrative should be reformulated to be more cautious about its causal conclusions. We have moderated the language in the entire manuscript to

impacts, and effectiveness in an single cross-sectional study. This allows them to test the first pathway since effectiveness is randomly manipulated and the causal effect on policy acceptance can be tested. However, they do not vary policy acceptance independently of effectiveness information and then see how increased policy support increases effectiveness perceptions, or alternately include multiple waves to assess the relationship longitudinally. I think this would be a tighter test of the two causal pathways and in the current setup the ability to make causal inferences is limited (and currently overstated).	avoid overstating claims about causal directionality beyond what our design supports (lines 146-148, 191, 385, 393-395, 434-437, 501-505 & 528) and transparently acknowledge its limitation (lines 470-474). To further clarify our design approach: In our factorial survey experiment, effectiveness perceptions and instrument support/acceptance are measured as outcomes following exposure to randomly varied vignette attributes, including effectiveness information. Because all experimental dimensions are independently randomized across respondents, this design allows for estimation of their respective causal effects on both perceived effectiveness and policy support while averaging over all other dimensions. For the post-treatment outcomes (policy support and perceived effectiveness), we then examine theoretically informed but experimentally anchored mediation patterns to assess plausible directional relationships with competitive model-comparison within the limits of a cross-sectional design. Indeed, confirming such processes would require longitudinal data, which we now explicitly acknowledge as a limitation.
5. Why is climate concern, but no other climate policy related variable, included as a control? Are results robust to its exclusion?	Climate concern was included as a theoretically relevant baseline covariate because it captures a general predisposition closely related to climate policy support and is commonly controlled for in this literature. Importantly, it also emerged as the strongest predictor among a broader set of climate-change evaluations in a Meta-Analysis (Bergquist et al., 2022), underscoring its centrality. In addition, we account for political orientation via party identification as a proxy for ideology, which is another impactful climate-policy-relevant covariate. We did not include a wider set of climate-specific controls, consistent with our aim of parsimony and standard practice, and we previously noted that climate change belief was not included due to its limited discriminatory power in the German context. We have edited lines 644-645 to make this clearer. To address concerns, we additionally re-ran our RQ1 and RQ2 analyses without climate concern as a control. The substantive pattern of results remains

	unchanged (see Supplementary Information Tables S6 and S11), indicating that the findings are robust to its exclusion. Because the exploratory analysis is a combination of the model components underlying RQ1 and RQ2, for which exclusion of climate concern and inclusion of dropped participants showed no substantive changes, we did not report separate additional robustness analyses for the exploratory models, which can be provided if desired.
6. The authors conclude that both pathways are at play, but that there is more evidence in favor of the alternative pathway. How should one think about the difference between SRMR=.01 and SRMR=.06. Is this a large and significant difference? How should it be interpreted given that the models have different number of paths, with the alternative having more?	Thank you for this important comment. We agree that the SRMR difference should not carry the comparative weight on its own. We have restructured the relevant passage so that AIC and BIC lead, both of which penalize model complexity. The differences exceed conventional thresholds. The lower SRMR for the alternative model is now reported as consistent with, not as the basis for, this comparative conclusion (see lines 240-246).
7. Line 246 suggests that the GCPS moderation is an exploratory analysis. Was the rest pre-registered then? Apologies if I missed this! This would be very helpful to see. Also on the GCPS, did the authors also try a continuous moderation? It was unclear to me why the median split was needed.	We apologize for any confusion created by our use of the term “exploratory”. By exploratory, we did not intend to imply that the remaining analyses were preregistered. Rather, we use “exploratory” to distinguish analyses that were not directly derived from our focal theory-driven research questions and hypotheses from those that were. Regarding GCPS moderation, our moderation research question (RQ2) is estimated using a continuous underlying construct. The respective simple slopes use then 3 groups in this continuum to illustrate the interaction. The multi-group structural equation modeling in the explorative analyses, however, requires grouping rather than continuous interaction. We have now adapted the groups to complement the 3 simple slopes and now distinguish 3 groups (supporters, opponents and the undecided). This specification is also empirically motivated by literature suggesting that neutral respondents may behave differently from both supporters and opponents. The three-group approach now captures more nuance than the median split and better approximates variation along the underlying continuum.

8. The authors find that both groups show evidence for the alternative causal pathways, but that the traditional causal pathway only applies to general climate policy supporters. It was not clear to me how to make sense of this result. In general, interpretation that is specific and points to the relevant literature would be helpful.	Thank you for this helpful comment. We agree that this finding warranted stronger theoretical interpretation. We revised lines 393-398 and 434-448 to clarify that the asymmetry does not lie in motivated reasoning being present in some groups but absent in others. Rather, all groups show evidence consistent with attitude-congruent processing via the alternative pathway, while only supporters and undecided respondents additionally show evidence of updating through the traditional pathway. We now situate this interpretation more explicitly in relation to the tension found in the literature.
9. The survey seemed to include many modules before the experiment—such as national climate targets, energy security considerations, and the potential contribution of these instruments to both objectives through reduced fossil fuel consumption in the building sector. Might these have impacted the results? In particular, this seems to already carry effectiveness information?	We thank the reviewer for raising this point. The introductory modules were shown to all respondents prior to random assignment and were included to ensure a shared baseline understanding of the policy context. We agree that such information may influence overall levels of perceived effectiveness. However, because all participants received identical information in the same order, any potential priming effects are orthogonal to treatment assignment. The experiment therefore remains valid, as treatment effects are estimated based on randomized manipulations with a common informational baseline. We have discussed the potential effects of the introductions in the limitations section (lines 475-480).
10. In the methods, I see that effectiveness information had four levels – level of effectiveness X type of impact (energy security or climate). This was not clear in the paper — and the motivation for this is unclear. Do the results look different if you include the different types of impact (energy or climate) in your analyses?	Thank you for this helpful suggestion regarding the further experimental conditions. To more clearly guide the reader that the presented analyses are part of a larger experimental set up and restricted to a subset of effectiveness conditions, we stated this earlier in the manuscript (right before the results and in the beginning of methods) with additional conditions being reported in the Supplementary Information (see lines 139-145 and 537-540). For each results section, we also added statements about how the results regarding the additional experimental conditions provide robustness to our focal analyses.

	Regarding motivation, the energy security effectiveness conditions were included as part of the broader experimental design. They allow to test whether the effects observed for climate effectiveness generalize across an alternative, policy-relevant impact framing. While the present manuscript focuses narratively on climate-target effectiveness, the energy security conditions serve as an additional robustness check and are therefore reported in the Supplementary Information.
11. Are the results robust to including excluded participants (e.g. due to fast completion times)?	To address this concern, we re-ran our RQ1 and RQ2 analyses including participants previously excluded due to fast completion times. The substantive pattern of results remains unchanged. We report these robustness checks, together with robustness checks excluding climate concern as a control, in additional Supplementary Information Tables S6 and S11. Because the exploratory analysis is a combination of the model components underlying RQ1 and RQ2, for which inclusion of participants previously excluded due to fast completion times showed no substantive changes, we expect the multi-group decomposition to behave similarly and have not added a separate check, which can be provided if desired.
12. In the paper there is a statement about how the sample does not look too far from the general population, but this is not how I would have interpreted the results of Table S1. You can show the t-tests and p-values so it is clear where there are significant differences. I don't think this is a big threat to inference given this is an experimental study, but I think it would help to be transparent about the sample.	Thank you for this fair note. We agree that the previous wording understated the sample deviations. To avoid the limitations of significance testing in large samples—where even small differences can yield statistically significant p-values that overstate substantive importance—we do not report t-tests against population benchmarks, but instead transparently describe where the sample deviates from the general population (lines 598-605). We retain the Reviewer's framing that the deviations are not a major threat to inference given the experimental design, while ensuring that readers can locate and judge the deviations themselves. We hope this addresses the reviewer's concerns.
13. The effect sizes are small – and while this is attributed to the debates around this policy, it could also be due to the vignettes. Can the authors compare effect sizes relative to other studies? Or show that they differ for a policy that is not currently the topic of debate?	Thank you for raising this point that prompted us to reflect more carefully on the experimental effects within the context of the FSE literature. We revised the manuscript to avoid characterizing the effects as simply “small,” as this may have been misleading. Instead, we now clarify that the observed effects are

	moderate in absolute magnitude, are not unusually small in the context of factorial survey experiments using multi-point evaluative scales (lines 475-477). Please also note that illustrative examples discussed in the factorial survey literature report effects of a comparable order of magnitude, which further suggests that the observed effects are not atypically small.
14. Lines 301-302 need citations. Especially as some literature does find that information about effectiveness increases support.	We thank the reviewer for catching this gap. Indeed, citations are needed and we have included them. We have also rephrased the opening of the practical implications passage to clarify that our findings qualify rather than overturn this evidence. Effectiveness information remains effective for supporters and the undecided, while being less likely to sway general opponents. The implications now contrast a uniform effectiveness-first approach with an audience-conditional communication strategy.
15. Lines 319-320 (“Effectiveness, in this view, is not just evaluated; it is constructed to be consistent with underlying judgments about a policy’s fairness and its impact on the self”) again raised the question for me about why we should not expect this to also be the case for perceptions of personal benefit and distributional impacts...	The comment raises a very similar conceptual question as in R3.2. As argued in our response above, we believe the key distinction lies in how tangible the instrument features informing the perceptions are: Impact on the self is anchored in more immediate outcomes (e.g., 'Will my heating bill increase by 20%?'). Fairness perceptions are tied to visible policy design features (e.g., 'Does the subsidy apply to everyone or only low-income households?'). In contrast, the effectiveness of a national climate policy is abstract and virtually impossible for a layperson to verify personally. Thus, while an individual might want to see a policy as fair because they like it, the objective design features act as a 'reality constraint.' Effectiveness lacks these immediate reality constraints. We have refined the section to better emphasize this distinction between 'constrained' and 'constructible' perceptions (lines 409-411).
16. On line 327, the authors state that “fairness and effectiveness cues are significantly more persuasive for those already inclined to support political climate action” yet on line 364 they state “To reach opponents and the undecided, policy design and communication must lead with fairness and the mitigation of personal costs”. To me, these seem a little inconsistent.	We thank the reviewer for highlighting this apparent inconsistency. The first passage describes the moderation effect identified in RQ2 (now lines 420-423), where we can show that, the effectiveness and fairness cues significantly increase when you increase GCPS (from opponents to the undecided to supporters, see Table S12). Specifically, simple slopes show that effectiveness

	cues and need-based principle cues are not significant for opponents. While the effect of the polluter-pays principle cue becomes stronger with increasing GCPS, it is not null for opponents. It is only the need-based principle that loses its positive effect on acceptance for opponents. The second passage additionally draws on the differentiated multi-group findings (now lines 434-440), where some design features, namely the polluter-pays principle, pull-oriented design, retain effects across all groups on acceptance and perceived effectiveness in the alternative pathway models. We have revised both passages to make this distinction explicit.
Minor Comments	
17. Lines 18-19, the first two lines of the introduction, do not currently read like a puzzle. Is the puzzle that the technically effective solutions are not politically feasible?	We have reformulated the phrase and hope it represents a puzzle clearly now.
18. I found the following assertion — Bolderdijk et al. ²⁰ argue, because the effectiveness of policy instruments—particularly when they are yet to be implemented—is difficult to gauge, individuals rely on other more readily available information. — confusing because then isn't this why communicating effectiveness information might be particularly valuable? Also, it seems that the effectiveness of climate policy instruments is often hard to assess even after implemented.	Indeed, in principle, communicating effectiveness information could be particularly valuable when the effectiveness of policy instruments is difficult to assess. However, we believe that uncertainty doesn't increase receptiveness to such information, but rather that it can increase reliance on prior attitudes. Individuals may interpret or discount effectiveness information in ways that are consistent with their prior policy preferences. We have revised the manuscript to clarify this mechanism in lines 57-61.
19. Line 314-316 did not make sense to me – what does need refer to here?	We have reformulated these lines and hope that now the reference to the need-based distribution principle is clearer.

Reviewer 4

All lines are indicated with regards to the manuscript showing tracked changes.

First, regarding the comparison between the traditional and alternative causal pathway, I believe a comparison of symmetrical models could strengthen the claims for evidence consistent with bidirectionality. There are papers that report a causal link between policy design features such as push vs. pull measures with perceived effectiveness (see e.g. Swim & Geiger, 2021). Seeing as other literature also points towards correlations between fairness perceptions, perceptions of personal impact and effectiveness perceptions are correlated, I think it is fair to say this link is established in the literature. As such, I think a symmetrical comparison added to the supplements could alleviate concerns that the specific model comparisons chosen by the researchers lead to evidence suggesting that the alternative causal pathway is the more general mechanism at play here.

Relatedly, I think the authors could point towards other ways at getting at causality such as instrumental variable estimation using a 2-Stage Least Squares approach (see e.g. Bastardoz et al., 2023; Sajons, 2020) or Directed Acyclic Graphs as avenues for future research to identify causal mechanisms in climate policy support research.

We appreciate the insights regarding the established links between instrument intrusiveness, fairness aspects and perceived effectiveness. We also were wary of these links and therefore had already extended the traditional pathway model to include the reviewer's suggested paths by also allowing for these concrete paths: from the instruments' personal impact- and fairness perception-relevant design features, namely type of instrument and fairness principle, to perceived effectiveness (see Figs. S1b and S3c in the supplement).

This addresses the concern that omitted theoretically plausible paths could favor the alternative model that also the literature mentioned by the reviewer raises. The results from the analyses (both regarding climate effectiveness and energy security effectiveness) showed that the alternative model still outperforms the extended, near-symmetric traditional model. We made now refer to this supplemental analysis more explicitly in the main manuscript (lines 208-209 and 247-248) and the SI (lines 54 and 76).

Regarding fully symmetrical models (where everything is exactly the same except for the order of acceptance and perceived effectiveness): because these models share the exact same covariance matrices, they produce mathematically identical fit indices. Consequently, a fully symmetrical model comparison cannot statistically differentiate between the two directions.

Moreover, thank you for the excellent reference suggestions about causal interpretations. We have included them in the respective section (lines 480-488).

Also, in line with that literature, we now use the broader empirical pattern across our findings, rather than overemphasizing fit statistics alone, to argue

	for our conclusion that the alternative pathway represents the more general mechanism. To clarify this logic, we have sharpened the text in Lines 403-408.
Second, I think the manuscript would benefit from a clearer definition of what the authors mean by policy acceptance and policy support. I can see the way the operationalized the concepts but I nevertheless think a somewhat more explicit discussion of what they are measuring and how policy acceptance differs from GCPS (and if what they call GCPS is actually support or also acceptance) would help the field as these terms are used relatively loosely throughout much of the literature. The authors do already cite a paper offering a framework/construct of how to differentiate between different constructs in climate policy support research (Kyselá et al., 2019), so it is somewhat surprising to see them not making use of the construct Kyselá and colleagues provide at all.	Thank you for raising this fair point regarding conceptual rigor. To better contextualize our terminological choices, we have on the one hand added a statement about the differentiation between acceptance and acceptability (lines 141-143). While hypothetical measures point toward acceptability, our focal instruments correspond directly to real-world policy measures already implemented in current energy and heating policy. On the other we have added lines 506-510 to the limitations section that acknowledge the scope of our operationalizations when it comes to strict, that is active, policy support. We believe these additions clarify the scope of our measures and the conceptual boundaries of our findings. Please also consider that our choice of the word "support" stems from the original survey items in German where a natural translation of "Befürworten" is "to support".
Finally, while I do think the findings on GCPS-opponents are highly relevant, they could be integrated more explicitly with research showing that highlighting the co-benefits of climate mitigation motivates climate action (see e.g. Bain et al., 2015). In my opinion, the results reported here and in papers such as the Bain et al. paper strongly suggest that we do not actually need to communicate about climate change or what a certain policy does in relation to climate change to convince people to support it. I think this point deserves somewhat more attention in the manuscript.	Thank you for this important point. We do indeed see that our results are in line with the recommendation to focus on co-benefits. The mentioned reference is therefore highly valuable and we integrated it in the discussion (lines 530-532).

Summary:

The authors use a survey experiment with German respondents to test the causal relationship between climate policy acceptance and perceived policy effectiveness. Specifically, they argue that most models of policy acceptance assume that effectiveness drives acceptance, whereas they find greater evidence for the alternative pathway: that acceptance predicts effectiveness. Additionally, they find that these results are moderated by whether respondents generally support climate policy.

Major Comments:

The potential bidirectionality of policy acceptance and effectiveness perceptions is an intriguing and interesting possibility, especially considering recent literature showing that correcting misperceptions about the effectiveness of proenvironmental behaviors and climate policy does in fact increase respondent willingness to take such behaviors and policy support. I had some concerns and questions as I read through the paper that I have enumerated below — I hope these can be helpful.

1. Introduction and Motivation.

- It would be helpful to see more motivation for why fairness and personal benefit are expected to work better than effectiveness information, and why these two factors over the many other determinants of policy support identified in the literature.
- Why is information about distributional impacts and personal benefits not also considered part of the traditional causal model, alongside effectiveness? (e.g. in the depiction of Figure 1, but also in the text) And why are they not also assumed to be impacted by policy acceptance in a similar bidirectional model to that of effectiveness? In general, the specific assumptions here seem somewhat *ad hoc* and would benefit from greater motivation make the case for bidirectionality or the alternative pathway in the case of effectiveness and not the other inputs to policy support.
- The moderation by general climate policy support (lines 81-83) was also lacking motivation — and in fact I could not quite see how it tied cleanly into the first question about effectiveness. I was also unclear about the predicted moderation directions in reading the paper. Was this study pre-registered? This would help mitigate concerns that some of these analyses seem a bit arbitrary or unmotivated. For example, in a caption I see that you assume that general policy support reduces the impact of type of instrument and increases responsiveness to distributional principles and effectiveness — however, why this is the case is not clearly laid out.

2. Experiment Design

- My biggest concern is with the design and analyses used to identify evidence in favor of the alternative causal pathway over the traditional one. In the current study, the authors manipulate policy instrument, distributional impacts, and effectiveness in a single cross-sectional study. This allows them to test the first pathway since effectiveness is randomly manipulated and the causal effect on policy acceptance can be tested. However, they do not vary policy acceptance independently of effectiveness information and then see how increased policy support increases effectiveness perceptions, or alternately include multiple waves to assess the relationship longitudinally. I think this would be a tighter test of the two causal pathways and in the current setup the ability to make causal inferences is limited (and currently overstated).
- Why is climate concern, but no other climate policy related variable, included as a control? Are results robust to its exclusion?
- The authors conclude that both pathways are at play, but that there is more evidence in favor of the alternative pathway. How should one think about the difference between SRMR=.01 and SRMR=.06. Is this a large and significant difference? How should it be interpreted given that the models have different number of paths, with the alternative having more?
- Line 246 suggests that the GCPS moderation is an exploratory analysis. Was the rest pre-registered then? Apologies if I missed this! This would be very helpful to see. Also on the GCPS, did the authors also try a continuous moderation? It was unclear to me why the median split was needed.
- The authors find that both groups show evidence for the alternative causal pathways, but that the traditional causal pathway only applies to general climate policy supporters. It was not clear to me how to make sense of this result. In general, interpretation that is specific and points to the relevant literature would be helpful.
- The survey seemed to include many modules before the experiment—such as national climate targets, energy security considerations, and the potential contribution of these instruments to both objectives through reduced fossil fuel consumption in the building sector. Might these have impacted the results? In particular, this seems to already carry effectiveness information?
- In the methods, I see that effectiveness information had four levels – level of effectiveness X type of impact (energy security or climate). This was not clear in the paper — and the motivation for this is unclear. Do the results look

different if you include the different types of impact (energy or climate) in your analyses?

- Are the results robust to including excluded participants (e.g. due to fast completion times)?
- In the paper there is a statement about how the sample does not look too far from the general population, but this is not how I would have interpreted the results of Table S1. You can show the t-tests and p-values so it is clear where there are significant differences. I don't think this is a big threat to inference given this is an experimental study, but I think it would help to be transparent about the sample.
- The effect sizes are small – and while this is attributed to the debates around this policy, it could also be due to the vignettes. Can the authors compare effect sizes relative to other studies? Or show that they differ for a policy that is not currently the topic of debate?

3. Conclusions

- Lines 301-302 need citations. Especially as some literature does find that information about effectiveness increases support.
- Lines 319-320 (“Effectiveness, in this view, is not just evaluated; it is constructed to be consistent with underlying judgments about a policy’s fairness and its impact on the self”) again raised the question for me about why we should not expect this to also be the case for perceptions of personal benefit and distributional impacts...
- On line 327, the authors state that “fairness and effectiveness cues are significantly more persuasive for those already inclined to support political climate action” yet on line 364 they state “To reach opponents and the undecided, policy design and communication must lead with fairness and the mitigation of personal costs”. To me, these seem a little inconsistent.

Minor Comments:

Throughout, there are various points where additional clarification or detail would have been helpful for me. These sentences are identified below.

- Lines 18-19, the first two lines of the introduction, do not currently read like a puzzle. Is the puzzle that the technically effective solutions are not politically feasible?
- I found the following assertion — *Bolderdijk et al.²⁰ argue, because the effectiveness of policy instruments—particularly when they are yet to be implemented—is difficult to gauge, individuals rely on other more readily available information.* — confusing because then isn't this why communicating effectiveness information might be

particularly valuable? Also, it seems that the effectiveness of climate policy instruments is often hard to assess even after implemented.

- Line 314-316 did not make sense to me – what does need refer to here?