Action-Effect Meta-Analysis

Reply to the Invitation to Revise and Resubmit

We would like to thank the editor and the reviewers for their useful suggestions and below we provide a detailed response as well as a tally of all the changes that were made in the manuscript. For an easier overview of all the changes made, we also provide a summary of changes.

Please note that the editor's and reviewers' comments are in bold while our answers are underneath in normal script.

A track-changes comparison of the previous submission and the revised submission can be found on: https://draftable.com/compare/HZyhGXGHEQNn

A track-changes manuscript is provided with the file: "PCIRR-RNR-Action-Inaction and Emotions Registered Report Meta Analysis Main Manuscript V7-G-trackchanges.docx"

Summary of changes

Below we provide a table with a summary of the main changes and our response to the editor and reviewers:

Section	**Actions taken**
Abstract	None
Introduction	CC, PR, and EC: Clarification of rationale and scope (e.g., differences between action-effect and other related but distinct effects) PR: 1) Discussed the possible differences between positive emotions and negative emotions, 2) Described the other dimensions of emotions, 3) Described issues of upward and downward counterfactuals, 4) Describe the issue regarding temporal pattern, negative emotions, and positive emotions. EC: 1) Elaborated more regarding theories and their possible links to moderators, 2) Elaborated more regarding the possible differences between positive emotions and negative emotions in action-inaction, 3) Elaborated further regarding the usage of the phrase “associated with”
Method/results	CC, and EC: Ensuring that the power analysis is tailored for multilevel main effect analyses EC, and DQ: Added that there will be another author responsible for screening and coding at Stage 2 if accepted DQ: 1) Elaborated more clearly regarding search process – i) pre-search, ii) listservs, iii) cut-off year, 2) Explained more regarding the usage of ResearchGate and Open Science Framework, 3) Added that plot digitiser tools will be used, 4) Elaborated more regarding multilevel model (used for analyses in the main manuscript), agg function (supplementary), and added that sensitivity analyses will be conducted (supplementary), 5) Added that Bayesian analyses may be conducted under certain scenarios at Stage 2 (but those scenarios are unlikely), 6) Elaborated more on funnel plot asymmetry tests EC: 1) Removed posteriori power analyses, 2) Changed the priori power analysis so that it is suitable for multilevel model, 3) Added categorical moderator analyses for temporal distance, 4) Clarified regarding effect size computations, 5) Elaborated more regarding multivariate models and added that associations between moderators will be reported, 6) Elaborated more regarding MetaForest
Supplementary	DQ: 1) Added the Template for Contacting Authors for Published and Unpublished Data on Listservs section, 2) Added analyses using agg function (with r = 0.3, r = 0.5, r = 0.7) EC: 1) Added the simulated results table for Chi Square Tests of Associations between moderators, 2) Added two-level model moderator results
Code	DQ: Added code for sensitivity analyses based on r = 0.3 and r = 0.7 with agg function EC: 1) Added code for calculation of Cohen’s d for within-subject studies given M and SD, 2) Added code for Chi Square Tests of Associations between moderators, 3) Added code for multilevel power analysis

Note. Editor: CC = Prof. Chris Chambers, Reviewers: DQ = Dr. Dan Quintana, PR = Dr. Priyali Rajagopal, EC = Dr. Emiel Cracco

Response to Editor Prof. Chris Chambers

> **Three reviewers with a range of methodological and field-specific expertise have now assessed the Stage 1 manuscript. As you will see, the evaluations are broadly positive, with the majority of comments identifying aspects of the proposal that would benefit from clarification and/or elaboration, as well as strengthening of the rationale, and ensuring tight linking between the sampling plan (power analysis) and analysis plans. Based on these reviews, we are pleased to invite a revised submission along with a point-by-point response to the reviews.**
> **As you know, the Managing Board has also been considering what level in the PCI RR bias control taxonomy is appropriate for your submission. This remains an ongoing discussion -- a unanimous position has not yet been agreed as there are arguments in favour of both positions (Level 6 *vs.* Level 4) -- but I wanted to let you know that we will reach a decision on this prior to the awarding of in-principle acceptance and will consult with you in due course. For now, there is no need to address this issue in your revised submission or response.**

Thank you very much for the positive opening note, the three helpful reviews obtained, and the valuable feedback. We appreciate the encouraging and highly constructive reviews.

We responded to each of the comments in detail below.

We understand the issue regarding level of taxonomy. We hope to see this as Level 6 for meta-analyses, we feel as though limiting meta-analyses to Level 4 would have implications for the future of meta-analyses as Registered Reports with PCIRR, and we are happy to discuss this further and provide any support in a discussion on this. We will continue with our project regardless, and will accept PCIRR's final decision.

Response to Dr. Dan Quintana

> **This is the Stage 1 of Registered Report submission describing a planned meta-analysis of the action effect literature. I would like to state upfront that I do not have experience with the action effect literature, so I cannot speak to the appropriateness of research questions in light of the prior research. Thus, I will be largely be commenting on the methodological aspects of this manuscript. Overall, this manuscript reports a comprehensive and well-considered plan for a meta-analysis. However, I have some comments that may improve future versions of this manscript.**

Thank you very much for your positive comments and very constructive suggestions. We very much appreciate your help and support.

> **"One of the most well-known effects in the action-inaction literature is the action-effect, which is the phenomenon that people imagine, associate, or experience stronger emotions for action compared to inaction" This section could be improved by providing a brief example after this sentence**

Thank you for the suggestion. We added a brief example there and referred readers to a later section for more details regarding the Kahneman and Tversky (1982) example.

> "It was first demonstrated by Kahneman and Tversky (1982) with a scenario describing two-investors who both ended up losing money following the same investment, with the main difference between the two investors being that one investor switched to that investment from a previous investment (action), whereas the other had considered switching the investment but had finally decided to stick with his original choice (inaction). Their findings were people perceived the action investor as experiencing stronger regret compared to the inaction investor, concluding that regret over negative outcomes is stronger when it involved an action decision rather than inaction decision."

> **"At the time of writing (July 2021), we identified 2466 citations of the article..." Using which database?**

We updated the citations and stated that it is based on Google Scholar.

> "At the time of writing (February 2022), we identified 2547 citations of the article (according to Google Scholar)"

> **Methods**
> ***"We conducted an initial unstructured pre-search..."* What was the purpose of this pre-search? To refine formal search strings?**

We added the following:

> "We conducted an initial unstructured pre-search to construct, test, and refine our search syntax."

> ***"and posted a notice on listservs..."* Provide an example or two of listservs. It sounds like this will be done so that the authors can be notified of possible related articles, but this kind of strategy is typically used to find unpublished studies—is this what the authors are intending?**

Great suggestion, thank you. We adopted this strategy to find both published articles and unpublished studies. We added/changed the following:

> "We conducted an initial unstructured pre-search to construct, test, and refine our search syntax. We then posted notices on listservs asking authors to alert us of possible related articles and unpublished studies: Society for Judgment and Decision Making (SJDM), European Association for Decision Making (EADM), European Association of Social Psychology (EASP), Society for Personality and Social Psychology (SPSP), and Social Psychology Network (SPN) on [Dates to be inserted in Stage 2] [Note: List will be updated in Stage 2 if changed].
>
> We provided a notice template in the Template for Contacting Authors for Published and Unpublished Data on Listservs section of the Supplementary. Systematic data collection has not been conducted for this project. There are no other unreported pre-registrations for this meta-analysis project. See Open Science Disclosures in Supplementary for details."

> ***"We validated and pre-tested the search pattern with 10 notable articles"* Was notability defined as the number of citations? Another metric?**

We added the following to clarify:

> "We validated and pre-tested the search pattern with 10 randomly selected articles out of all articles related to action-effect mentioned in this manuscript"

> ***"Third, we included both published or unpublished studies, from 1982"* State why this particular year was used as the cutoff (i.e., the Kahneman and Tversky paper)**

Thank you, we added the following:

> "We chose 1982 to be the cut-off year as Kahneman and Tversky (1982), the first study on Action-Effect, was published in 1982."

> ***"See Supplementary Materials template for contacting authors subsection."* I appreciate the comprehensiveness of including this information**

Thank you.

> ***"We set up a project on ResearchGate and added all identified articles as references, where possible, to notify authors about this project, and to provide an open-access list of available studies"*** **Why was ResearchGate chosen for this? I'm unsure about the longevity of the platform (although it seems appropriate in the short term as one way for notifying authors, assuming they actively use the platform). In other words, I think ResearchGate is useful as one approach for contacting promoting the meta-analysis, but another platform with (more or less) guaranteed longevity (e.g., OSF) should be used for providing a list of studies**

To be clear, the use of ResearchGate is in addition to sharing all that we're doing on the OSF, and in addition to us using the obvious known channels. We noted that we are using this to simply connect to people in one more way.

We realized that we may have needed to make clearer our use of the OSF in the manuscript, which is ofcourse our main tool for open-science. Materials/data/code storage wise, our solution is the OSF which we also used in this project in our initial submission (e.g., list of studies, materials, code, outputs, etc.).

ResearchGate, for better or worse, seems to be where some scientists connect. It has existed almost as long as Facebook and Twitter (since 2008) and is just a common channel that some academics use for connecting to other researchers and research and keeping track of new papers and developments.

Though we have our own reservations regarding the platform, as we do with many other social media platforms, but in the past, we found it to be useful in some contexts in connecting with hard-to-reach researchers. It also allows for some social-network features that are typically lacking for academics, in keeping track of people's work and related articles/projects/authors.

We added clarifications regarding this in the following:

> "We set up a project on ResearchGate and added all identified articles as references, where possible, as another method of trying to reach and notify authors about this project. We used ResearchGate and OSF to keep track related of articles, projects, and authors, and to provide an open-access list of available studies (links: [insert link]). The OSF project was also used to store our the Datafile (with all articles and studies), the RMarkdown code and outputs, and preprints."

> ***"If we were not able to obtain the required statistics, we excluded the articles"*** **Have the authors considered using plot digitiser tools to extract data if the raw data is not reported in text?**

Yes, this is one of the ways we use to extract data from articles, and we appreciate the call to add that transparently in the text. We added the following:

> "In cases of missing statistical data, we first attempted to contact the authors. When plots are provided, we would attempt to use plot digitizers tools such as Web Plot Digitizer (Drevon et al., 2017), GraphClick (Arizona Software Inc., 2010), DataThief III (Tummers, 2006; also see Flower et al., 2016 for reliability and validity of GraphClick and DataThief III), or digitize R package (Poisot et al., 2016) to extract data from plots. We will document the methods used for obtaining required

statistics in the coding sheet. When we were not able to obtain the required statistics or raw data from the article or through emailing the authors, we excluded the studies, even if the studies met all other search criteria. We excluded all non-experimental studies."

"When we could not reach agreements on certain inclusion/exclusion, a moderator would make the final decision" **There is two authors on the study, has a moderator been identified yet?**

Thanks for bringing this up. Yes, the corresponding author will supervise the process but will not be part of the coding procedure, and so for the coding we plan to add a co-author to screen and code studies to join the lead author in Stage 2 after receiving in-principle acceptance. We will update this section with the exact details in Stage 2, but for now we added the following clarification to simulate what it might look like:

"A collaborator ([insert name of coauthor by Stage 2]) joined the team at after receiving an in-principle acceptance to help with search, screening, and coding together with the lead author [will be updated in Stage or 2 if recruitment was not successful]. When the two coders (Siu Kit Yeung and [insert name of coauthor by Stage 2]) could not reach an agreement on certain inclusion/exclusion, the corresponding author (Gilad Feldman) would make the final decision as a moderator."

If we are unsuccessful in our recruitment, the lead author will conduct all coding, and the corresponding author will verify the coding, a method the corresponding author employed in other published meta-analysis work with guided students (e.g., Kutscher & Feldman, 2019). Regardless, we will update this after Stage 2.

"All statistics were converted to Hedges g effects." **This should be "Hedges' g"**
"Chi-square is Converted to Cohen d with chies function of compute.es v0.2-5 (Re, 2020)." **This should be "Cohen's d"**

Thank you, changed.

It appears the authors are planning two different approaches to account for effect size dependencies: Three level multivariate models and effect size aggregation (via the 'agg' function). From what I can gather, three level models will be used for moderator analysis and aggregation will be used for main-effect analysis, is that correct? Why not use three level models for all analysis, considering that you will lose some precision with effect size aggregation? I'm not entirely opposed to effect size aggregation, but I just want to better understand the reasoning here. I may have missed something here, but this description of how effect size dependencies will be dealt with is currently unclear ***"and assumed the correlation between the measures to be 0.5..."*** **For a sensitivity analysis, I would choose two other correlations to make sure that the conclusions don't differ according to the assumed correlation.**

We agree that it is better to adopt multivariate three-level models for both main analyses and moderator analyses, which is what we did and reported in the main manuscript. We adopted the agg function but simply reported these results in the supplementary Table 8 as robustness checks. Our experience has been that the two yield very similar results.

Thank you for the very useful sensitivity analyses suggestion. In our experience the correlations between the measures have very little to no impact on the analyses, yet we 100% agree that it would be best to communicate this clearly and transparently in the manuscript.

We found very similar results (tiny difference of <0.02 in *g*) assuming correlations to be 0.3, 0.5 or 0.7, or using multivariate three-level model.

We therefore added the following:

> "Apart from conducting multivariate three-level analyses for both main effects and moderator analyses, we adopted agg function (MAd package, Re & Hoyt, 2014) for effect aggregation of two-level models as there might be a few studies with multiple dependent effect sizes. In general, assuming the correlation between the measures to be 0.5 (as most studies did not report correlations, Wampold et al., 1997) is the common practice but we also conducted sensitivity analyses assuming correlations to be 0.3 and 0.7. We reported results based on agg function in the supplementary Table 8 and reported results based on multivariate three-level model in the main manuscript. As suggested by the reviewer Dan Quintana, the aggregation method may result in loss of precision, so we prefer using the multivariate three-level model to account for effect size dependencies. That said, the results using agg function, assuming correlations to be 0.3, 0.5, or 0.7, are very similar to results using multivariate three-level model."

> ***"We stated our planned preferred effect size adjustment methods under different scenarios in Supplementary Table 5"*** **I appreciate the comprehensiveness of this approach**
> **A big strength of this article is the use of simulated data in the results section**

Thank you.

> ***"We conducted posteriori power analyses with Tiebel (2018) tool"*** **Should this be "Tiebel's tool"?**

Thank you. We decided to remove the section on posteriori power analyses given feedback that it is not informative, and it may lead to confusion and misunderstanding.

Please see our reply to reviewer Dr. Emiel Cracco on this point.

> **The authors should also consider robust bayesian meta-analysis, which addresses many limitations of frequentists approaches to publication bias (e.g., how to interpret a non-significant publication bias test, dealing with conflicting conclusions from different publication bias tests). See this primer on the RoBMA R package from Bartos et al https://psyarxiv.com/75bqn/**

Thank you for the suggestion and the helpful reference. This seems to be a rather new approach, but one that might be useful here, and we would be glad for the opportunity to see how this might help further improve our planned meta.

Based on our knowledge of the literature and experience with several high-power pre-registered replications of effects in this literature, it is likely that the action-effect is a robust and meaningful effect in general.

That said, we like this suggestion plan to conduct Bayesian analyses under certain (even if unlikely) conditions at Stage 2:

> "[Note 2: As suggested by the reviewer Dan Quintana, we intend to conduct Bayesian Analyses in cases of disagreements of publication bias tests mentioned above. This means that for experimental and/or comparison studies, if the agreement rate between adjustment tests falls below 5/6, we will conduct Bayesian Analyses. For example, a substantial disagreement occurs when four tests find support for an effect with confidence intervals not overlapping with null whereas two tests fail to find support for an effect. In such scenarios, we will conduct Bayesian analyses. We expect this to be very unlikely. Based on our knowledge of the literature, even though action-effect is weakened in some conditions and reversed in a few conditions, action-effect is generally a robust and replicable effect.]"

> ***"With rank correlation tests and Egger's regression tests, which are based on funnel plot asymmetry (see Figure 3 for funnel sunset plot), we found support for evidence of publication bias"*** **Funnel plot asymmetry approaches are technically tests of small study bias, which encompasses publication bias but can also include other sources of bias**

Great point. We clarified in the following:

> "Funnel plot asymmetry may also be due to other factors, such as low methodological quality in studies with smaller sample sizes and inflated effect sizes, artefactual causes or pure chance, not always due to publication bias (Egger et al., 1997; Page et al., 2021)."

> "With rank correlation tests and Egger's regression tests, which are based on funnel plot asymmetry (see Figure 3 for funnel sunset plot), there appears to be support for evidence of publication bias, but we are not uncertain as there are other possible causes of funnel plot asymmetry."

> ***"We set this threshold arbitrarily, as no study has compared performances between MetaForest, traditional mixed effects two-level model, and traditional multivariate three-level model given different numbers of studies. We would appreciate constructive feedback from reviewers."*** **I think this is a reasonable threshold, as long as the authors are explicit in the paper that this is arbitrary, as no comparison studies exist.**

Yes, it is essential to be transparent that this is an arbitrary threshold.

Thank you very much for the very helpful feedback.

Response to Dr. Priyali Rajagopal

> **The proposed meta-analysis is well designed and deals with an interesting topic area – the action effect. The authors do a nice job of summarizing the current literature and proposing a set of moderators to explore the action effect. A few suggestions are noted below.**

Thank you very much for the positive opening note and the constructive feedback.

> **The authors articulate their objective as focusing on action-inaction asymmetries with respect to two specific outcomes – emotions and counterfactual thoughts. Some justification or reasoning for the selection of these two types of outcomes would be useful for the reader.**

Thank you for the suggestion and feedback encouraging us to elaborate further. We added the following:

> "In this meta-analysis, we investigated action-inaction asymmetries of emotions and counterfactual thoughts. We included counterfactual thoughts as those have been commonly studied with and have been shown as being associated with emotions. We focused on these dependent variables as these were the first initial demonstrations of action-inaction asymmetries and based on our experience in conducting replications in this domain these dependent variables seem to be the most commonly studied in the action-inaction domain.
>
> The current project will not cover action-inaction asymmetries regarding moral judgments and decisions. Some of those asymmetries have already been investigated in other meta-analyses, such as in the meta-analysis on omission bias by Yeung et al. (2021). Omission bias and action-effect are related but distinct, with different dependent variables, moderators of interest, and theoretical paradigm (see Feldman et al., 2020 and Yeung et al., 2021 for more details). Norm theory is a framework that has been used to try and align omission bias with action-effect (Feldman et al., 2020) yet this remains to be studied further, and there are other competing theoretical explanations for both effects (Decision Justification Theory, Action-Effort Link, etc.) that seem more suitable for explaining one effect yet not the other. "

> **Within positive and negative emotions (Table 2), are there any expected differences? Research has found that emotions can vary on many dimensions (e.g., arousal, control etc.) even when they are similarly valenced (positive or negative). Hence, within the context of the action effect, will all positive or all negative emotions respond similarly? Will the authors test for differences between specific emotions if there is sufficient data?**

Thank you. We agree that discussing this in the introduction can help guide readers about our expectations.

There are very limited number of studies in the literature on joy and regret investigated together with mixed findings and we therefore consider this direction exploratory. The earlier classic findings suggest similar effects for joy and regret, yet the more recent findings indicated differences in strength of action-effect for positive and negative emotions.

As far as we are aware the vast majority of the action-effect literature has been focused on regret, with few studies on joy, but no other emotions that vary on degrees of arousal, so we are not sure what to expect regarding other emotions at this stage. We would like to try and keep the scope and aim of this meta-analysis narrow and well-defined, yet we added that suggestion as a direction for future research to be discussed in the discussion at Stage 2 (see note in the discussion section).

We added the following:

> "A recent study by Fillon et al. (2022b), one of the very few studies that investigated both regret and joy, found stronger effects for regret and weaker effects for joy, possibly because of negativity bias (bad seems to have stronger impact than good). Their findings were different than the more classic Landman (1987) who found similar effects for joy and regret. We therefore conclude mixed findings and this direction as exploratory. The vast majority of the studies we know from the literature investigated action-effect for negative emotions, with very few studies for positive emotions (e.g., Fillon et al., 2022b; Landman, 1987). If we find sufficient studies (minimum: 5) for action-inaction asymmetries in positive emotions, we will conduct exploratory moderator analyses comparing positive emotions studies and negative emotions studies.
>
> Exploratory hypotheses (given sufficient studies): Action-effect is stronger/weaker for negative emotions compared to positive emotions."

While the authors focus on the numbers of counterfactual thoughts, it may be helpful to consider the type of counterfactual too (e.g., upward vs downward).

Thank you.

Similarly to the above point about joy, we expect limited (if any) studies for downward counterfactuals in the action-effect literature. That said, if we do come across 5 or more studies for both types of counterfactual thoughts, we will conduct exploratory analyses.

We added the following:

> "Another exploratory direction is regarding possible differences between upward counterfactuals and downward counterfactuals. To the best of our knowledge, the vast majority of studies measured upward counterfactuals, thought there might be studies on downward counterfactuals. Therefore, if we find at least 5 studies on action-inaction asymmetries on downward counterfactual thoughts and at least 5 studies for upward counterfactual thoughts, then we would conduct exploratory analyses for a moderation of upward versus downwards. We do not have any directional expectations for findings."

> **Will H2 hold for both positive and negative emotions? The research cited for supporting the moderating role of temporal distance seems rather specific to regret alone – why would the authors expect it to replicate for all emotions?**

We are honestly not sure. As far as we know there are no studies for temporal pattern for positive emotions or any study for positive emotions action-effect in the long-run (let us know if there are studies on this). It is difficult to develop expectations or hypotheses regarding this issue.

We added the following clarification:

> "To the best of our knowledge, the vast majority of studies for temporal pattern of action-effect measured negative emotions but not positive emotions, and we therefore do not have any a-priori expectation regarding such temporal pattern for positive emotions."

Response to Dr. Emiel Cracco

> **Let me start by saying that literature on the action effect is not one I am familiar with. As a result, I cannot comment much on the literature review. From an outsider's perspective, I thought the literature review was clear and provided a good overview of the different theories and findings in the field**

Thank you for the positive opening note and the constructive feedback.

> **On P. 6, the authors write that they focus on emotions and counterfactual thoughts and not on other action-inaction effects, which they argue are different. It was not clear to me, however, why these other effects were different. Are they not explained by the same theories? If they are explained by the same theories, then why not include them?**

Thank you. There are a number of challenges here, and we tried to be careful to not wander into this potential minefield of issues when planning and writing this meta-analysis.

First, there is a long list of action-inaction effects, with separate literatures that do not interact much with each other. There is also the related issue of how action and inaction are defined, and some of the confusion and confounds in the action-inaction literatures (see Feldman, et al., 2020). In addition, the literature on action-inaction is based on the judgment and decision-making paradigm mainly focused on phenomena, where the effect was first demonstrated, then repeatedly interrogated, and only after an accumulation of evidence several competing broad theories were proposed to try and explain these effects. There seems to be no clear winner, and no one theory has been able to explain all action-inaction effects, likely because of the lingering issues.

The one that seemed to come closest, was norm theory, but the post-hoc generalized theory far extends beyond action-inaction, is much broader in scope, and somewhat lacking in specificity, with ongoing debates regarding its testable hypotheses and how those relate to the action-effect literature. In our meta-analysis we were therefore phenomenon focused, with the moderators focusing on observed effects we identified in the literature.

To explain a bit more about the differentiation from other action-inaction effect, we added the following:

> "In this meta-analysis, we investigated action-inaction asymmetries of emotions and counterfactual thoughts. We included counterfactual thoughts as those have been commonly studied with and have been shown as being associated with emotions. We focused on these dependent variables as these were the first initial demonstrations of action-inaction asymmetries and based on our experience in conducting replications in this domain these dependent variables seem to be the most commonly studied in the action-inaction domain.
>
> The current project will not cover action-inaction asymmetries regarding moral judgments and decisions. Some of those asymmetries have already been investigated in other meta-analyses, such as in the meta-analysis on omission bias by Yeung et al. (2021). Omission bias and action-effect are related but distinct, with different dependent variables, moderators of interest, and theoretical paradigm (see Feldman et al., 2020 and Yeung et al., 2021 for more details). Norm

> theory is a framework that has been used to try and align omission bias with action-effect (Feldman et al., 2020) yet this remains to be studied further, and there are other competing theoretical explanations for both effects (decision justification theory, action-effort link, etc.) that seem more suitable for explaining one effect yet not the other."

We have already completed and published the omission-bias meta-analysis, where we differentiated between the harm/morality focused omission-bias effect from the emotions focused action-effect. Our replication of the omission bias with an exploratory extension attempting to tie omission bias to the action-effect (Jamison et al., 2020) proved very tricky with findings that we and the literature still need to sort out. The findings were opposite to our expectations. We therefore still view the two literatures as separate and focused on the part of the literature yet to be summarized and aggregated.

> **Most moderators appeared rather atheoretical. The authors write on P. 10 that current theories are imprecise and therefore difficult to test, so I assume this explains why. Nevertheless, I was wondering if some of the theories do make different predictions about some of the included moderators? In addition, I was wondering if it would be possible to directly test norm theory by conducting a meta-analytical correlation between counterfactual thoughts and positive/negative emotions, where the latter forms a proxy for regret?**

Thank you for the suggestion. As we noted above, given the complex literature, we did aim to focus on the phenomenon without going in-depth into the theoretical paradigms that have been suggested to explain the action-effect.

We do not view the suggested theories as contradictory. We added the following clarification:

> "However, these theories are not necessarily contradictory and may jointly explain the phenomenon. An action may be regretted more because of several different reasons - it is perceived as abnormal, it is perceived as more blameworthy or responsible, it is perceived as wasted effort, or it is perceived as unjustified. It is possible all these factors contribute to action-effect in some situations, but one of these accounts may be more relevant in some situations. That said, very limited studies have compared these different contributing factors of action-effect and there is much need for more research on the intersection of the theoretical paradigms."

Regarding the connections of theories to moderators, in the prior outcome section, we added the following regarding its link to Decision Justification Theory:

> "This is in line with decision justification theory (Connolly & Zeelenberg, 2002), arguing that decisions (action given negative prior outcomes) that are better socially justified are regretted less."

In our prior version, we already mentioned norm theory in normality as moderator section.

Moreover, in the meaning of action-inaction section, we added that the change or no change definition "…which seem to be more related to norm theory and the concept of normality, see Yeung & Feldman, 2022)", as "changing a prior decision" can be considered as "deviating from past behavioral norm".

Regarding correlational analyses between counterfactual thoughts and regret, we would prefer to keep the scope of this meta-analysis specific to the action-effect, focusing on action and inaction. We wish to avoid adding other elements looking at broader theories looking at the link between emotions and counterfactuals, which may confuse readers and shift attention away from our main aim. These directions are important and deserve a much broader investigation that examines all links going beyond the action-effect literature.

If needed, we are willing to reconsider this given clear editorial guidelines.

> **Why not test H1a and H1b in a single model, where "type of emotion" is included as a moderator? This will give the authors more power and would have the benefit that it also**
> **allows them to test if the action-inaction effect might be stronger/weaker for negative than for positive emotions.**

Thank you for the suggestion. We kept H1a and H1b separate, yet will also test overall action-effect in a single model, comparing effects of positive emotions and negative emotions. Please see below.

> "We expected the evidence to be in support of action-effect with a positive effect meaningfully different from null (null not included in confidence intervals) ($H1_{overall}$)."
>
> "A recent study by Fillon et al. (2022b), one of the very few studies that investigated both regret and joy, found stronger effects for regret and weaker effects for joy, possibly because of negativity bias (bad seems to have stronger impact than good). Their findings were different than the more classic Landman (1987) who found similar effects for joy and regret. We therefore conclude mixed findings and this direction as exploratory. The vast majority of the studies we know from the literature investigated action-effect for negative emotions, with very few studies for positive emotions (e.g., Fillon et al., 2022b; Landman, 1987). If we find 5 studies for action-inaction asymmetries in positive emotions, then we will conduct exploratory moderator analyses comparing positive emotions studies and negative emotions studies.
>
> Exploratory hypotheses (if 5 studies or more on positive emotions): Action-effect is stronger/weaker for negative emotions compared to positive emotions."

> **Temporal distance: I have a sense that coding this in terms of # years will result in a very skewed distribution with a number of large outliers on the right side of the distribution. This could potentially bias this analysis. How will the authors deal with this? More generally, how realistic is the assumption that the effect of temporal distance is linear? For example, do we really expect the difference between a 1-week or a 2-week interval to matter? If not, perhaps it makes sense to code this variable categorically. Could the authors comment on this?**

Yes, it is possible that coding this in terms of number of years would result in a skewed distribution, and to the best of our knowledge, there are no studies examining whether the effect of temporal distance is linear or not, so we are not sure.

We agree that coding this variable categorically is also a reasonable option. We plan to report both categorical analysis and continuous analysis.

We added the following (based on simulated data):

> "Additionally, we also conducted analyses with temporal distance as a categorical variable. Fifteen studies with current or events with lower temporal distance (within a year) had a mean effect of g = -0.26, CI [-0.65, 0.14] but confidence intervals overlap with null. Eleven studies with events that are more than 1 year ago but not lifetime had a mean effect of g = 0.32, CI [-0.06, 0.70] but confidence intervals overlap with null. Seven studies measuring emotions or counterfactual thoughts for major lifetime events had a mean effect of g = 0.46, CI [-0.31, 1.24]. With multivariate three-level model, we found support for a moderating effect of temporal distance. Studies with more distant events had a larger effect."

> **The screening procedure wasn't entirely clear to me. P. 24 mentions that the lead author will screen the papers, but P. 26-27 seem to suggest that screening will be done together**

Thank you for this feedback.

We planned to add an additional collaborator as a co-author who will screen the papers with the lead author at Stage 2. We added the following:

> "A collaborator ([insert name of coauthor by Stage 2]) joined the team at after receiving an in-principle acceptance to help with search, screening, and coding together with the lead author [will be updated in Stage 2 or if recruitment was not successful]."

> **The effect size computation is potentially problematic. Cohen's d can be calculated in different ways for repeated measures designs (Lakens, 2013) and it's not entirely clear how the authors will calculate it in the different scenarios they identify. Based on my reading, I fear they might be collapsing different types of cohen's d for the repeated measures studies. The formula based on the t-test mentioned in Table 4 suggests that they will calculate dz, which corrects the SD for the correlation between measures. How they will calculate cohen's d from descriptive information in repeated measures designs is, however, not clearly described and if I understand their code correctly, it suggests that cohen's d will be calculated there as if it were a between-subjects study, therefore not correcting the SD for the correlation between measures. I think the authors should ensure that cohen's d is always calculated in the same way. Given that both within- and between-subject studies are included, this means that they should always calculate cohen's d without correcting the SD for the correlation between measures (Lakens, 2013). This is especially important when comparing within- and between-subject studies (H3), because otherwise any difference can be trivially attributed to the fact that in repeated measures designs, cohen's d was calculated differently.**

This is a good point, and this is an issue typically not addressed well in meta-analyses. We are not aware of implemented best-practices directly addressing this issue, aside from being fully transparent about all decisions and calculations. For whatever it is worth, our experience regarding those is that the differences between the different calculations are for the most part not very large, yet this may depend on various study-related factors (sample size, design, etc.)

We revisited Lakens (2013) and recognized some key differences between different types of Cohen's d. Lakens (2013) stated that "Cohen's d_z is only rarely used in meta-analyses, because researchers often want to be able to compare effects across within and between-subject designs." (p. 4)

An alternative solution suggested by Lakens (2013) for within-subject studies, is calculation of d_{av} (similar to d_s in general). However, that is only possible if *M* and *SD* are available, and possibly the correlations between the dependent measures.

We do plan to try and obtain raw information from the authors if they are not provided in the original article. If *M* and *SD* are not provided, we would calculate d_z based on t-statistics and sample size. We understand this is not always ideal. However, we believe this is a reasonable solution given insufficient information in some studies. We are making all our effects and calculations transparent, so

Furthermore, we added code (see our latest revised RMD file: Action-effect-meta-syntax-markdown) and information in the main manuscript for calculations of Cohen's *d* for within-subject studies based on *M* and *SD*, with the MOTE package (Buchanan et al., 2019).

We clarified the above issues in Table 4:

> **"**1) For between-subject studies, if the original studies only provided *M* and *SD*, we calculated the Cohen's d_s with escalc function of metafor (Viechtbauer, 2010). For within-subject studies, if the original studies only reported *M* and *SD*, we calculated Cohen's d_{av} using d.dep.t.avg function of MOTE v1.0.2 (Buchanan et al., 2019). According to Lakens (2013), d_{av} is more similar (relative to d_{rm}) to d_s under most situations.
> 2) Between-subject t-statistics was converted into Cohen's d_s with esc_t function of esc v0.5.1 (Lüdecke, 2019), whereas within-Subject t-statistics is converted into Cohen's d_z with Lakens (2013) Formula 7: t-statistics/sqrt(n). That said, as Lakens (2013) mentioned, d_z is less preferred in meta-analyses especially when researchers need to compare effects between within-subject studies and between-subject studies. We planned to obtain *M* and *SD* to calculate d_{av} (more similar to d_s), a suggested solution by Lakens (2013), from authors for within-subject studies if t-statistics is provided in the article but *M* and *SD* are not provided. If we couldn't obtain *M* and *SD*, we would calculate d_z, with t-statistics and sample size, a less ideal solution."

Building on the above, it's not clear to me how the cohen's d calculated from binary choices relates to the cohen's d calculated from continuous variables. As mentioned above, the authors should ensure that cohen's d always means the same thing across studies. For this reason, I also think it is not a good idea to use the cohen's d reported in the analyzed papers (as the authors propose on P. 32), because it will not always be clear which type of cohen's d is reported.

For binary choices studies, we calculated effect sizes for comparison studies separately (from experimental studies). We calculated Hedges' g (converted from Cohen's d) as this allows us to compare effects of different designs (within-subject studies, between-subject studies, comparison studies). We added the following:

"All statistics were converted to Hedges' *g* effects[1], as this allowed for a generalized comparison of effect sizes between studies of different designs (see similar methods by Fillon et al., 2020; Jachimowicz et al., 2019)."

Regarding the usage of Cohen's d reported in the analyzed articles, we understand our concerns. After some further considerations, we made the following changes:

"In most cases if *M, SD* and sample sizes are given, for between-subject studies, we analyzed using d_s, whereas for within-subject studies, we analyzed using d_{av} as they are more comparable to d_s (see our Table 4, Lakens, 2013). However, if *M*, *SD*, and sample sizes are not given, we either calculate d_z based on the t-statistics and the sample size, or we use the reported Cohen's d from the original study. These are not ideal solutions, but these solutions are likely reasonable estimates given insufficient information."

We believe that in meta-analyses with different kinds of designs and different kinds of information available, it is not **possible** to ensure Cohen's d or Hedges' g always mean**s** the same exact thing across all studies. This sometimes depends on the type of statistical information available (**e.g.,** *M*, *SD*, Cohen's *d* only, t-statistics, proportions of choices, often limited in older studies), which may vary between studies. We can only provide rough estimate given limited available information.

We added the following in footnotes:

"We note that the formulae of Hedges' *g* or Cohen's *d* vary among different studies. In a meta-analysis, it is not practical to ensure the effect size measure of all studies having the same meaning or formula across different types of designs. We can only roughly estimate Hedges' *g* of different studies based on available information (which may be *M*, *SD*, t-statistics, Cohen's *d*, Chi-Square statistics, proportions of choices) which may be limited in some studies and different across studies."

The statistical approach could use some more explanation. For example, how do three-level models correct for confounding among moderators? Relatedly, would it make sense to report the relationships between the moderators to assess confounding? (e.g., Hofmann et al., 2010).

The *multivariate* three-level model accounts for confounding relationships among moderators. This can be done in metafor: https://www.metafor-project.org/doku.php/analyses:konstantopoulos2011, which We referred to Lipsey (2003) in our manuscript, and added below to clarify:

"Furthermore, we used three-level models (with metafor, Viechtbauer, 2010) to account for dependencies of effect sizes within the same article (Cheung, 2019). The three-level models are also multivariate, accounting for confounding relationships between moderators(Lipsey, 2003),

[1] We note that the formulae of Hedges' *g* or Cohen's *d* vary among different studies. In a meta-analysis, it is not practical to ensure the effect size measure of all studies having the same meaning or formula across different types of designs. We can only roughly estimate Hedges' *g* of different studies based on available information (which may be *M*, *SD*, t-statistics, Cohen's *d*, Chi-Square statistics, proportions of choices) which may be limited in some studies and different across studies.

since some moderators (e.g., hypothetical vs real life experience and temporal distance) may be related to each other."

Thank you for the suggestion of reporting the relationships between moderators. Regarding the relationship between moderators, we plan to conduct Chi-Square tests. We added a table for Chi Square Test of Association between moderators in Supplementary Table 7 (p. 12). We added the following in the main manuscript.

"We conducted chi-square tests for the associations between moderators (referring to Hofmann et al., 2010; Lipsey, 2003 Table 2). This is reported in the Supplementary Table 7."

I usually use RVE to deal with effect size dependence and so am not very familiar with threelevel models, but if my understanding is correct, three-level models only deal with "hierarchical dependence", not with the type of dependence arising from the same sample providing multiple effect sizes (e.g., a study reporting different measures of negative emotion; Tanner-Smith et al., 2016). This latter type of dependence strikes me as more important than the hierarchical dependence. How common do the authors estimate multiple effect sizes from the same sample will be? If common, perhaps it makes sense to use RVE (Tanner-Smith et al., 2016) instead of averaging together those effect sizes as proposed now on P. 32.

Thanks for the suggestion. Since majority of studies, based on numerous studies we read, only measured one negative emotion (but not positive emotions or counterfactual thoughts, or different measures of negative emotions), it is more likely that a study only has one effect size but not multiple effect sizes (even though there will be a small proportion of studies with two or more effect sizes), so we believe RVE is not needed. Also, multivariate multi-level models can deal with dependence within the same article, some articles have 2 or more studies) and possible confounding relationships between moderators.

The authors report a power analysis, which is great, but I was wondering whether their power analysis is appropriate for the multivariate three-level models they aim to fit.

Thank you for the reminder. We adjusted the power analysis method so that it is tailored for three-level models we aim to fit. Please see the following:

"We conducted a priori-power calculation based on the lower confidence intervals and lower limits of our estimates above (k = 20, sample size = 80, g = 0.2) with code developed by Vembye et al. (2022) for multi-level meta-analyses. The expected estimated power is 76.3% given the above estimations. We expect statistical power for both comparison studies and experimental studies (assuming the effect sizes of comparison studies and experimental studies are both 0.2 and above, with 20 or more samples for both types of studies) to be adequate. We note that such estimations are conservative, and the power is likely higher. For example, the statistical power would be >99.9% if we assume k = 35, sample size = 140, g = 0.5, a more optimistic estimation."

> **I didn't understand the reported posteriori power analyses. What effect sizes are these based on? How can there be high posteriori power for a non-significant effect size?**

Thank you, that is valuable feedback.

After further consideration, we agree, and removed the posteriori analyses.

> **Given that we have data to support when one type of publication bias correction is better than another, why not report the best type given the parameters of the data and report the other corrections in supplementary material? Reporting all types of correction next to each other gives the impression that they are all equally good, which the authors themselves say is not the case.**

We thought it best to report all tests in the main manuscript to be more transparent. Our current plan is to report all publication bias results, but to highlight and clearly specify which approach seems the most relevant method given the information available at Stage 2.

We added the following:

> "3) We will highlight and clearly specify the most relevant approach given information available by Stage 2 and elaborate further in the discussion section."

We are happy to make further modifications, if given clear editorial guidelines.

> **The authors request feedback on when to use random forests. Given that there is no research on this, I personally think the arbitrary threshold they propose is reasonable. Alternatively, if that makes sense, they could run a power analysis to assess the power they would have with a three-level model and then use a cut-off based on the outcome of this analysis.**

We are not familiar with packages to aid with statistical power of moderator analyses of *multivariate three-level* models, and were unable to identify those. We would have appreciated some citations or links to such resources, or to manuscripts and/or published articles that employed such techniques.

There is a growing literature employing random-forests. We implemented the number of studies threshold as the simplified approach to this issue. That said, if such packages or related work are published by Stage 2 or if provided with additional information on this point, we would be happy to revise that.

> **Why include 2-level model results in the moderator section?**

Yes, we agree, these were only meant as robustness and error checks, expecting some alignment between the 2 and 3 level models.

We moved the 2-level model results to the supplementary (Table 5 of the supplementary). We will only report three-level model results in the main manuscript only. If, as we expect, the two are very similar and there is no value in the 2-level model then we will only provide it with the code+outputs and remove that from our manuscript altogether.

> **Metaforest is not a very common method (yet), so it would be helpful if the authors could provide some more guidance in interpreting the output. For example, they write "the main model indicator, R-squared (R-OOB) was -0.02". What does this mean?**

Thank you, great point. We added the following. This is based on simulated data so R-squared (R-OOB) is negative.

> "The negative value means that the model is overfitting with inclusion of some noise predictors (van Lissa, 2017)."

We also added this regarding variable importance value:

> "A positive variable importance value implies that the variable is a meaningful moderator whereas a negative variable importance implies the variable is not a meaningful moderator."

> **The authors often seem to drop the article when speaking of the action-inaction effect. I found this a bit awkward to read.**

We followed the way this is commonly referred to in many JDM and social-psychology articles which did not use "the" when naming the phenomenon. We consider this a matter of personal preference.

We kept it the same way in the revision but would be happy to revise this given clear editorial guidelines.

> **Table 2: I don't really understand what the authors mean with "associated with". That is, I have difficulties relating the "description" to the "term" here.**

We added "3) "Action is associated with stronger emotions than inaction" means that most people perceive and/or experience stronger emotions acting compared to not acting." In Table 2 (p. 18 of main manuscript) to clarify. I hope this is clear enough. Let us know if you or the editor have better suggestions.

> **P. 22: the authors refer to table 1, but it should be table 3.**
> **P. 41: experimental studies → comparison studies?**

Thank you for going through our manuscript carefully. We changed.

> **P. 44: "We recognize that the median power of studies is 12.7%" → of which studies?**

Thanks, we meant comparison studies. However, we later removed the section on posterior power analyses.

> **P. 50: "Nine studies with between-subject design had a positive mean effect" → this sounds as if there were nine studies that had a positive effect. I would rephrase throughout the results section.**

Great suggestion, thanks. We fixed throughout the results section. For example:

“Nine studies with between-subject design had a mean effect of *g* = 0.43, CI [0.01, 0.84]” (p. 57 and p. 58 of main manuscript)

We really appreciate your patience in going through our manuscript, and very constructive comments with helpful references. We learnt a lot revising based on your review.

References

Althouse, A. D. (2021). Post hoc power: not empowering, just misleading. *Journal of Surgical Research*, *259*, A3-A6. https://doi.org/10.1016/j.jss.2019.10.049

Arizona Software Inc. (2010). GraphClick 3.0.2. http://www.arizona-software.ch/graphclick/

Barreiro-Hurle, J., Espinosa-Goded, M., Martinez-Paz, J. M., & Perni, A. (2018). Choosing not to choose: A meta-analysis of status quo effects in environmental valuations using choice experiments. *Economía Agraria y Recursos Naturales-Agricultural and Resource Economics*, *18*(1), 79-109. https://doi.org/10.7201/earn.2018.01.04

Buchanan, E. M., Gillenwaters, A., Scofield, J. E., & Valentine, K. D. (2019). MOTE: Measure of the Effect: Package to assist in effect size calculations and their confidence intervals. *R Package Version 1.02, 2019*

Cheung, M. W. L. (2019). A guide to conducting a meta-analysis with non-independent effect sizes. *Neuropsychology Review,* 1-10. https://doi.org/10.1007/s11065-019-09415-6

Connolly, T., & Zeelenberg, M. (2002). Regret in decision making. *Current Directions in Psychological Science, 11*(6), 212-216. https://doi.org/10.1111/1467-8721.00203

Drevon, D., Fursa, S. R., & Malcolm, A. L. (2017). Intercoder reliability and validity of WebPlotDigitizer in extracting graphed data. *Behavior modification, 41*(2), 323-339. https://doi.org/10.1177/0145445516673998

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629-634.

Feldman, G., Kutscher, L., & Yay, T. (2020). Omission and commission in judgment and decision making: Linking action-inaction effects using the concept of normality. *Social and Personality Psychology Compass*. https://doi.org/10.1111/spc3.12557

Fillon, A., Kutscher, L., & Feldman, G. (2021). Impact of past behaviour normality: meta-analysis of exceptionality effect. *Cognition and Emotion*, *35*(1), 129-149. https://doi.org/10.1080/02699931.2020.1816910

Fillon, A., Strauch, L., Evans, A. M., van de Calseyde, P., & Feldman, G. (2022b). *Inaction agents are preferred and perceived more competent than action agents in risky situations: Role of expectations and social norms.* [Manuscript in Preparation]

Flower, A., McKenna, J. W., & Upreti, G. (2016). Validity and reliability of GraphClick and DataThief III for data extraction. *Behavior Modification, 40*(3), 396-413. https://doi.org/10.1177/0145445515616105

Hoenig, J. M., & Heisey, D. M. (2001). The abuse of power: the pervasive fallacy of power calculations for data analysis. *The American Statistician*, *55*(1), 19-24. https://doi.org/10.1198/000313001300339897

Jachimowicz, J. M., Duncan, S., Weber, E. U., & Johnson, E. J. (2019). When and why defaults influence decisions: A meta-analysis of default effects. *Behavioural Public Policy*, *3*(2), 159-186. https://doi.org/10.1017/bpp.2018.43

Kahneman, D., & Tversky, A. (1982). The psychology of preferences. *Scientific American*, *246*(1), 160-173. https://doi.org/10.1038/scientificamerican0182-160

Knobe, J. (2003). Intentional Action and Side Effects in Ordinary Language. *Analysis*, *63*(3), 190-194. https://doi.org/10.1093/analys/63.3.190

Landman, J. (1987). Regret and elation following action and inaction: Affective responses to positive versus negative outcomes. *Personality and Social Psychology Bulletin, 13*(4), 524-536. https://doi.org/10.1177/0146167287134009

Lipsey, M. W. (2003). Those confounded moderators in meta-analysis: Good, bad, and ugly. *The Annals of the American Academy of Political and Social Science, 587*(1), 69-81. https://doi.org/10.1177/0002716202250791

Lüdecke, D. (2019). Package 'esc'. *R Package Version 0.5.1, 2019*

Page, M. J., Sterne, J. A., Higgins, J. P., & Egger, M. (2021). Investigating and dealing with publication bias and other reporting biases in meta-analyses of health research: A review. *Research Synthesis Methods, 12*(2), 248-259.

Poisot, T., Sachse, R., Ashander, J., & Galili, T. (2016). Package "digitize". *R Package Version 0.0.4, 2016*

Re, A. C. D., & Hoyt, W. T. (2014). Package "MAd": Meta-Analysis with Mean Differences. *R Package Version 0.8-2, 2014*

Tummers, B. (2006). DataThief III v.1.1. Available from http://www.datathief.org/

van Lissa, C. J. (2017). MetaForest: Exploring heterogeneity in meta-analysis using random forests. Retrieved from https://psyarxiv.com/myg6s/

Vembye, M. H., Pustejovsky, J. E., & Pigott, T. (2022). Power Approximations for Meta-Analysis of Dependent Effect Sizes. https://osf.io/6tp9y/download

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*, 1-48. https://doi.org/10.18637/jss.v036.i03

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A metaanalysis of outcome studies comparing bona fide psychotherapies: Empirically, "all must have prizes." *Psychological Bulletin, 122*(3), 203–215. https://doi.org/10.1037/0033-2909.122.3.203

Yeung, S. K., & Feldman, G. (2022). Revisiting the Temporal Pattern of Regret: Replication of Gilovich and Medvec (1994) with extensions examining responsibility. https://doi.org/10.13140/RG.2.2.11150.08000/1

Yeung, S. K., Yay, T., & Feldman, G. (2021). Action and inaction in moral judgments and decisions: Meta-analysis of Omission-Bias omission-commission asymmetries. http://dx.doi.org/10.13140/RG.2.2.28468.58249

Zhang, Y., Hedo, R., Rivera, A., Rull, R., Richardson, S., & Tu, X. M. (2019). Post hoc power analysis: is it an informative and meaningful analysis?. *General psychiatry*, *32*(4). https://dx.doi.org/10.1136%2Fgpsych-2019-100069