**Peer Review and Communication History**

**MS Title**: Alone with our Thoughts: Investigation of Autonomy Supportive Framing as a Driver of Enjoyment During Quiet Time in Solitude

**Author Names**: Thuy-vy Nguyen, Netta Weinstein, Edward Deci

**Submitted:** Jan 9, 2021

**Editor First Decision**: Revise & Resubmit

Jan 17, 2021

Dear Dr. Nguyen,

I have now reviewed your manuscript, as well as the previous peer reviews and decision letter and your response to them. My assessment is that your manuscript has important strengths and also that there are some issues that need to be addressed. I therefore encourage you to submit a revised version for further consideration at Collabra: Psychology.

I outline my major concerns below. In your resubmission, please include a document with a point-by-point response to both the points I list here, outlining each change made in your manuscript or providing a suitable rebuttal.

In addition, I am attaching a Word document with tracked changes and comments – please consider all of those edits and comments, and describe any that you are not accepting or addressing in your resubmission letter.

1. I found it quite difficult to compare the studies to the pre-registrations. First, the hypotheses presented at the end of the introduction do not map on clearly to the hypotheses in the pre-registrations for the two studies. In the document, I suggest some ways to address this. However, what is most important is that the results pertaining to the pre-registered hypotheses are all reported in the manuscript, that readers can clearly tell where to find the results pertaining to each pre-registered hypothesis, and that any results that do not directly map on to a pre-registered hypothesis are clearly marked as not pre-registered. But anything you can do above and beyond this to make it easier for readers to tell which results map on to which hypothesis in your pre-registrations and which hypothesis at the end of your introduction, the better. I want to emphasize that, while I raised some question in my tracked changes about where the tests were that correspond to the pre-registered hypotheses, these comments are not meant to be exhaustive and I would like you to carefully check that every verbal hypothesis in the pre-registration and in the introduction corresponds to a specific statistical test (or more than one) in the results, and that it is always clear to the reader whether the hypothesis and statistical test were pre-registered, and whether the results support the prediction. As another example not mentioned in my comments on the manuscript, you do not report the result for the end-of-day measures because the effects were not there – that is not a good reason not to include these results in the results section. Indeed, pre-registered hypotheses that did not pan out are just as important to report! I had to search pretty hard to find your explanation for why you didn’t report these results, and many readers will not check the pre-registration or spend as much time searching the manuscript as I did, so it is the author’s responsibility to make sure that all readers will get a clear and accurate picture of what hypotheses were pre-registered and what the results were for those pre-registered hypotheses. As currently written, I did not come out of my reading of your manuscript with a clear sense of that, and I think you will need to make major changes to the structure of the paper to accomplish this. I know it is a very hard task, and most papers reporting pre-registered studies that I read also fall short, but I would like to see this change if you resubmit to Collabra.

Besides this issue of matching the pre-registered hypotheses to the results and the hypotheses listed in the introduction, there were many other points where there seemed to be deviations from the pre-registration that were not emphasized. Please make sure all deviations are clearly marked. In addition to deviations, you should also make it clear when the pre-registration did not specify in enough detail what you would do, and so you had flexibility in how you approached the decision. These also count as non-pre-registered decisions even though they are not exactly deviations, and readers should know that they were not pre-registered. Finally, I would like you to mention the pre-registrations and link to them at the beginning of each study’s method section, and also to state that the pre-registrations were completed after data collection had begun, but that you had not seen any of the data when you submitted the pre-registration (assuming this is accurate).

1. I am concerned that the evidence is not as strong as or maybe even doesn’t support your claims. In the end, this is not such a big deal because your overall conclusion is that you have not found a way to change people’s affect or behavior, but even the perceived autonomy and enjoyment results may be weaker than you present them as being. I have several reasons for this skepticism. First, the results from Study 1 make me worry that the effects are driven by a negative effect of the controlling condition rather than a positive effect of the autonomy supportive condition – the pattern of means seems very consistent with that possibility. By removing the neutral condition in Study 2, we cannot judge this anymore, so I think it is still very much a live possibility, and one that should temper your conclusions. Second, for perceived autonomy, you did not pre-register a hypothesis or analysis about this, I don’t think (I may be misreading or have missed something in your pre-registration). This means that you should treat the result with a lot more uncertainty, and a one-tailed test is not appropriate, in my opinion (see also point #4 below about one-tailed tests more generally). Thus, the result now becomes a p of .046 without any pre-registration, which, to my mind, is a very equivocal result. Moreover, the effect size is quite small, and falls outside the minimum effect size that you powered your study to detect (d = .30). Combined with my first concern, I don’t think you have strong evidence that the autonomy supportive condition enhanced people’s motivation for autonomy. I feel quite strongly that a conclusion that you were able to increase perceived autonomy at all, and certainly not beyond a trivial amount. With regards to enjoyment, I have more mixed feelings about that. Again, the effect size in Study 2 was smaller than what you aimed to power your study to detect, and may be too small to be practically meaningful. I am also concerned about the potential for flexibility in data analysis to have boosted this effect, because the pre-registration is not clear about how enjoyment would be operationalized – it implies, to me, that you would examine the effect of the manipulation on three separate items. To address these concerns, I suggest you run robustness tests examining the effect in the individual items, as well as the aggregate, and that you present a visualization showing the individual data points so readers can evaluate how robust they think the effect is. I would also like you to adjust your conclusions if these further analyses present reasons for skepticism about the effect.
2. Given that many of your conclusions rest on null results, I think you should be more systematic about how you draw inferences from the null results. I realize that you only pre-registered null significance hypothesis tests, but as these don’t allow you to evaluate the amount of evidence for the absence of an effect, you need another approach to do that. You sometimes follow up with a discussion of effect sizes, but as I note in my comments on the document, this is not enough (I am not a fan of reporting point estimates without confidence intervals, and the point estimates alone aren’t enough to judge the evidence for absence of an effect). I would encourage you to explicitly use a technique such as equivalence testing or Bayesian statistics, and apply it consistently to null results throughout the manuscript. You would need to be explicit that these analyses were not pre-registered, and that any inference criteria or cutoffs you choose will be post-doc and could be influenced by your knowledge of the data/results, but that it still probably better than the haphazard treatment that the null results receive now. Given your sample size in Study 2, I think it is likely that you have relatively strong evidence of absence of meaningful effects for many of the outcomes variables, but it would be good to be more systematic in your assessment of this.
3. As you’ll see in my comment on page 14, I don’t think the use of a one-tailed test here is justified. To be transparent, I am not a fan of one-tailed tests in most of social and personality psychology, where we rarely have reason to be confident that there is no chance of a meaningful effect in the opposite direction to our prediction (though I can imagine some exceptions). More importantly, I think we as a field need to be very disciplined about the use of one-tailed tests, and they should only be used when pre-registered, to avoid fooling ourselves into using them when we need them to make the results looks better. Thus, I would very strongly prefer that you use two-tailed tests, which would also allow you to run the analyses you pre-registered. If you have reasons you think I have not considered for choosing a one-tailed test, you are welcome to lay out those reasons, but I’ll be honest that I am very reluctant to set a precedent that post-hoc one-tailed tests are acceptable outside of situations where an effect in the opposite direction would be almost completely absurd.
4. Can you provide any explanation for why the sample sizes across conditions in Study 1 are so unequal? I am especially worried (in both studies) about participant dropout – if participants dropped out after the instructions were given, this would threaten the random assignment and therefore causal inference. But even beyond that, it would be good to know what could have caused such unequal cell sizes.
5. I wonder whether the pencil sorting activity may not have been as unpleasant and banal as you describe, if participants thought they were helping the researchers by sorting pencils. Can you rule this out? If it could have been perceived as a pro-social activity, this changes the interpretation of this measure quite a bit.
6. In the discussion you report that the effects of the manipulation on (increasing) low arousal negative affect are weaker than on (increasing) low arousal positive affect, but this is a claim that could and should be tested statistically if you are going to discuss it. Please either remove it, or conduct the appropriate statistical test (clearly marked as not-pre-registered)
7. I also wondered whether the fact that the study was advertised as a “solitude” study could lead to selection bias, and thus affect the validity of inferences, including but not limited to causal inferences and generalizability.

In summary, I think this is a promising manuscript and, I hope you will revise it for further consideration at Collabra: Psychology. I look forward to receiving your revision.

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This is the last opportunity for major editing, therefore please fully check your file prior to re-submission.

If you have any questions or difficulties during this process, please contact the editorial office at editorialoffice@collabra.org.

We hope you can submit your revision within the next six weeks. If you cannot make this deadline, please let us know as early as possible.

Sincerely,

Simine Vazire
Editor in Chief
Collabra: Psychology

**Author Response**
May 11, 2021

Dear Dr. Vazire,

We are grateful to your extensive attention to our manuscript and constructive and detailed feedback. Your perspective has reinvigorated our passion for this paper and sense that it is an authentic expression our work, following a depleting previous peer-review process. We have italicized your comments and respond to each, in turn, below. You will see if we have attempted to make all requested changes, and describe our rationale in the one case when we do not do so.

***Editor’s comment 1:*** *I found it quite difficult to compare the studies to the pre-registrations. First, the hypotheses presented at the end of the introduction do not map on clearly to the hypotheses in the pre-registrations for the two studies. In the document, I suggest some ways to address this. However, what is most important is that the results pertaining to the pre-registered hypotheses are all reported in the manuscript, that readers can clearly tell where to find the results pertaining to each pre-registered hypothesis, and that any results that do not directly map on to a pre-registered hypothesis are clearly marked as not pre-registered. But anything you can do above and beyond this to make it easier for readers to tell which results map on to which hypothesis in your pre-registrations and which hypothesis at the end of your introduction, the better. I want to emphasize that, while I raised some question in my tracked changes about where the tests were that correspond to the pre-registered hypotheses, these comments are not meant to be exhaustive and I would like you to carefully check that every verbal hypothesis in the pre-registration and in the introduction corresponds to a specific statistical test (or more than one) in the results, and that it is always clear to the reader whether the hypothesis and statistical test were pre-registered, and whether the results support the prediction. As another example not mentioned in my comments on the manuscript, you do not report the result for the end-of-day measures because the effects were not there – that is not a good reason not to include these results in the results section. Indeed, pre-registered hypotheses that did not pan out are just as important to report! I had to search pretty hard to find your explanation for why you didn’t report these results, and many readers will not check the pre-registration or spend as much time searching the manuscript as I did, so it is the author’s responsibility to make sure that all readers will get a clear and accurate picture of what hypotheses were pre-registered and what the results were for those pre-registered hypotheses. As currently written, I did not come out of my reading of your manuscript with a clear sense of that, and I think you will need to make major changes to the structure of the paper to accomplish this. I know it is a very hard task, and most papers reporting pre-registered studies that I read also fall short, but I would like to see this change if you resubmit to Collabra.*

*Besides this issue of matching the pre-registered hypotheses to the results and the hypotheses listed in the introduction, there were many other points where there seemed to be deviations from the pre-registration that were not emphasized. Please make sure all deviations are clearly marked. In addition to deviations, you should also make it clear when the pre-registration did not specify in enough detail what you would do, and so you had flexibility in how you approached the decision. These also count as non-pre-registered decisions even though they are not exactly deviations, and readers should know that they were not pre-registered. Finally, I would like you to mention the pre-registrations and link to them at the beginning of each study’s method section, and also to state that the pre-registrations were completed after data collection had begun, but that you had not seen any of the data when you submitted the pre-registration (assuming this is accurate).*

**Comment 1 response:** We apologize for the disorganization and mismatch between our preregistration and the writeup of this research. We now have returned to our original preregistrations for both studies, and once again include the outcomes that we originally had in the manuscript, but which we had removed after reviewer requests to do so in previous rounds of peer review at a previous journal.

You will see that, now in Study 1, we looked at condition effect on: H1 – Autonomous Motivation for Solitude, H2 – Task Experience and Future Intention, H3 – Free-Choice Behavior, H4 – End-of-Day Well-being. We have used headings to clearly organise these hypotheses in ‘The Present Study section (in Introduction) and the Results section. We also separated between Preregistered Analyses and Exploratory Analyses (i.e., investigating Changes in Affect).

In Study 2, we looked at condition effect on: H1 – Changes in Affect, H2 – Free-Choice Behavior, H3 – Free-Choice Solitude Experience, H4 – End-of-Day Relatedness, H5 – Future Engagement with Solitude. We again used headings to clearly organise these hypotheses in the Results section of Study 2. These sections are now aligned with the pre-registered document and our initial goals for the project.

***Editor’s comment 2:*** *I am concerned that the evidence is not as strong as or maybe even doesn’t support your claims. In the end, this is not such a big deal because your overall conclusion is that you have not found a way to change people’s affect or behavior, but even the perceived autonomy and enjoyment results may be weaker than you present them as being. I have several reasons for this skepticism. First, the results from Study 1 make me worry that the effects are driven by a negative effect of the controlling condition rather than a positive effect of the autonomy supportive condition – the pattern of means seems very consistent with that possibility. By removing the neutral condition in Study 2, we cannot judge this anymore, so I think it is still very much a live possibility, and one that should temper your conclusions. Second, for perceived autonomy, you did not pre-register a hypothesis or analysis about this, I don’t think (I may be misreading or have missed something in your pre-registration). This means that you should treat the result with a lot more uncertainty, and a one-tailed test is not appropriate, in my opinion (see also point #4 below about one-tailed tests more generally). Thus, the result now becomes a p of .046 without any pre-registration, which, to my mind, is a very equivocal result. Moreover, the effect size is quite small, and falls outside the minimum effect size that you powered your study to detect (d = .30). Combined with my first concern, I don’t think you have strong evidence that the autonomy supportive condition enhanced people’s motivation for autonomy. I feel quite strongly that a conclusion that you were able to increase perceived autonomy at all, and certainly not beyond a trivial amount. With regards to enjoyment, I have more mixed feelings about that. Again, the effect size in Study 2 was smaller than what you aimed to power your study to detect, and may be too small to be practically meaningful. I am also concerned about the potential for flexibility in data analysis to have boosted this effect, because the pre-registration is not clear about how enjoyment would be operationalized – it implies, to me, that you would examine the effect of the manipulation on three separate items. To address these concerns, I suggest you run robustness tests examining the effect in the individual items, as well as the aggregate, and that you present a visualization showing the individual data points so readers can evaluate how robust they think the effect is. I would also like you to adjust your conclusions if these further analyses present reasons for skepticism about the effect.*

**Comment 2 response:** We have now added the robust tests to the manuscript, and included the test statistics for individual items of enjoyment composite to *Table 2,* and *Table 6,* following the plans we had preregistered. We presented Cohen’s *d* effect size and 95% confidence intervals for the composite and the individual items in both tables. In Study 2, similar measures from Study 1 were not preregistered so we included the test statistics for those variables in the Supplementary Materials so the reader can see which effects were or were not replicated. Overall, as you have suggested, there is not strong evidence for differences between conditions as the effect sizes we obtained were not bigger than the smallest effect size that our samples had adequate power to detect. We stated this throughout the manuscript for each outcome variable.

***Editor’s comment 3:*** *Given that many of your conclusions rest on null results, I think you should be more systematic about how you draw inferences from the null results. I realize that you only pre-registered null significance hypothesis tests, but as these don’t allow you to evaluate the amount of evidence for the absence of an effect, you need another approach to do that. You sometimes follow up with a discussion of effect sizes, but as I note in my comments on the document, this is not enough (I am not a fan of reporting point estimates without confidence intervals, and the point estimates alone aren’t enough to judge the evidence for absence of an effect). I would encourage you to explicitly use a technique such as equivalence testing or Bayesian statistics, and apply it consistently to null results throughout the manuscript. You would need to be explicit that these analyses were not pre-registered, and that any inference criteria or cutoffs you choose will be post-doc and could be influenced by your knowledge of the data/results, but that it still probably better than the haphazard treatment that the null results receive now. Given your sample size in Study 2, I think it is likely that you have relatively strong evidence of absence of meaningful effects for many of the outcomes variables, but it would be good to be more systematic in your assessment of this.*

**Comment 3 response:** Following careful reflection, we have decided not to pursue equivalence testing or Bayesian statistics because, as you have said, any equivalence null or cutoffs now would be post-hoc. We selected instead to report the statistics as compared to the smallest effect size our sample has adequate power to detect. As such, for any effect sizes that were smaller than that number, we concluded that we did not have evidence to support our hypotheses. Overall, our research did not find evidence suggesting that supporting autonomy or controlling instructions administered in laboratory settings creates substantial effects on participants’ experiences when sitting alone with thought, their intention to do it again, the well-being benefits they gained from it, nor subsequent engagement with this experience. We explore why that may be in the Discussion section.

***Editor’s comment 4:*** *As you’ll see in my comment on page 14, I don’t think the use of a one-tailed test here is justified. To be transparent, I am not a fan of one-tailed tests in most of social and personality psychology, where we rarely have reason to be confident that there is no chance of a meaningful effect in the opposite direction to our prediction (though I can imagine some exceptions). More importantly, I think we as a field need to be very disciplined about the use of one-tailed tests, and they should only be used when pre-registered, to avoid fooling ourselves into using them when we need them to make the results looks better. Thus, I would very strongly prefer that you use two-tailed tests, which would also allow you to run the analyses you pre-registered. If you have reasons you think I have not considered for choosing a one-tailed test, you are welcome to lay out those reasons, but I’ll be honest that I am very reluctant to set a precedent that post-hoc one-tailed tests are acceptable outside of situations where an effect in the opposite direction would be almost completely absurd.*

**Comment 4 response:** Thank you for your honest comment. We have now removed the one-tailed t-test and went back to follow our original preregistration. Originally, we believed the one-tailed t-test is justified because our hypotheses were solely based on self-determination theory, which had clear prediction of how autonomy support and control would influence the outcomes. However, given that our one-tailed t-tests were post-hoc and that two-tailed t-tests would provide true tests for theory testing; if we had found the effect in the opposite direction it would directly contradict with self-determination theory. Switching back to two-tailed t-tests did not change the results, but we agreed with your argument.

***Editor’s comment 5:*** *Can you provide any explanation for why the sample sizes across conditions in Study 1 are so unequal? I am especially worried (in both studies) about participant dropout – if participants dropped out after the instructions were given, this would threaten the random assignment and therefore causal inference. But even beyond that, it would be good to know what could have caused such unequal cell sizes.*

**Comment 5 response:** We have now added more details on how we recruited participants. The original write-up had little detail that described this process, and we apologize for that. Recruitment decisions were made based on practical considerations; to make sure we have the sample size we aim for before the window for SONA recruitment closed (at the institution where this work was conducted, SONA recruitment was closed in the summer). The unequal sample sizes between conditions occurred because we set Qualtrics algorithm to randomly assign participants to conditions, and allowed full randomization. We now state this in the Recruitment and Participants sections in both studies; thank you, we agree this is important detail to convey to readers.

***Editor’s comment 6:*** *I wonder whether the pencil sorting activity may not have been as unpleasant and banal as you describe, if participants thought they were helping the researchers by sorting pencils. Can you rule this out? If it could have been perceived as a pro-social activity, this changes the interpretation of this measure quite a bit.*

**Comment 6 response:** We could not test whether participants engaged in the pencil-sorting task for prosocial purposes, but we conducted mixed ANOVAs on both Study 1 and Study 2 data to investigate whether participants perceived the pencil-sorting any more or less enjoyable than the alone-with-thought experience. We reported the result of these exploratory analyses in the Supplementary Materials, and we not mention this issue in the Discussion section of the main manuscript on page 27. Overall, there was not significant difference between ratings of enjoyment for pencil-sorting task compared to sitting alone with thoughts.

***Editor’s comment 7:*** *In the discussion you report that the effects of the manipulation on (increasing) low arousal negative affect are weaker than on (increasing) low arousal positive affect, but this is a claim that could and should be tested statistically if you are going to discuss it. Please either remove it, or conduct the appropriate statistical test (clearly marked as not-pre-registered).*

**Comment 7 response:** In response to this recommendation, we have now removed the claim that the increase of low-arousal negative affect is smaller than the increase of low-arousal positive affect.

***Editor’s comment 8:*** *I also wondered whether the fact that the study was advertised as a “solitude” study could lead to selection bias, and thus affect the validity of inferences, including but not limited to causal inferences and generalizability.*

**Comment 8 response:** We did not have any data to speak to the possibility that the effect of advertising the study as one focused on “solitude” could have caused selection bias, but now highlight this issue in Footnote 1 on page 8. In addition, undergraduate samples such as ones we recruited here definitely limit generalizability. We have added to the Discussion section (page 28) that future research should extend the paradigm outside of laboratory setting. We argue that doing so would make possible the recruitment of more diverse samples other than undergraduate students, particularly those from different age groups and living arrangements.

***Editor’s comment 9:*** *In summary, I think this is a promising manuscript and, I hope you will revise it for further consideration at Collabra: Psychology. I look forward to receiving your revision.*

*Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This is the last opportunity for major editing, therefore please fully check your file prior to re-submission.*

**Comment 9 response:** We greatly appreciate your (the Editor’s) comments, both in the manuscript and in general. These comments have allowed to re-evaluate our work more critically and gained ownership over the conclusions we can make from the evidence we have obtained.

Sincerely,

Thuy-vy Nguyen

Netta Weinstein

Edward Deci

**Editor First Decision**: Revise & Resubmit

Jul 14, 2021

Dear Dr. Nguyen,

Thank you for submitting your revised manuscript, “Alone with our Thoughts: Supporting Autonomy as a Driver of Enjoyment during Quiet Time in Solitude” along with the detailed cover letter. I appreciate your responsiveness to the points I raised. I have read your revision carefully, and it is close, but there are still some minor changes and one major change that I would like to see before accepting it for publication. I therefore encourage you to submit a revised version for further consideration at Collabra: Psychology.

I want to begin by apologizing for taking a while to send you this decision. However, I also have to confess that I feel that I have put an unusual amount of time into each version of this manuscript, and I cannot promise that I will be able to do the same for the next version (or if I need to, it will once again take several months before I can set aside the large swath of time necessary to go through the manuscript so closely). I know it may seem as though I am being overly nit-picky - I can assure you that I typically don’t edit manuscripts by tracking changes in Microsoft Word (I typically only decribe the changes in narrative format in my decision letters). In the case of the original submission, I felt that there was a strong set of studies wrapped in a manuscript that did not show off their strength and did not always represent what was done in a clear and transparent manner, so I wanted to spend the time to make it clear how the studies would need to be presented in order to meet the criteria for publication in Collabra: Psychology. In the case of the revision, the structure of the manuscript is there, and the reporting is accurate, but there were still a lot of inconsistencies, typos, and other errors that made it difficult to get through the manuscript. I continue to think this is a very strong set of studies testing an important question, but that more care needs to be put into the writing. I am concerned that I am finding rather obvious errors (e.g., questions ending in periods rather than question marks, sentences with extra words or missing important words, etc.). I realize a few errors like this is normal, but I am seeing a lot even in the revision. This makes me nervous about whether the authors have proofread the mansucript, and whether there are errors I’m not catching (e.g., I think I found at least one error in a statistic reported (a positive number on page 18 that I think should be negative, though I’m not sure), but I am not in a position to catch most statistical errors, as I did not examine the data and code). For the next revision, please ensure that several people have carefully checked the manuscript for errors, including computational errors and errors in transcribing results. Please include a thorough check of the tables by at least two people. (Note that I have not looked at the tables - I ran out of time and do not want to delay this decision any longer, so I will plan to look at the tables more closely next time. Please make sure they are clear and free of errors. The best way to do this is by having a few people who aren’t familiar with the paper try to make sense of the tables and give you feedback about what is unclear.)

In addition to the smaller issues raised in my comments in the document attached, the remaining large issue is how to interpret null results. I believe that you attempted to address my concern by adding more discussion of whether the effects you observed were at least as large as the smallest effect that you had 80% power to detect. Unfortunately, that is not a great way to address my concern. First, it confuses two things: the effect size from your sensitivity power analysis and the inference criterion you pre-registered (alpha = .05, two-tailed). These do not match up, because the smallest effect necessary to obtain p < .05 is smaller than the smallest effect you have 80% power to detect (by definition, you have 50% power to detect a just-significant effect), so you should not use the effect size from you sensitivity power analysis to decide what threshold to use for a “meaningful” effect, because you already commited to interpreting any effect with a p-value < .05 as meaningful (or at least “real”). Moreover, comparing non-significant effects to the critical effect size from your sensivity power analysis does not help us to interpret them (in part because we alreay know they must be below this threshold since they were not significant). The key question I would like you to consider in interpreting your null effects is: do you think this results is evidence that there is no effect (or, that any effect that may exist is so close to zero as to be trivial)? Or do you think that this null result is inconclusive (it is consistent with both the possibility of a no/trivial effect and also with the possibility of a non-trivial effect). There isn’t one right way to do this, but it requires grappling with the boundary for how small you think an effect could be and still be meaningful. That isn’t the same as your sensitivity power analysis (indeed, in Study 1, you had 80% power only to detect relatively large effect sizes - I assume you might not be prepared to say that any effect smaller than .41 is trivial). It’s a conceptual question, not a mathemetical one (though once you pick your answer to the conceptual question, it has implications for the statistics). Once you have an idea of the boundary, or range, of how small an effect has to be before it’s closer enough to zero to be trivial, then you have a basis for discussing and interpreting your null effects, by comparing that boundary to the ends of the 95% confidence intervals for your non-significant effects. For example, if I am studying phenomenon X and I decide that an effect smaller than d = .20 is probably not practically meaningful, then I can examine my non-signifcant results and see if the 95% CIs only contain values smaller than .20. If so, then I would argue that the results actually provide evidence for an absence of an effect - I can confidently say that I can rule out any meaningfully-sized effect (to oversimplify a bit). However, if the 95% CIs on my non-significent effects include effect sizes larger than .20, then I would say that the result is inconclusive - maybe there’s no effect, or maybe there is a meaninful effect and I just failed to detect it.

In this way of thinking, there is often a gap between your sensitivity power analysis result (e.g., in your Study 1, d = .41) and the effect size you decide is the boundary for trivial vs. meaningful effects. For example, if you decide anything above d = .30 could be meaningful, then you have a zone (between .30 and .41) where you don’t quite have adequate power, but you still think a true effect in that range would be meaningful. That’s why null results can be inconclusive - we don’t always have as much power as we would need to always have high power to detect an effect size just large enough to be meaningful. This is just life - we don’t always have as much power/precision as we would like. It’s ok, but it’s important to include a discussion, for null results, of whether we think they fall in the “evidence of absence (i.e., 95% CI includes only trivially small effects)” or “absence of evidence (i.e., not significant, but 95% CI still includes effect sizes that are big enough to be meaningful)” category, and explain why. I’m oversimplifying quite a bit here (and that’s why equivalence tests or Bayesian analyses are better), but I think this logic is enough to get us where we need to be for your manuscript.

In summary, I think your manuscript is close to publishable and I hope you will revise it for further consideration at Collabra: Psychology. I look forward to receiving your revision.

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This is the last opportunity for major editing, therefore please fully check your file prior to re-submission.

If you have any questions or difficulties during this process, please contact the editorial office at editorialoffice@collabra.org.

We hope you can submit your revision within the next six weeks. If you cannot make this deadline, please let us know as early as possible.

Sincerely,

Simine Vazire
Editor in Chief
Collabra: Psychology

**Author Response**
Aug 18, 2021

*Hi Dr. Vazire, I just want to first say that I really appreciate the work and efforts you have put into reviewing this manuscript. It was my dissertation work in 2018, and revising the manuscript to meet the criteria for publication in Collabra has allowed me to think more reflectively about my findings. This process also makes me prouder of my works and I am committed to improving its transparency and clarity further if necessary. Both the second author and I feel that we have learned a lot from this process, and we would like to take this opportunity to thank you for that.*

Dear Dr. Nguyen,

Thank you for submitting your revised manuscript, "Alone with our Thoughts: Supporting Autonomy as a Driver of Enjoyment during Quiet Time in Solitude" along with the detailed cover letter. I appreciate your responsiveness to the points I raised. I have read your revision carefully, and it is close, but there are still some minor changes and one major change that I would like to see before accepting it for publication. I therefore encourage you to submit a revised version for further consideration at Collabra: Psychology.

I want to begin by apologizing for taking a while to send you this decision. However, I also have to confess that I feel that I have put an unusual amount of time into each version of this manuscript, and I cannot promise that I will be able to do the same for the next version (or if I need to, it will once again take several months before I can set aside the large swath of time necessary to go through the manuscript so closely). I know it may seem as though I am being overly nit-picky - I can assure you that I typically don't edit manuscripts by tracking changes in Microsoft Word (I typically only decribe the changes in narrative format in my decision letters). In the case of the original submission, I felt that there was a strong set of studies wrapped in a manuscript that did not show off their strength and did not always represent what was done in a clear and transparent manner, so I wanted to spend the time to make it clear how the studies would need to be presented in order to meet the criteria for publication in Collabra: Psychology.

***Response:*** *For this revision, I submitted both the tracked change version which has my co-author’s edits (to check for typos and sentence structure) as well as parts where I highlighted to address your comments in the manuscript. Below, I will summarize those revisions and explain how we have responded to your feedback.*

In the case of the revision, the structure of the manuscript is there, and the reporting is accurate, but there were still a lot of inconsistencies, typos, and other errors that made it difficult to get through the manuscript. I continue to think this is a very strong set of studies testing an important question, but that more care needs to be put into the writing. I am concerned that I am finding rather obvious errors (e.g., questions ending in periods rather than question marks, sentences with extra words or missing important words, etc.). I realize a few errors like this is normal, but I am seeing a lot even in the revision. This makes me nervous about whether the authors have proofread the manuscript, and whether there are errors I'm not catching (e.g., I think I found at least one error in a statistic reported (a positive number on page 18 that I think should be negative, though I'm not sure), but I am not in a position to catch most statistical errors, as I did not examine the data and code). For the next revision, please ensure that several people have carefully checked the manuscript for errors, including computational errors and errors in transcribing results. Please include a thorough check of the tables by at least two people. (Note that I have not looked at the tables - I ran out of time and do not want to delay this decision any longer, so I will plan to look at the tables more closely next time. Please make sure they are clear and free of errors. The best way to do this is by having a few people who aren't familiar with the paper try to make sense of the tables and give you feedback about what is unclear.)

***Response:*** *My second author and I have reviewed the manuscript multiple times. First, I reviewed the whole manuscript, corrected any typos or unclear sentences, and addressed your comments. Then, Dr. Netta Weinstein reviewed the manuscript the second time to check for typos, made sure the manuscripts and tables follow APA format. Finally, I reviewed the manuscript again. I hope now all the typos have been corrected. Further, we also made sure to consistently add hyphens to “autonomy-supportive” and “controlling-instruction” when addressing the conditions.*

In addition to the smaller issues raised in my comments in the document attached, the remaining large issue is how to interpret null results. I believe that you attempted to address my concern by adding more discussion of whether the effects you observed were at least as large as the smallest effect that you had 80% power to detect. Unfortunately, that is not a great way to address my concern. First, it confuses two things: the effect size from your sensitivity power analysis and the inference criterion you pre-registered (alpha = .05, two-tailed). These do not match up, because the smallest effect necessary to obtain p < .05 is smaller than the smallest effect you have 80% power to detect (by definition, you have 50% power to detect a just-significant effect), so you should not use the effect size from you sensitivity power analysis to decide what threshold to use for a "meaningful" effect, because you already committed to interpreting any effect with a p-value < .05 as meaningful (or at least "real"). Moreover, comparing non-significant effects to the critical effect size from your sensitivity power analysis does not help us to interpret them (in part because we already know they must be below this threshold since they were not significant). The key question I would like you to consider in interpreting your null effects is: do you think this results is evidence that there is no effect (or, that any effect that may exist is so close to zero as to be trivial)? Or do you think that this null result is inconclusive (it is consistent with both the possibility of a no/trivial effect and also with the possibility of a non-trivial effect). There isn't one right way to do this, but it requires grappling with the boundary for how small you think an effect could be and still be meaningful. That isn't the same as your sensitivity power analysis (indeed, in Study 1, you had 80% power only to detect relatively large effect sizes - I assume you might not be prepared to say that any effect smaller than .41 is trivial). It's a conceptual question, not a mathematical one (though once you pick your answer to the conceptual question, it has implications for the statistics). Once you have an idea of the boundary, or range, of how small an effect has to be before it's closer enough to zero to be trivial, then you have a basis for discussing and interpreting your null effects, by comparing that boundary to the ends of the 95% confidence intervals for your non-significant effects. For example, if I am studying phenomenon X and I decide that an effect smaller than d = .20 is probably not practically meaningful, then I can examine my non-significant results and see if the 95% CIs only contain values smaller than .20. If so, then I would argue that the results actually provide evidence for an absence of an effect - I can confidently say that I can rule out any meaningfully-sized effect (to oversimplify a bit). However, if the 95% CIs on my non-significant effects include effect sizes larger than .20, then I would say that the result is inconclusive - maybe there's no effect, or maybe there is a meaningful effect and I just failed to detect it. In this way of thinking, there is often a gap between your sensitivity power analysis result (e.g., in your Study 1, d = .41) and the effect size you decide is the boundary for trivial vs. meaningful effects. For example, if you decide anything above d = .30 could be meaningful, then you have a zone (between .30 and .41) where you don't quite have adequate power, but you still think a true effect in that range would be meaningful. That's why null results can be inconclusive - we don't always have as much power as we would need to always have high power to detect an effect size just large enough to be meaningful. This is just life - we don't always have as much power/precision as we would like. It's ok, but it's important to include a discussion, for null results, of whether we think they fall in the "evidence of absence (i.e., 95% CI includes only trivially small effects)" or "absence of evidence (i.e., not significant, but 95% CI still includes effect sizes that are big enough to be meaningful)" category, and explain why. I'm oversimplifying quite a bit here (and that's why equivalence tests or Bayesian analyses are better), but I think this logic is enough to get us where we need to be for your manuscript.

***Response:*** *I thank you very much for this thoughtful feedback. I have now included a section “Evaluation of Effects on In-Lab Measures Across Studies” using an interval null between -0.30 and 0.30. I compared the effect sizes of the differences between autonomy-supportive and controlling-instruction conditions against this interval. I determined that an effect size smaller than 0.30 either in favor of the autonomy-supportive or controlling-instruction condition would not be considered practically meaningful. I presented our observed effects and their confidence intervals in Figure 1. I believe this now has made the manuscript much clearer to inform future research about which effects were equivocal and would need larger sample sizes to determine presence or absence of an effect.*

*In addition to those revisions described above, I also addressed the comments that you provided in the manuscript with tracked changes, submitted along with a clean version with no tracked changes. The manuscript now has been much improved and I hope it meets the standards for publication in Collabra: Psychology.*

*Best wishes,*

*Thuy-vy Nguyen*

**Editor First Decision**: Revise & Resubmit

Sep 5, 2021

Dear Thuy-vy ,

Thank you for submitting your revised manuscript, “Alone with our Thoughts: Investigation of Autonomy Supportive Framing as a Driver of Enjoyment During Quiet Time in Solitude” along with the response letter. I appreciate the detailed attention that you and your co-author gave to the issues I raised, and your very gracious response. I have a few more remaining concerns and suggestions, but I hope this will be the last round of revisions. I therefore encourage you to submit a revised version for further consideration at Collabra: Psychology.

I should preface this by saying that I did not read the entire manuscript carefully - I focused on the points that have changed since the last version. I appreciate that the writing is clearer, the new figure is great, and I think the new section on interpreting null effects is also very helfpul. I made some comments (and a few small edits) using tracked changes in the attached document. Below I list a few more issues (some overlapping with my comments in the manuscript) that I would like you to address as well:

1. I don’t think the rationale for your smalles effect size of interest is very strong. You say that you chose d = .30 because that is the average effect size in the literature according to a meta-analysis, but you do not explain the reasoning behind why that makes it an appropriate cutoff for practically meaningful effects. Specifically, the average effect size from a meta-analysis could, in principle, be very large, in which case there is no reason to assume that any effect smaller than the meta-analytic average is not practically meaningful, or very small, in which case there is no reason to assume that any effect at least as big as the meta-analytic average is necessarily practically meaningful. That is, to my mind, the smallest effect size of interest and the average effect according to a meta-analysis can be completely independent - I would want more explanation for why you think the meta-analysis result can be used as guidance about the smallest practically meaningful effect. I provide a few options in my comment in the manuscripts - I don’t think there is a right answer about what the smallest effect size of interest should be, and I think d = .30 is reasonable, but I think you should either explain the rationale in more detail (or give a different rationale), or else state that you didn’t have a good rationale, but give some indication for why you think effects bigger than this matter, and effects smaller than this may not matter.
2. Please tell readers that null results will be interpreted in a later section of the manuscript, at the moment when you report these null results in the results sections of each study.
3. It was not clear to me why you excluded the outside-the-lab measures from the section interpreting null effects. I think you should probably include them, but if you have a good reason for excluding them I’d be open to it, but it would need to be explained better than your current explanation.
4. The figure is very helpful, but it needs a caption. Please explain all components of the figure that need explaining (one obvious thing that needs explaining is the dotted lines at d = -.30 and d = .30, but I suspect there are other things as well). The figure should be able to stand alone (with the caption), even if the reader does not read the main text.

In summary, I hope you will revise your manuscript for further consideration at Collabra: Psychology, and I hope this will be the last revision. I look forward to receiving it.

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This may be the last opportunity for major editing, therefore please fully check your file prior to re-submission.

If you have any questions or difficulties during this process, please contact the editorial office at editorialoffice@collabra.org.

We hope you can submit your revision within the next six weeks. If you cannot make this deadline, please let us know as early as possible.

Sincerely,

Simine Vazire
Editor in Chief
Collabra: Psychology

**Author Response**
Dec 17, 2021

Dear Editor,

I would like to thank again for another opportunity to strengthen this manuscript; I hope you will still consider this manuscript for publication at Collabra. I would also like to apologize for the time it has taken me to return this revision to you. I misread your letter and thought the deadline for submission was within the next 6 months, and realized after I have missed the deadline that it was indeed 6 weeks. I have sent an email on November 23rd to <notifications@email.scholasticahq.com> to let the journal know that I would try my best to submit it as soon as I can, but I am not sure the email was received.

I have now addressed all the comments you raised in the most recent review. Specifically, I have made changes accordingly to your suggestions as follow:

1. I don't think the rationale for your smallest effect size of interest is very strong. You say that you chose d = .30 because that is the average effect size in the literature according to a meta-analysis, but you do not explain the reasoning behind why that makes it an appropriate cutoff for practically meaningful effects. Specifically, the average effect size from a meta-analysis could, in principle, be very large, in which case there is no reason to assume that any effect smaller than the meta-analytic average is not practically meaningful, or very small, in which case there is no reason to assume that any effect at least as big as the meta-analytic average is necessarily practically meaningful. That is, to my mind, the smallest effect size of interest and the average effect according to a meta-analysis can be completely independent - I would want more explanation for why you think the meta-analysis result can be used as guidance about the smallest practically meaningful effect. I provide a few options in my comment in the manuscripts - I don't think there is a right answer about what the smallest effect size of interest should be, and I think d = .30 is reasonable, but I think you should either explain the rationale in more detail (or give a different rationale), or else state that you didn't have a good rationale, but give some indication for why you think effects bigger than this matter, and effects smaller than this may not matter.

**Response:** I have now included further rationales for the null interval on page 25. Specifically, I was honest that the choice was ambiguous and not a mathematical one. This is what I put:

“We compared effect sizes observed to an interval null between -0.30 and 0.30; that is, we determined that an effect size smaller than 0.30 either in favor of the autonomy-supportive or controlling-instruction condition would not be practically meaningful. The interval null between -0.30 and 0.30 was an ambiguous choice, as there was no clear reference for which effect sizes to expect from the previous literature for our manipulation. This interval null was also not a mathematical choice, as we entirely based our decision of the interval null on what the first author subjectively considered to be practically meaningful. Given that our studies used a rather subtle manipulation, we did not expect to get an effect size as large as .40 or .50, but we considered .20 too small to be practically meaningful. We settled for d = .30 as the large-enough effect size to aim for for this type of manipulation. Based on a meta-analysis by Ntoumanis et al. (2021), the smallest effect size of SDT-based interventions was also estimated to be around .30.”

1. Please tell readers that null results will be interpreted in a later section of the manuscript, at the moment when you report these null results in the results sections of each study.

**Response:** This has now been added to the manuscript on page 22 at the end of the Statistical Analyses section of Study 2: “In a later section at the end of the Results section of Study 2, we will use the ‘forestplot’ package (version 2.0) to compare all effect sizes of non-significant results in Study 1 and Study 2 to evaluate consistencies across two studies.”

1. It was not clear to me why you excluded the outside-the-lab measures from the section interpreting null effects. I think you should probably include them, but if you have a good reason for excluding them I'd be open to it, but it would need to be explained better than your current explanation.

**Response:** I have now added the outside-the-lab measures to the section interpreting null effects at the end of page 26. I used the same interval null for *ß* coefficients and produced a similar forest plot for those effects in Figure 2.

1. The figure is very helpful, but it needs a caption. Please explain all components of the figure that need explaining (one obvious thing that needs explaining is the dotted lines at d = -.30 and d = .30, but I suspect there are other things as well). The figure should be able to stand alone (with the caption), even if the reader does not read the main text.

**Response:** I have now added the caption to both Figure 1 and Figure 2: “These forest plots compare the differences between pairs of conditions on end-of-day measures against the interval null of -.30 and .30. In these plots, the blue squares represent the standardized coefficients (ß) observed in Study 1 and Study 2, with 95% confidence intervals (CI) around those coefficients. The dotted vertical lines represent the upper and lower bounds of the interval null. The effects that have 95% CI falling within the interval null were determined too small to be practically meaningful.”

Additionally, I have also read through the whole manuscript again to check for errors and typos before resubmitting it. I look forward to hearing back from you.

Best wishes,

Thuy-vy Nguyen

**Editor Final Decision: Accept**

Dec 21, 2021

Dear Thuy-vy ,

I have now had a chance to read over your revised manuscript “Alone with our Thoughts: Investigation of Autonomy Supportive Framing as a Driver of Enjoyment During Quiet Time in Solitude”, along with the letter describing the changes you made. Thank you very much for your responsiveness to the concerns I raised and for the extra changes you and your co-authors made - the manuscript is even clearer now. I am happy to say that your paper is now officially accepted for publication in Collabra: Psychology. Congratulations on this excellent work, I think it will make an important contribution to the literature and I look forward to seeing it published! I hope your experiences with Collabra: Psychology have been positive and that you will continue to consider it as an outlet for your work.

As there are no further reviewer revisions to make, you do not have to complete any tasks at this point. Our managing editor will contact you in case there are any pre-prodution file related questions. You will have an opportunity to check the page proofs before we publish your article. Thank you again for publishing in Collabra: Psychology.

Sincerely,
Simine Vazire
Editor in Chief
Collabra: Psychology

Note:

Jan 25, 2022

The handling editor for this manuscript, Simine Vazire, realized, after the peer review process had completed, that she has recently published with one of the authors (IJzerman, Lewis, Przybylski, Weinstein, DeBruin, Ritchie, Vazire, Forscher, Morey, Ivory, & Anvari, 2020). She regrets not noticing this earlier and not recusing herself from evaluating this manuscript, as this would be considered a potential Conflict of Interest. We believe the best course of action, so as not to penalize the authors unnecessarily, is to disclose Vazire's error here. Readers can view the peer review history for this manuscript to judge whether editor Vazire did a thorough job.