nature portfolio

Peer Review File

Social Technology Use and Life Satisfaction in a Five-Wave Panel Study of U.S. Adults

Corresponding Author: Professor Kostadin Kushlev

Version 0:

Decision Letter:

2nd February 2026

Dear Professor Kushlev,

Thank you for submitting your revised manuscript, "Social Technology Use Is Not Associated with Greater Subsequent Life Satisfaction: A Five-Wave Panel Study of U.S. Adults". After careful consideration and discussion with my colleagues, I am sorry to have to tell you that we do not feel that the referees' comments have been sufficiently addressed to justify sending this revision back for peer review.

This unusual course of action is taken occasionally to avoid unproductive rounds of review that result in reviewer fatigue and damage the chances of the manuscript obtaining a fair and objective evaluation. Such situations are not in an author's best interest so we try to avoid them when it seems prudent to do so.

We appreciate how the revisions have addressed the analytical/reporting issues the reviewers highlighted. However, we are not persuaded that the issues the reviewers raised regarding appropriate interpretations of the findings have been fully addressed. To prevent a negative outcome due to this issue, we ask that you further revise your manuscript and provide a revised point-by-point response.

Should you be able to adequately respond to these and the reviewers' other concerns, we would be happy to look at a revised manuscript.

We shall hope to receive your revised version as soon as possible. If you anticipate a delay of more than four weeks, however, please let us know. We will be happy to consider your revision so long as nothing similar has been accepted for publication at Nature Human Behaviour or published elsewhere. Should your manuscript be substantially delayed without notifying us in advance and your article is eventually published, the received date may be that of the revised, not the original, version.

If you are not interested in submitting a suitably revised manuscript in the future please let me know immediately so we can close your file. If you have any questions, please contact me.

Please use the link below to submit a suitably revised manuscript and updated response to referees when they are ready.

Link Redacted

Sincerely,

[redacted]

[redacted]
Nature Human Behaviour

Version 1:

Decision Letter:

24th March 2026

Dear Professor Kushlev,

Thank you once again for your revised manuscript, entitled "Social Technology Use Is Not Associated with Greater Subsequent Life Satisfaction: A Five-Wave Panel Study of U.S. Adults," and for your patience during the peer review process.

The manuscript has now been evaluated by the same reviewers as in the previous round. As you will see, the reviewers continue to raise several concerns that were also highlighted during the earlier review at Nature Communications and in our request for pre-review revisions. We therefore ask that you address these points comprehensively, following the reviewers' guidance as well as the editorial instructions outlined below.

1) Please revise language throughout the manuscript that implies a causal question or interpretation. For example, the Abstract currently frames the research question as "is it increasing our life satisfaction?", which suggests causality. This should be rephrased to reflect the correlational nature of the study.

2) Please ensure that all conclusions and interpretive statements rely solely on results that have been corrected for multiple comparisons. In instances where corrected and uncorrected results diverge, the corrected results must form the basis of the interpretation.

3) Given that the study yields almost only null results and that concerns remain regarding the conclusiveness of these findings, we ask that you report Bayes Factors and credibility intervals to strengthen the interpretations.

Finally, your revised manuscript must comply fully with our editorial policies and formatting requirements. Failure to do so will result in your manuscript being returned to you, which will delay its consideration. To assist you in this process, I have attached a checklist that lists all of our requirements. If you have any questions about any of our policies or formatting, please don't hesitate to contact me.

In sum, we invite you to revise your manuscript taking into account all reviewer and editor comments. We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

We hope to receive your revised manuscript within two months. I would be grateful if you could contact us as soon as possible if you foresee difficulties with meeting this target resubmission date.

With your revision, please:

• Include a "Response to the editors and reviewers" document detailing, point-by-point, how you addressed each editor and referee comment. If no action was taken to address a point, you must provide a compelling argument. When formatting this document, please respond to each reviewer comment individually, including the full text of the reviewer comment verbatim followed by your response to the individual point. This response will be used by the editors to evaluate your revision and sent back to the reviewers along with the revised manuscript.

• Highlight all changes made to your manuscript or provide us with a version that tracks changes.

• EXTENDED DATA FIGURES
When re-submitting your manuscript, please ensure that any supplementary figures and tables that are crucial to the manuscript's conclusions are converted into Extended Data figures and tables to increase visibility of these data. Extended Data figures and tables are online-only (present in the online PDF and full-text HTML versions of the paper), peer-reviewed display items that provide essential background to the article but are not included in the main article due to space constraints. A maximum of ten Extended Data display items (figures and tables) is permitted.

Please use the link below to submit your revised manuscript and related files:

Link Redacted

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your work. Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further.

Sincerely,



[redacted]
Nature Human Behaviour

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author):

I appreciate the authors' revisions here. Many of them are very positive, and I do believe the authors have been serious in taking the reviewer recommendation. However, I am still concerned by the overinterpretation of some near zero (indeed inconclusive) effect sizes and the authors got seduced by some bad arguments (i.e., Orth,. Indeed, it looks like the authors basically find null effects in both directions, and they should simply acknowledge that.

The title should be changed to reflect the bidirectional null results, "Social Technology Use Is Not Associated with Greater OR REDUCED Subsequent Life Satisfaction." (caps just highlight the needed addition, not yelling, for the record).

The sentence in the abstract that begin: "Negative associations emerged only for specific…" is going to need to be edited or removed to, once again, reflect that these effects are basically zero. You can maybe get away with saying something like "predict less than 1% of the variance in outcomes and may be unreliable".

Ok, now that we have the standardized effect sizes (thank you), we can see that these are near zero and, indeed, unreliable effects. This is where I tried to save me and the authors time by noting I'd heard all these counterarguments about SESOIs before. Unfortunately, the authors try to raise them again, but they are nonsense. I mean come on…surely you can just eyeball those standardized regression coefficients and see they are nearly zero? How much closer can we possibly get to zero before we can just, as human beings with common sense (and perhaps a heavy heart as our bias is typically in the direction of that beloved $p < .05$), say "these are nothing." This is why I advised you not to go searching for defensive papers…I'm aware they exist and they all produce self-gratifying rubbish arguments. I mean, really, .03, .07, .12 as small, medium and large effects? Sorry if a bit of pique is coming through (and I mean no venom toward the authors who have merely been seduced by silly arguments), but this is clear foxes guarding the henhouse rationale.

Setting benchmarks against percentiles is an awful approach to effect size interpretation. Again, that's foxes guarding the henhouse. Most effect sizes in psychology are near zero. Using that to interpret tiny effects as "large" is just unscientific. And yes, if you thought that my recommendations means most social science longitudinal effects are inconclusive, that is precisely the perspective I take. Indeed, most of social science involves scratching away at wisps of fantasy, which is why we remain as much a pseudoscience as a real science. Only by embracing greater standards of rigor can we move beyond this.

Whether effect sizes are cross-sectional or cross-lagged doesn't matter. The problem is the precision of our measures, which are low (and which the authors now specifically acknowledge which I thank them for), as well as the numerous biases that creep into our methodologies, which tend to favor hypotheses. Even with cross-lagged, these easily explain results as small as those you find.

The section that begins: "People who reported greater frequency of using text messaging…" you will need to note that all such effects (Save 1 I think) are below a minimum SESOI (if you use the more generous .10 level) and should be regarded as inconclusive/unreliable.

Similarly, the paragraph that begins: "However, some platforms' use was associated with lower subsequent life satisfaction in some subgroups" also will need interpretive editing along the same lines. First, you can't make that kind of opening claim as these results are all too small/unreliable for such a claim. Second, again, you're going to need to acknowledge that these are all (again except maybe 1, but even that was probably chance) below any reasonable SESOI and are likely unreliable.

Again, just look at your own tables 2 and 3. Surely you can see, looking at your own bolds, that this is what expect from just some random statistical fluctuations, not "real" effects?

So, your effect sizes are still basically zero, and you need to interpret them accordingly.

The authors also suggest that they shouldn't use a SESOI because they didn't have one a priori. Failing to set one a priori (a mistake) is not a rationale for not correcting that mistake post-hoc. Yes you SHOULD have set your SESOI a priori, but alas you did not. Arguing that you don't need to use one now is like saying that you didn't say you were going to correct for multicollinearity a priori, thus you're not going to correct it when the problem is pointed out to you by a reviewer. I think you are fine to transparently say that you hadn't set the SESOI a priori, but that the unreliability of tiny effect sizes was subsequently pointed out during the review process and you put one in place according. *This is literally why we have reviewers and even preregistered papers are still at the mercy of peer review.*

And I'll repeat from my first review…this is a MUST for my endorsement of the paper. Yes, I am very aware you can find

papers arguing for a complete lack of any serious critical thinking when it comes to interpreting effect sizes in a way that is emotionally satisfying for social scientists. Do not get led down those rabbit holes. No science lives there.

So: 1) edit the title, 2) edit the abstract and discussion accordingly, and 3) put in that SESOI.

Cordially,
Chris Ferguson

Reviewer #2 (Remarks to the Author):

I was Reviewer 1 in the previous round of reviews. My general impression of the manuscript was somewhat mixed. I agreed with the authors' limitations/conjectures that motivated the research and I stated that the method (5-wave longitudinal) and sample (diverse, US) are of very high quality and could provide new answers. I also had some more critical comments. Please find below my most relevant thoughts and questions regarding the revised version of the manuscript.

Let me start with the new title:
The current discussion in academic and applied settings revolves around negative effects of social media. Against this background, I think changing the title to "Social Technology Use Is Not Associated with Greater Subsequent Life Satisfaction: A Five-Wave Panel Study of U.S. Adults" makes little sense. It sounds to me like some straw man line of argumentation. Likewise (abstract) "is it increasing our life satisfaction?" should better be "is it decreasing or increasing our life satisfaction?"

Now to my earlier points and how I assess the present manuscript and the changes made:

1. General remarks
All points were sufficiently addressed.

2. Missing effect sizes
This point was sufficiently addressed.

3. Subgroup analyses without testing for moderation effects
Okay, makes more sense now.

4. Narrative / Framing and 6. Alpha-error inflation
My points 4 and 6 are connected, so I discuss them together. In my first review I mentioned the authors' tendency to overplay the results that fit a negative story and to downplay the results contradicting it. I strongly encouraged the authors to convey the complexity of their results and set a more balanced tone. I am still not satisfied with some of the methodological decisions and interpretations:

The authors did not specify a priori what exactly they aimed to test (e.g., by pre-registering specific associations) but rather analyzed the longitudinal associations in an exploratory fashion. Table 2 shows the resulting 100 longitudinal associations. Only a few of the associations are significantly different from zero at the conventional .05 alpha level. This is what we would expect by chance alone.
I appreciate that the manuscript now includes an alpha error correction. It is first introduced here (p. 9): "Due to the large number of statistical tests, it is possible that the negative relationship between earlier Twitter/X use and later life satisfaction was due to chance (i.e., Type 1 error). To provide a more conservative test of the relationships, therefore, we applied a false discovery rate (FDR) correction. Given that we examined 10 social technology platforms, we adjusted the critical alpha by the following formula: adjusted-alpha = (0.05*k)/m, where m = 10 representing the total number of tests, and k is the rank-order of each of the ten tests by p-value, from 1 being the smallest to 10 being the largest. This gives us the following ten adjusted alpha values: .005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04, 0.045, 0.05. By these adjusted values, the negative relationship of earlier Twitter/X use and later life satisfaction did not reach statistical significance (beta = -.033, p = .0312, padj ≈ .312)."
Five comments:
- The description of the method as cited is confusing, it seems that in the first part the critical value is adjusted, in the second part in which the method is applied, the empirical p-value is adjusted. Both may be feasible and lead to the same conclusion anyway, but please be clearer in the presentation.
- If the BH-FDR correction is taken seriously, then the sentence in the first section on p.9 … "people who reported higher Twitter/X use compared to their own average subsequently experienced lower life satisfaction" is incorrect.
- The graphs should be in line with the correction (e.g., Figure 2). As it is, they are contradictory. This also applies to necessary changes regarding subgroup differences (see my next point).
- Subgroup differences: The BH-FDR is conducted based on 10 tests in each column. However, 100 tests were conducted, so m should be 100 and the critical p-value (for the largest association) would be .0005. It seems that no association is significantly different from zero.

In sum, my take-away is that the 'cross-lagged effects of earlier self-reported frequency of social technology use on later life satisfaction' did not exceed levels expected from chance alone (H0 stands).
In stark contrast, the abstract and other sections of the manuscript suggest otherwise: from the abstract […] "younger adults' Facebook and Instagram use and racial minorities' Twitter/X use predicted lower life satisfaction. No social technology,

including texting or video calling, predicted improved subsequent life satisfaction in any group." […]. This citation is taken out of the context, but still, I expect that this is what many superficial readers, journalists and LLMs perceive as the take-away message.

Other changes in response to comments 4 and 6 (like adding the effect sizes and related benchmarks) have improved the manuscript. I accept that the authors prefer not to highlight the between-subjects-associations.

5. Sample and Age
Okay, fine with me.

7. Activities/content pursued
Okay.

In sum, I think the manuscript is improved, but the authors' interpretation of the data is misleading. The data do not show within-person effects of technology use on life satisfaction. This needs to be changed.

Version 2:

Decision Letter:

Our ref: NATHUMBEHAV-26010366B

6th July 2026

Dear Dr. Kushlev,

Thank you for submitting your revised manuscript "Social Technology Use and Life Satisfaction: A Five-Wave Panel Study of U.S. Adults" (NATHUMBEHAV-26010366B). It has now been seen by the original referees and their comments are below. As you can see, the reviewers find that the paper has improved in revision. We will therefore be happy in principle to publish it in Nature Human Behaviour, pending minor revisions to satisfy the referees' final requests and to comply with our editorial and formatting guidelines.

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements within a couple of weeks. Please do not upload the final materials and make any revisions until you receive this additional information from us.

Please do not hesitate to contact me if you have any questions.

Sincerely,

[redacted]

[redacted]
Nature Human Behaviour

Reviewer #1 (Remarks to the Author):

This version of the paper is much, much better and I think best reflects the data. I like the addiction of the Bayesian analysis (which I think was the editor's idea if I understood).

I admit I probably gave the authors a bit of a hard time throughout, but I appreciate their diligence and the paper is much better for it.

I have no further comments.

Reviewer #2 (Remarks to the Author):

I was Reviewer 2 in the previous round of reviews. All my remarks were sufficiently addressed. I am not an expert in the Bayesian analyses that were conducted. As a non-expert reader, the analyses seem reasonable.
One final comment: The authors start by introducing the field of social media effects among adolescents and they return in the discussion section and mention social media bans. Their participants, however, are adults. This is kind of a disconnect. I encourage the authors to clarify upfront why a study on adults (rather than adolescents) is relevant and I also encourage the authors to acknowledge more explicitly that results may differ for adolescents.

INTERNAL USE ONLY AIP modules required: "Behavioural Science"

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.
In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.
The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.
To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

POINT-BY-POINT RESPONSES TO REVIEWERS

We thank the reviewers for their thoughtful reviews. Below, we provide detailed responses to each of the reviewers' comments (our responses in *Italics*). We are happy to make further revisions as deemed necessary.

We have addressed the two key issues raised by both reviewers about (1) providing corrections for Type 1 error for multiple comparisons and (2) providing the standardized effect sizes. Specifically, we provide *Benjamini–Hochberg false discovery rate (BH-FDR) corrections.*

We have also addressed other comments in the revision, as detailed and explained point by point below. Finally, we provide a rationale below for why we were not able to directly address three specific comments in the revised manuscript.

Reviewer #1 (Remarks to the Author)

The authors present a 5-wave longitudinal study on the relationship between social media use and life satisfaction. The sample is 1,966 adults (Wave 1) from the Understanding America Study (UAS) panel, with an average age of around 40 years. I agree with the authors that more large-scale longitudinal studies on this topic are required, in particular those that examine different platforms and/or different behaviors and/or user groups. Thus, the current project could make an important contribution. That said, my general impression of the manuscript is somewhat mixed. Please find below my most relevant thoughts and questions:

We are glad the reviewer thinks this project could make an important contribution.

1. General remarks
A) I believe that the limitations/conjectures that motivate the research (pages 3-4 of my PDF, no page numbers are provided) are valid, even if they are not particularly novel. For example, the idea that the influence of social media use should – based on current theory and evidence – vary for different platforms/activities is common ground (see, for example, work by Valkenburg, or Bayer et al., 2020). Please soften the novelty aspect of these points.

As requested, we have softened the novelty of addressing each of the four gaps we have outlined as follows (Page 5): "While previous research has tackled each of those four gaps (Beyens et al., 2020; Orben et al., 2019; Valkenburg et al., 2021), we are not aware of any studies that have addressed them together particularly in a diverse sample of adults."

We have also made sure to add citations when identifying each gap so that it is clearer that we are not identifying those gaps for the first time. For example, on Page 2, we state and cite: "This is problematic because, of course, different technologies feature distinct experiences (Bayer et al., 2020; Valkenburg, 2022)."

B) The method (5-wave longitudinal) and sample (diverse, US) are of very high quality and could provide new answers (see below). The method and analysis approach are clearly presented. The tables and figures are helpful. I know that RI-CLPM has been criticized (e.g., by Lucas, Roberts). I am not a true expert on the methodological discussion; that said, my impression is that it is defensible to use it at this point.

Thank you. We agree.

2. Missing effect sizes
Until late in the discussion, effect sizes were neither mentioned nor reported. I guess we cannot infer effect sizes from the unstandardized coefficients reported throughout. This prohibits readers from drawing firm conclusions.

We now report standardized coefficients and use state-of-the-art benchmarks to interpret their size in the context of RI-CLPM. See the new Tables 2 and 3.

3. Subgroup analyses without testing for moderation effects
In experiments or meta-analyses, it is a standard procedure to test for moderation effects before reporting simple main effects / simple slopes. If the moderation is non-significant, the conclusion is that groups do not differ; simple main effects /simple slopes are usually not reported. I was wondering why such tests of moderation were not conducted. Also, the subgroup analyses were not pre-registered (see Narrative / Framing).

This is the approach we typically would take. However, the only way to test for moderation in the RI-CLPMs is by comparing constrained and unconstrained models. All moderation effects are statistically significant. These are reported in the supplementary materials. We did not want to highlight the significance of these effects too much because these types of moderation analyses are much less conservative than typical moderation analyses, which are not possible here without adopting a completely different analysis approach.

4. Narrative / Framing
The social media – well-being link is contested and complex, whereas bestselling popular science books tell a one-sided, negative story (Haidt, Twenge, etc.). Some scientists may feel pressure to follow the negative narrative. I recognized a tendency in this manuscript to overplay the results that fit a negative story and to downplay the results contradicting it. I strongly encourage the authors to convey the complexity of their results and set a more balanced tone. This may include the following:

After our revisions, we believe we have further emphasized the lack of both positive and negative effects. We provide more details on how we have addressed each point below.

a) Acknowledge (small, very small?) effect sizes

We have reported standardized effect sizes (See Tables 2 and 3). However, based on past research with RI-CLPM, we conclude that most effect sizes can be characterized as small to medium rather than very small. This is because within-subjects effects in RI-CLPM are expected to be considerably smaller than the corresponding between-subjects effects that we typically see in the psychological literature. That said, we have taken care to emphasize that most of the cross-lagged effects from social technology use to life satisfaction were null, especially after accounting for Type 1 error and multiple comparisons.

b) Note that more of the cross-sectional relationships are positive than negative, as early as in the abstract

Abstracts in the current journal are limited to 150 words. As such, we were not able to incorporate the cross-sectional findings in the abstract. That said, we have completely revised the abstract to take a much more nuanced tone about the (lack of) effects of social technology on life satisfaction.

We also note that highlighting and interpreting the cross-sectional results comes in direct conflict with the opinion of Reviewer 2 about not interpreting any cross-sectional effects smaller than r = .2. All of the cross-sectional effects were closer to r = .1. Though we have argued against the appropriateness of setting a post-hoc smallest effect of interest (SESOI) after having seen the results, we are also hesitant to over-interpret these results given Reviewer 2's concerns.

For the reviewer's convenience, the revised abstract is pasted below.

> "Abstract: Social technology—from texting to Facebook and TikTok—has transformed our social lives. But is it increasing our life satisfaction? Using five waves of probability-based panel data from U.S. adults (N = 1,966), we examined bidirectional associations between changes in the use of ten common social technologies and subsequent changes in life satisfaction over one year—both overall and within demographic subgroups. Overall, social technology use showed little evidence of predicting subsequent life satisfaction. Negative associations emerged only for specific platforms within specific groups: younger adults' Facebook and Instagram use and racial minorities' Twitter/X use predicted lower life satisfaction. No social technology, including texting or video calling, predicted improved subsequent life satisfaction in any group. In the reverse direction, increases in life satisfaction predicted modest increases in calling behavior (e.g., FaceTime), but not in scrolling behavior (e.g., browsing social media). While the present work calls into question narratives that social technology harms adult life satisfaction, it also highlights their consistent failure to make people happier."

c) Interpret the alternative causal pathways

The revised manuscript now includes much more consistent discussion and interpretation of the reverse causal pathways from earlier life satisfaction to later social technology use. These include:

Page2 (Abstract): "We also found that, although an increase in life satisfaction was subsequently associated with more self-reported calling behavior (e.g., on FaceTime), it was never associated with meaningful changes in scrolling behavior (e.g., on social media) among subgroups."

Page 9 (Results): "Beyond examining how social technology use is related to subsequent life satisfaction, the RI-CLPMs also examine the reverse: how earlier life satisfaction is related to subsequent technology use (see Figure 2, middle panel). After correcting for multiple comparisons, we find that higher life satisfaction preceded higher self-reported use of FaceTime ($\beta = .072$, $p = .0011$, $p_{adj} \approx .011$), but was unrelated to subsequent changes in the self-reported use of any other social technologies (see Table 3)."

Page 12 (Results, Subgroup Differences Subsection): "We can conduct the same comparisons using reverse cross-lagged effects—from life satisfaction to subsequent self-reported frequency of use. Here, we find that an increase in life satisfaction in many groups subsequently predicted greater frequency of FaceTime, including in females ($\beta = .093$, $p = .0008$, $p_{adj} \approx .008$), people aged 30 or older ($\beta = .074$, $p = .0045$, $p_{adj} \approx .045$), people without a Bachelor's degree ($\beta = .114$, $p < .0001$, $p_{adj} < .001$), people with lower incomes ($\beta = .111$, $p < .0001$, $p_{adj} < .001$), and racial or ethnic minorities ($\beta = .122$, $p = .0011$, $p_{adj} \approx .011$). See Table 3 for all standardized coefficients. These effects were consistent after excluding FaceTime nonusers, suggesting that the effects cannot be explained by differences between Android and iPhone users (see SOM for details). This pattern suggests that greater life satisfaction subsequently predicts more frequent calling behavior (e.g., on FaceTime) but not more frequent scrolling behavior (e.g., on social media platforms)."

*Page 21 (Discussion, First Paragraph): "*We also observed one consistent pattern in the reciprocal direction: People—both overall and across various subgroups—who had higher life satisfaction were subsequently more likely to engage in calling behavior (e.g., Facetiming others more often), but not more likely to engage in scrolling behavior (e.g., on social media). This is consistent with evidence that happier people are more likely to invest time in relationships and social interactions (Elmer, 2021; Luhmann et al., 2013)."

d) Explain in greater detail what the three effects (between-subjects associations, the two within-subjects paths) tell us, keeping an impartial standpoint.

In the revised manuscript, we dedicate a full page to reporting and contrasting the between-subjects associations with the cross-lagged associations. To more clearly illustrate what each tells us, we have included the following paragraph (Page 10): "To illustrate the value of distinguishing between cross-lagged and between-subjects effects, we can contrast the results from all three panels of Figure 2 for TikTok, where the between-subjects and cross-lagged effects diverged. Using TikTok more does not predict subsequent changes in life satisfaction (left panel), nor does having greater life satisfaction predict changes in subsequent TikTok use

(middle panel); but the kind of people who use TikTok more tend to have less life satisfaction (right panel), though this relationship could be driven, of course, by other factors such as age."

e) Be explicit about the danger of alpha-error cumulation and why it may or may not affect your results

We have corrected for Type 1 error in the revised manuscript.

5. Sample and Age

Although much of the theory and research around the social media – well-being link is focused on adolescents and/or emerging adults, the sample here is adults with an average age of 40 years. Why? Importantly, I see an opportunity for new insight. Instead of splitting the sample at age 30, the authors are encouraged to compare younger and older adults. Much of the theory and research suggests that older age groups profit from social media use (see Hwang et al., 2021; Francis-Levin et al., 2025).

We investigated the practicability of this possibility. However, as noted in the paper, while people under 30 were oversampled, older adults were not. Thus, for example, less than 10% of the sample were 65 or older, and even a more liberal cut-off, such as 60+ as used in Francis-Levin et al. (2025), would only comprise less than 15% of the total sample. While enough to conduct meaningful analyses, this will be substantially smaller than any of the other groups analyzed in the present study. Given the FDR-adjustment for Type 1 error, our rough power analyses indicated that we will need the largest effect size to be beta = .26 to detect a significant effect for the 262 individuals in the 60+. This would be substantially larger than any of the cross-lagged effects we observed in any of the other groups, making this an unfair comparison. In short, we do not feel that the study has enough power to actually test this intriguing possibility. We also worry that, at this stage, re-running the analyses with a different age cut-off may be perceived by some as fishing for results, potentially increasing Type 1 errors.

6. Alpha-error inflation?

This could be considered a fatal flaw of the manuscript. Given the many variables examined, I wonder whether results (e.g., the effect of Twitter/X among men as shown in Figure 3a) could simply be products of chance. The project is not very theory-guided, which is okay in my opinion, but this makes this potential problem worse, I guess. For example, the analyses underlying Figure 3 include 20 associations, one of them significantly different from zero. This one association is mentioned in the discussion and may be the target of media attention, etc. If alpha error corrections were made, I missed them.

The new manuscript has been substantially revised to include FDR-adjustments for multiple comparisons. We have also been careful to revise the manuscript so that it is clear that most of the effects are null rather than negative (or positive).

7. No study is perfect. Still, the limitations section should include the caveat that the study is silent on the activities/content pursued, arguably the largest source of variance in social media effects (e.g., on messengers, users can regularly have pleasant conversations with best friends or regularly be the target of bullying)

*We have revised the manuscript to discuss this point explicitly in the Discussion as part of the study limitations. Specifically, on Page 22, we now discuss as follows: "*In addition to being self-reported, our measure of social technology use captured only overall use rather than how and for what purpose people use these platforms. Our approach, therefore, does not allow us to establish whether the null associations we observed are due to a lack of any effects of a given platform or due to the presence of both positive and negative effects of different types of use that cancel each other out."

Reviewer #2 (Remarks to the Author)

1. I appreciated the opportunity to review this manuscript. I think there are a lot of strengths and, in particular, they do a nice job with their literature review, given controversies in this area. However, I did not find the database to be convincing and am concerned that they overinterpret very weak results. Overall, I find the evidence here more convincing that there is no impact for social media use, given the limits of the dataset, rather than that there are either "benefits" or "harms". These effects are just too weak to support any hypothesis, as I'll get to below.

In the revised manuscript, we have further emphasized the lack of both positive and negative effects.

2. First, minor issue, but don't use words like "harm" (or, for that matter, "benefit"). That is causal language, and this field, including this study, are overwhelmingly correlational. There's no means to turn correlational data into causal evidence, despite how some may argue. Further, such language is manipulative and unscientific. It does not belong in a scientific article.

We have removed references to "harm" and "benefit" or similar terms from the revised manuscript.

3. As mentioned, the authors offer a good lit review, and I appreciate that.

We appreciate the positive note.

4. The authors need to provide standardized coefficients. These unstandardized coefficients look pretty unimpressive, and I'm concerned that the authors fail to interpret some of them in a

rigorous way (indeed, this appears confirmed by their Pearson correlations). But without standardized coefficients, it can be hard to be sure.

We have provided standardized coefficients in the text and in Tables 2 and 3. Using state-of-the-art benchmarks for RI-CLPM, we can see that the standardized coefficients of most relationships that remained significant after the FDR corrections approached or surpassed the benchmark for cross-lagged effects of medium effect size.

5. As a main issue…and this is non-negotiable. The authors need a smallest effect size of interest (SESOI). A common problem throughout social science, but including this field specifically, is false positive interpretations based on weak effect sizes. Unfortunately, with larger sample sizes (like this one), almost everything becomes "statistically significant," which can mislead scholars into interpreting results as evidence supportive when they shouldn't. We can now be sure that many of those "statistically significant" but weak effect sizes are false positives driven by methodological noise. Please read Ferguson & Heene (2021) carefully to understand this common problem, one which likely maintains a lot of social science as pseudoscientific and maintains a high false positive rate (Ioannidis, 2005). At this juncture, I am recommending a r or β of .20 as a minimum threshold for interpreting results as hypothesis supportive. Please note this is not a "practical significance" issue (though that also should be considered), but simply an internal validity problem issue. Our methodological precision (and particularly for a rather crude dataset like this one) just isn't sufficient for interpreting results smaller than that as anything other than inconclusive. Note once again, this isn't something I'm interested in a counter-argument on…I've heard them all before, and they are wrong. This is a requirement for me to endorse the article.

We agree that articulating a smallest effect size of interest is valuable in principle because of the "crud factor" problem defined and demonstrated in Ferguson & Heene (2021). *However, we were not able to set a SESOI in the revised manuscript for two related reasons. First, doing so post hoc would in effect be analogous to selecting an inferential threshold after inspecting the results. Second, unlike the inferential threshold of alpha = .05, we could not find any direct evidence for what this SESOI threshold should be, let alone widely agreed upon guidelines. We note that Ferguson and Heene's (2021) proposed "basement cut-off" (e.g., r ≈ .10) is derived from large cross-sectional, between-subjects zero-order correlations—rather than cross-lagged effects. As a result, defining a SESOI at this stage of the process could potentially create, or at least be seen as creating, the very same issue that it is trying to resolve. Specifically, adopting a post hoc SESOI risks replacing one form of interpretive flexibility (overreliance on p-values in large samples) with another (retrofitting an evidentiary cutoff after inspecting the results), which would not meaningfully address the underlying crud-factor concern.*

We hope, however, that the revised manuscript partially addresses the reviewer's underlying concern in several ways.

First, we have provided benchmarks for small, medium, and large effects and standardized effect sizes so that readers can more readily interpret the meaningfulness of the results. Based on state-

of-the-art guidance for interpreting the size of cross-lagged effects in RI-CLPM (Orth et al., 2024), we consider .03 as a small effect, .07 as a medium effect, and .12 as a large effect.

Second, after the corrections for Type 1 error, only three subgroup effects from earlier social technology use to later life satisfaction remained significant, and there were no significant omnibus associations. Thus, whether relying on SESOI or a traditional alpha = .05, our conclusions remain largely the same: the association of social technology use and life satisfaction is largely null.

Third, we have taken care not to overinterpret or overclaim anything about any one specific association, but rather documenting that a few negative but weak relationships emerged. We have revised the manuscript to emphasize this overall pattern in our results.

Regarding the specific proposed SESOI (β = .20) suggested by the reviewer, we think this may not be appropriate for several reasons.

First, β = .20 is calibrated to the wrong statistical object. Cross-lagged effects are not between-subjects correlations. In cross-lagged models, the prospective path is estimated while controlling for (i) prior levels of the outcome (autoregressive stability) and (ii) the concurrent association between constructs at baseline. Orth et al. explicitly caution that "conventions established for correlations should not be applied to cross-lagged effects" because these controls mechanically limit the plausible magnitude of cross-lagged betas and make them systematically smaller than cross-sectional associations.

Second, the best available evidence on typical cross-lagged effect sizes directly contradicts β = .20 as a "minimum." Orth et al. (2024) provide empirical benchmarks from a quasi-representative sample across major psychology subfields. For CLPM cross-lagged effects, the 25th/50th/75th percentiles are .03/.07/.12, and for RI-CLPM they are .02/.05/.11. Put bluntly: β = .20 is not a "minimum threshold"; it is approximately the 90th percentile for CLPM cross-lagged effects (and even higher relative to RI-CLPM, where the 90th percentile is .17). Setting the SESOI at .20 would therefore imply that most of the longitudinal cross-lagged literature—including effects Orth et al. classify as "large" (≈ .12)—should be treated as "inconclusive by default," which is not a methodological principle so much as a redefinition of the entire field's effect-size scale. That position would require strong empirical justification specific to within-person cross-lagged parameters, which, to our knowledge, has not been established.

6. The authors also need Bonferroni corrections due to the vast number of analyses they conduct. There's a lot of capitalizing on chance here.

We have provided adjustments for Type 1 error. However, we have provided FDR corrections. Type 1 is important, but Bonferroni is notoriously overly conservative, and can result in inflating Type 2 errors, which is also important. FDR corrections balance these two competing issues.

That said, we have substantially revised the manuscript to highlight the largely null findings rather than overfocusing on the few negative relationships observed.

7. As mentioned, this dataset appears crude to me. We're relying on a few crude questions that are self-reports. This raises issues of demand characteristics and single-responder bias, clear sources of exactly the kind of noise I was referring to above. We also know that people who are more life dissatisfied tend to overestimate their time online, adding a source of noise/bias. That this dataset uses self-report rather than an objective measure of time spent raises yet another concern.

We have addressed these points as limitations in the Discussion. Specifically, we have added the following paragraph on Page 22 of the Discussion: "Our study is not without limitations. Our measures rely on a small number of self-report items, which introduces well-known sources of noise and bias. Because both social technology use and life satisfaction were reported by the same individuals, observed associations may partly reflect single-responder bias, whereby correlations are inflated or distorted by shared response tendencies rather than substantive relationships. Relatedly, prior work suggests that individuals who are more dissatisfied with their lives tend to provide more inaccurate estimates of their social media use (Sewall et al., 2020), introducing systematic measurement error that may further obscure true effects."

Please note that we could not find a citation that directly supports the statement that "people who are more life dissatisfied tend to overestimate their time online, adding a source of noise/bias." Thus, we did our best to make the same argument with similar evidence.

8. I would delete the sentence "This lack of improvement in wellbeing stands in stark contrast to the measurably positive effects of in-person social activities…" As it happens, I'm actually working on a dataset that finds that engagement with outside activities actually is associated with negative mental health (likely due to overscheduling).

We have deleted this sentence.

9. The authors offer only a poor rationale for small effects: "Even these statistically small effects, however, may be practically significant for two reasons…." These are not true. More likely, effects are overestimated, creating false positives due to issues I outlined above. I really wish that scholars wouldn't do this "every effect size is sacred" routine. It's self-serving and a common source of misinformation and low rigor. Please, please, I know it is common…but it's pseudoscientific and due to bad training. Do not do this. You have "crud" effects. That's ok. You don't have to try to make them something they are not.

We have now revised the paragraph in question to highlight the crudeness of the measures (rather than make much ado about nothing). As seen on Page 22, we have completely removed the argument that "statistically significant effects may be practically significant." Instead, we explicitly made the opposite argument that we could be overestimating the size of the effects. See response to Point 7 for details.

10. The over time reliability is low. The authors try to wave this aside, but I am unconvinced. This seems a problem to me.

We appreciate the reviewer's concern regarding the reliability of within-person change. In the revised manuscript, we now cite the following paper to show that our interpretation of the reliability is based on best practice rather than on a casual waving aside: Nezlek, J. B. (2017). A practical guide to understanding reliability in studies of within-person variability. Journal of Research in Personality, 69, 149–155. https://doi.org/10.1016/j.jrp.2016.06.020.

As discussed by Nezlek, it is important to consider contextual differences between the reliability of within-person change (Rcn) and traditional temporal stability (i.e., test–retest reliability). Test–retest reliability assesses whether a construct remains stable across time, whereas Rcn evaluates the extent to which observed wave-to-wave variation reflects systematic within-person change rather than random noise. An Rcn of .47 indicates that nearly half of the observed within-person variation in life satisfaction reflects reliable change.

As further discussed by Nezlek, in longitudinal designs, within-person reliability is expected to be lower than between-person reliability because individuals typically vary within a relatively narrow range over time and because within-person measures are often intentionally brief, using fewer items to reduce participant burden across repeated waves. As a result, distinguishing true change from measurement error is inherently more challenging than in between-person comparisons. Consistent with this expectation, between-person reliability in the present study was excellent (.94).

Nezlek further shows that occasion-level reliability is often substantially lower than person-level reliability because such measures are designed to capture dynamic, context-sensitive states rather than stable traits. Moreover, Nezlek (2017) emphasizes that low occasion-level reliability does not render within-person analyses invalid. Contemporary multilevel models explicitly account for unreliability via variance decomposition, which reduces the influence of noisy estimates on fixed effects and significance tests. As a result, lower within-person reliability primarily attenuates effect sizes rather than generating spurious associations. In this sense, the observed estimates should be interpreted as conservative. Of course, given the reviewer's strong opinion about the reasons why the study's design could overestimate effect sizes, we have refrained from making this argument anywhere in the manuscript.

In light of these considerations, we believe the observed reliability provides sufficient signal to examine systematic within-person processes (cross-lagged effects), which was the primary aim of the study.

11. Overall, the results here look to be null in both directions. That's fine. But the authors searching through subgroup analyses and overinterpreting near-zero effects are adding a lot of capitalization on chance and bias that's going to cause more misinformation than anything useful. I hope these comments are helpful for the authors.

We believe that the revised manuscript now much more clearly highlights the overall lack of both positive and negative effects of social technology platforms. Here, we highlight some of the places where we have taken care to highlight this lack of effects throughout the manuscript.
Abstract (Page 2): "…Overall, social technology use showed little evidence of predicting subsequent life satisfaction."… "While the present work calls into question narratives that social technology harms adult well-being, it also highlights their consistent failure to make people happier."
Results (Page 9): "In short, we found no robust evidence of effects—positive or negative—of social technology use on subsequent life satisfaction in this aggregate analysis."
Discussion, opening sentence (Page 21): "Among a diverse sample of U.S. adults, we found little evidence that social media, texting, email, and video calling either harm or improve life satisfaction."

Discussion, opening paragraph (Page 21): "Specifically, we found no social technology platform wherein usage predicted higher subsequent life satisfaction, either across or within any specific demographic group. Instead, we found some evidence that social technology use predicted lower life satisfaction in select groups, but the use of most platforms in the majority of people did not predict lower life satisfaction."

In conclusion, we thank the reviewers and the editors for their consideration of our manuscript.

Reviewer #1 Comments

Opening Remarks

I appreciate the authors' revisions here. Many of them are very positive, and I do believe the authors have been serious in taking the reviewer recommendation. However, I am still concerned by the overinterpretation of some near zero (indeed inconclusive) effect sizes and the authors got seduced by some bad arguments (i.e., Orth,. Indeed, it looks like the authors basically find null effects in both directions, and they should simply acknowledge that.

Response:

We greatly appreciate your push for higher methodological rigor. We agree that the previous version overinterpreted near-zero associations. In this revision, we have completely abandoned those counterarguments and fully embrace the bidirectional null findings. As detailed in our overarching response, we have transitioned to a Bayesian parameter estimation framework that explicitly evaluates whether associations are practically meaningful beyond crud. Consequently, we now explicitly acknowledge and focus on the lack of credible evidence for bidirectional associations throughout the manuscript.

Comment 1

The title should be changed to reflect the bidirectional null results, "Social Technology Use Is Not Associated with Greater OR REDUCED Subsequent Life Satisfaction." (caps just highlight the needed addition, not yelling, for the record).

Response:

We have revised the title to *"Social Technology Use and Life Satisfaction: A Five-Wave Panel Study of U.S. Adults."*

Note that NHB requires that null effects are described in terms of no credible evidence as follows: "*For studies reporting statistically null results, make sure that you do not interpret the absence of evidence as evidence of absence. There is no statistical test that can demonstrate the absence of an effect. Statements such as 'There is no association between x and y.' or 'X has no effect on Y.' must be revised to read '[No/Little] credible evidence of an association between x and y.' or '[No/Little] credible evidence that X affects Y.'*"

Comment 3

The sentence in the abstract that begin: "Negative associations emerged only for specific…" is going to need to be edited or removed to, once again, reflect that these effects are basically zero. You can maybe get away with saying something like "predict less than 1% of the variance in outcomes and may be unreliable".

Response:

We have completely removed the sentence regarding negative associations from the abstract. The abstract has been entirely rewritten to reflect that the cross-lagged associations from earlier social technology use to life satisfaction were not large enough to indicate any kind of meaningful association, aligning with your assessment that these associations are basically zero and represent methodological noise.

Comment 4a

Ok, now that we have the standardized effect sizes (thank you), we can see that these are near zero and, indeed, unreliable effects. This is where I tried to save me and the authors time by noting I'd heard all these counterarguments about SESOIs before. Unfortunately, the authors try to raise them again, but they are nonsense. I mean come on…surely you can just eyeball those standardized regression coefficients and see they are nearly zero? How much closer can we possibly get to zero before we can just, as human beings with common sense (and perhaps a heavy heart as our bias is typically in the direction of that beloved $p < .05$), say "these are nothing." This is why I advised you not to go searching for defensive papers…I'm aware they exist and they all produce self-gratifying rubbish arguments. I mean, really, .03, .07, .12 as small, medium and large effects? Sorry if a bit of pique is coming through (and I mean no venom toward the authors who have merely been seduced by silly arguments), but this is clear foxes guarding the henhouse rationale.

Response:

We hear your frustration and sincerely appreciate your candid guidance that has produced a much stronger and more rigorous manuscript in our opinion. We have removed all of the previous defensive arguments and rationales regarding smaller effect sizes. We have fully adopted your recommendation to use a Smallest Effect Size of Interest. We implemented a Region of Practical Equivalence in our Bayesian models based exactly on the a |0.10| threshold recommended by Ferguson and Heene 2021. Our interpretations now reflect that these near-zero standardized coefficients represent nothing beyond methodological noise.

Comment 4b

Setting benchmarks against percentiles is an awful approach to effect size interpretation. Again, that's foxes guarding the henhouse. Most effect sizes in psychology are near zero. Using that to interpret tiny effects as "large" is just unscientific. And yes, if you thought that my recommendations means most social science longitudinal effects are inconclusive, that is precisely the perspective I take. Indeed, most of social science involves scratching away at wisps of fantasy, which is why we remain as much a pseudoscience as a real science. Only by embracing greater standards of rigor can we move beyond this.

Response:

We agree that adopting a standard of rigor is necessary, and we now rely exclusively on a ±0.10 threshold to interpret our findings using Bayesian ROPE approach.

Comment 4c

Whether effect sizes are cross-sectional or cross-lagged doesn't matter. The problem is the precision of our measures, which are low (and which the authors now specifically acknowledge which I thank them for), as well as the numerous biases that creep into our methodologies, which tend to favor hypotheses. Even with cross-lagged, these easily explain results as small as those you find.

Response:

We agree with this assessment. By applying the uniform 0.10 threshold across our analyses via the Bayesian Region of Practical Equivalence, we avoid making excuses for small effect sizes based on the cross-lagged design and appropriately attribute them to negligible fluctuations or measurement imprecision.

Comment 4d

The section that begins: "People who reported greater frequency of using text messaging…" you will need to note that all such effects (Save 1 I think) are below a minimum SESOI (if you use the more generous .10 level) and should be regarded as inconclusive/unreliable.

Response:

This comment refers to our between-subjects associations. After applying the Bayesian approach, we conclude that there is evidence for positive associations for texting, r = 0.134 (88.50% probability of falling beyond ROPE bounds) and email, r = 0.129 (86.10%), and negative associations for TikTok, and YouTube, r = -0.107 (59.53%), r = -0.106 (59.43%). Several other platforms, such as Snapchat, r = 0.007 (0%), and X (Twitter), r = -0.005 (0%) showed null association, while Facebook, r = 0.065 (7.97%), and Instagram, r = 0.034 (0.47%), showed weak association with little credible evidence that these relationships are due to a meaningful signal.

Please note that the raw Pearson correlations that we had reported in the previous version of the manuscript were indeed slightly smaller because they represent cross-sectional snapshots that conflate stable trait variance, temporary state fluctuations, and random measurement error. By utilizing the Bayesian RI-CLPM, our models separate these components to isolate the pure, between-person trait variance (the random intercepts). Because state-level "noise" and measurement error have been mathematically stripped away—a process known as disattenuation—it is expected that these latent RI correlations will be slightly stronger and more accurate than crude Pearson estimates. Consequently, while we agree with the reviewer that several platforms show unreliable or negligible signals below the Smallest Effect Size of Interest (SESOI), our fully specified Bayesian models demonstrate that texting, email, TikTok, and YouTube do indeed exceed the 0.10 threshold, representing meaningful trait-level associations. We have clarified this distinction between the raw Pearson estimates and the latent RI correlations in the revised manuscript. Since these are between-subjects associations, we do not overinterpret them.

We did adopt the more generous 0.10, defined as the bare minimum in Ferguson & Heene (2021). Given the mostly null results even under this more generous threshold, we do believe that this provides even stronger demonstration of the overall pattern of lack of credible evidence for any associations, thereby addressing your underlying concern about highlighting the null effects (i.e., a reader already skeptical of the Bayesian approach might dismiss the findings if we had set a minimum meaningful effect size of, say, 0.2).

Comment 4e

Similarly, the paragraph that begins: “However, some platforms’ use was associated with lower subsequent life satisfaction in some subgroups” also will need interpretive editing along the same lines. First, you can’t make that kind of opening claim as these results are all too small/unreliable for such a claim. Second, again, you’re going to need to acknowledge that these are all (again except maybe 1, but even that was probably chance) below any reasonable SESOI and are likely unreliable.

Response:

This refers to the cross-lagged associations. We now report that “Across all ten of the subgroups, we found little credible evidence that earlier social technology use predicts more or less subsequent life satisfaction.” We have updated all sections of the paper to reflect this pattern by clearly stating that we lack sufficient credible evidence to conclude these represent real, reliable effects.

Comment 4f

Again, just look at your own tables 2 and 3. Surely you can see, looking at your own bolds, that this is what expect from just some random statistical fluctuations, not “real” effects?

Response:

We agree entirely. Under our new Bayesian Region of Practical Equivalence framework, our interpretation of the data in these tables now explicitly concludes that these represent negligible statistical fluctuations and measurement noise rather than real psychological signals. Rather than reporting p values with each coefficient, we now report the percent of posterior distribution exceeding the ROPE bounds. The tables (see Table 2) clearly shows that these probabilities are 0 or almost 0 for all coefficients going from earlier tech use to later life satisfaction. Only the associations of earlier life satisfaction to later FaceTime had point estimates larger than .10 and a posterior probability greater than 50% in some groups to be larger than .10.

Comment 4g

So, your effect sizes are still basically zero, and you need to interpret them accordingly.

Response:

We have revised the abstract, results, and discussion sections from top to bottom to interpret these effect sizes as basically zero and practically negligible.

Comment 4h

The authors also suggest that they shouldn't use a SESOI because they didn't have one a priori. Failing to set one a priori (a mistake) is not a rationale for not correcting that mistake post-hoc. Yes you SHOULD have set your SESOI a priori, but alas you did not. Arguing that you don't need to use one now is like saying that you didn't say you were going to correct for multicollinearity a priori, thus you're not going to correct it when the problem is pointed out to you by a reviewer. I think you are fine to transparently say that you hadn't set the SESOI a priori, but that the unreliability of tiny effect sizes was subsequently pointed out during the review process and you put one in place according. *This is literally why we have reviewers and even preregistered papers are still at the mercy of peer review.*

Response:

We have completely adopted the SESOI approach via our Bayesian analyses using ROPE.

Comment 4i

And I'll repeat from my first review…this is a MUST for my endorsement of the paper. Yes, I am very aware you can find papers arguing for a complete lack of any serious critical thinking when it comes to interpreting effect sizes in a way that is emotionally satisfying for social scientists. Do not get led down those rabbit holes. No science lives there.

Response:

We discarded the previous rationales, implemented the Smallest Effect Size of Interest of 0.10, and evaluated our Bayesian models against this standard.

Concluding Remarks

So: 1) edit the title, 2) edit the abstract and discussion accordingly, and 3) put in that SESOI.

Response:

We have completed all three of these directives. We updated the title to avoid suggesting an association, completely overhauled the abstract and discussion to focus on the bidirectional null findings, and formally integrated the 0.10 Smallest Effect Size of Interest into our primary Bayesian framework.

Reviewer #2 Comments

I was Reviewer 1 in the previous round of reviews. My general impression of the manuscript was somewhat mixed. I agreed with the authors' limitations/conjectures that motivated the research and I stated that the method (5-wave longitudinal) and sample (diverse, US) are of very high quality and could provide new answers. I also had some more critical comments. Please find below my most relevant thoughts and questions regarding the revised version of the manuscript.

Response:

Thank you for your continued engagement with our manuscript. We deeply appreciate your valuable feedback, which has improved the rigor of our methodology and the accuracy of our interpretations.

Opening Remarks

Let me start with the new title: The current discussion in academic and applied settings revolves around negative effects of social media. Against this background, I think changing the title to "Social Technology Use Is Not Associated with Greater Subsequent Life Satisfaction: A Five-Wave Panel Study of U.S. Adults" makes little sense. It sounds to me like some straw man line of argumentation. Likewise (abstract) "is it increasing our life satisfaction?" should better be "is it decreasing or increasing our life satisfaction?"

Response:

We completely agree with this assessment. Based on your feedback and the feedback from the other reviewer, we recognize that the previous title and framing created an unnecessary straw man and failed to accurately capture the bidirectional null findings. We have extensively revised the manuscript to reflect this. We have revised the title to *" Social Technology Use and Life Satisfaction: A Five-Wave Panel Study of U.S. Adults."* As detailed in our overarching response, our transition to a Bayesian parameter estimation framework confirms that the associations are practically negligible in both directions. Note that NHB requirements specifically require that null associations are interpreted in terms of credible evidence and not stated as "is not associated."

Comment 1

1. General remarks: All points were sufficiently addressed.

Response:

Thank you. We are glad to hear tour revisions addressed your previous concerns.

Comment 2

2. Missing effect sizes: This point was sufficiently addressed.

Response:

Thank you. We are glad to hear tour revisions addressed your previous concerns.

Comment 3

3. Subgroup analyses without testing for moderation effects: Okay, makes more sense now.

Response:

Thank you.

Comment 4a

4. Narrative / Framing and 6. Alpha-error inflation: My points 4 and 6 are connected, so I discuss them together. In my first review I mentioned the authors' tendency to overplay the results that fit a negative story and to downplay the results contradicting it. I strongly encouraged the authors to convey the complexity of their results and set a more balanced tone. I am still not satisfied with some of the methodological decisions and interpretations.

Response:

We concede that the previous version of the manuscript overplayed the negative findings and failed to adequately account for the sheer volume of tests conducted. To holistically address your concerns about alpha-error inflation and the over-interpretation of noise, we have switched primarily to a Bayesian approach This transition allows us to focus on parameter estimation rather than traditional p-values, resolving the multiple comparisons problem while confirming your intuition that the data do not support a negative story.

Comment 4b

The authors did not specify a priori what exactly they aimed to test (e.g., by pre-registering specific associations) but rather analyzed the longitudinal associations in an exploratory fashion. Table 2 shows the resulting 100 longitudinal associations. Only a few of the associations are significantly different from zero at the conventional .05 alpha level. This is what we would expect by chance alone.

Response:

You were entirely correct in your assessment. When evaluating these 100 associations under our new Bayesian framework, we found that they are indeed indistinguishable from chance or methodological noise. By defining a Smallest Effect Size of Interest of an absolute value of b = 0.10 and evaluating the posterior distributions against this region of practical equivalence, we can now confidently state that none of the cross-lagged associations from earlier social technology use to later life satisfaction represent a meaningful psychological signal.

Per your point more directly, however, we have also included results from frequentist analyses as a robustness check. Those analyses lead to the same conclusions as the Bayesian models (see Extended Data Table 4). Per your recommendation, we applied an FDR correction for 110 tests. We corrected for 110 to include the omnibus analyses as well as the demographic subgroup tests. We note, even with this correction, the frequentist analyses were slightly less conservative than the Bayesian analyses, showing some relationships as significant even when the Bayesian analyses indicated little credible evidence that the associations are large enough to be considered meaningful beyond methodological noise.

Comment 4c

I appreciate that the manuscript now includes an alpha error correction. It is first introduced here (p. 9): "Due to the large number of statistical tests, it is possible that the negative relationship between earlier Twitter/X use and later life satisfaction was due to chance (i.e., Type 1 error). To provide a more conservative test of the relationships, therefore, we applied a false discovery rate (FDR) correction. Given that we examined 10 social technology platforms, we adjusted the critical alpha by the following formula: adjusted-alpha = (0.05*k)/m, where m = 10 representing the total number of tests, and k is the rank-order of each of the ten tests by p-value, from 1 being the smallest to 10 being the largest. This gives us the following ten adjusted alpha values: .005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04, 0.045, 0.05. By these adjusted values, the negative relationship of earlier Twitter/X use and later life satisfaction did not reach statistical significance (beta = -.033, p = .0312, padj ≈ .312)." The description of the method as cited is confusing, it seems that in the first part the critical value is adjusted, in the second part in which the method is applied, the empirical p-value is adjusted. Both may be feasible and lead to the same conclusion anyway, but please be clearer in the presentation.

Response:

We have explained the adjustment in terms of the p-values rather than the alpha level.

Comment 4d

If the BH-FDR correction is taken seriously, then the sentence in the first section on p.9 … "people who reported higher Twitter/X use compared to their own average subsequently experienced lower life satisfaction" is incorrect.

Response:

We completely agree, and that sentence has been removed. In our revised manuscript, we explicitly state that we find little credible evidence that any social technology use is positively or negatively associated with subsequent life satisfaction.

Comment 4e

The graphs should be in line with the correction (e.g., Figure 2). As it is, they are contradictory. This also applies to necessary changes regarding subgroup differences (see my next point).

Response:

All graphs and figures have been completely overhauled to align with our new Bayesian approach. Instead of displaying misleading p-value significance markers, our figures now plot the full posterior density distributions alongside a shaded vertical band representing our region of practical equivalence from -0.10 to 0.10. This visually demonstrates to the reader how the vast majority of the effect estimates fall squarely within the range of negligible associations.

Comment 4f

Subgroup differences: The BH-FDR is conducted based on 10 tests in each column. However, 100 tests were conducted, so m should be 100 and the critical p-value (for the largest association) would be .0005. It seems that no association is significantly different from zero.

Response:

As previously noted, we applied a correction for 110 tests to include the omnibus associations and all subgroup analyses associations. You are absolutely correct that across all 110 tests, the cross-lagged associations from earlier social tech use to later life satisfaction are not distinguishable from zero. Our Bayesian analysis of the subgroup differences confirms this precisely. Across all ten demographic subgroups, we report that we lack sufficient evidence to conclude that these specific subgroup associations reflect practically meaningful changes in subsequent life satisfaction, as the posterior probabilities overwhelmingly fail to exceed our predefined threshold. We also note the Bayesian models show that the relationship of earlier life satisfaction to later FaceTime showed some evidence of meaningful associations above .10 in some demographic groups. Specifically, we find credible evidence for these relationships in people without a Bachelor's degree, $\beta = .112$ (67.07%), people with lower incomes, $\beta = .101$ (51.77%), and racial or ethnic minorities, $\beta = .111$ (62.03%). For the associations between earlier life satisfaction and later FaceTime use, the frequentist models with 110 FDR corrections concluded that all of the subgroup associations for FaceTime deemed potentially meaningful by the Bayesian analyses reached statistical significance (lower education, $p_{adj} = 0.0055$; lower income: $p_{adj} = 0.0055$; racial ethnic minority: $p_{adj} = 0.0257$), including flagging the association for females as significant ($p_{adj} = 0.0303$).

Comment 4g

In sum, my take-away is that the 'cross-lagged effects of earlier self-reported frequency of social technology use on later life satisfaction' did not exceed levels expected from chance alone (H0 stands). In stark contrast, the abstract and other sections of the manuscript suggest otherwise: from the abstract […] "younger adults' Facebook and Instagram use and racial minorities' Twitter/X use predicted lower life satisfaction. No social technology, including texting or video calling, predicted improved subsequent life satisfaction in any group." […]. This citation is taken out of the context, but still, I expect that this is what many superficial readers, journalists and LLMs perceive as the take-away message.

Response:

We completely agree with your takeaway. The manuscript has been revised from top to bottom to make this exact point the central conclusion of the paper. We clearly state that the associations are negligible and cannot be differentiated from statistical noise. The abstract now unequivocally states that we found little credible evidence that the use of social technology is associated with later life satisfaction, ensuring that readers, journalists, and language models will correctly interpret the takeaway message as a lack of evidence. To avoid sensationalizing these null results, however, the abstract also notes some of the between-subjects associations that were flagged as meaningful by the Bayesian models and remained significant after FDR corrections for 110 tests. We believe this is in the spirit of your comment by avoiding sensationalizing the null results (and ensuring that the overall pattern of associations at each level is correctly represented).

Comment 4h

Other changes in response to comments 4 and 6 (like adding the effect sizes and related benchmarks) have improved the manuscript. I accept that the authors prefer not to highlight the between-subjects-associations.

Response:

Thank you. We continue to report these standardized estimates transparently while maintaining our primary focus on the within-person cross-lagged parameters.

Comment 5

5. Sample and Age: Okay, fine with me.

Response:

Thank you

Comment 6

7. Activities/content pursued: Okay.

Response:

Thank you.

Concluding Remarks

In sum, I think the manuscript is improved, but the authors' interpretation of the data is misleading. The data do not show within-person effects of technology use on life satisfaction. This needs to be changed.

Response:

We completely agree with this final assessment and have made this the core focus of our revision. By adopting the rigorous Bayesian parameter estimation framework, setting a threshold for practical significance, and overhauling our results and discussion sections, the manuscript now clearly and consistently interprets the data as showing little credible evidence for meaningful within-person associations of technology use and life satisfaction.