Social media positivity bias, or just positivity bias?

*A recommendation by **Veli-Matti Karhulahti** ⓘD based on peer reviews by **Marcel Martončik** ⓘD, **Julius Klingelhoefer** ⓘD and 1 anonymous reviewer of the STAGE 2 REPORT:*

Both research and public debates around social media use tend to involve a premise of positivity bias, which refers to presenting one's life in an overly positive light by various different means. This premise contributes to multiple potentially important follow-up hypotheses, such as the fear of missing out and low self-image effects, due to repeated consumption of positive social media content (e.g., Bayer et al. 2020, for a review). The positivity bias of social media use, itself, has received limited research attention. In the present study, Masciantonio and colleagues (2025) tested positivity bias in the context of three social media platforms: Facebook, Instagram, and X. The experiment involved recruiting participants (n=312) into platform-specific user groups and crafting posts to be shared with friends as well as respective social media audiences. For social media positivity bias to differ from everyday positivity bias, posts in the former should introduce more positive valence in comparison to offline sharing—and if the platforms differ in their encouragement of positivity bias, they should introduce significant between-platform differences in valence. Based on how the participants reported events, the study found no significant differences between everyday and social media positivity bias, but messages posted on Twitter/X had a more negative valence than posts in other social media platforms. The results would be consistent with the implication that people's positive actions in social media simply follow the natural human tendency to present oneself positively to others—a hypothesis that should be investigated in follow-up work. More research attention should also be given to specific design features, which may contribute to platform-specific differences in user habits, as suggested by the distinct valence rate found in relation to Twitter/X. The Stage 2 manuscript was evaluated by three experts (areas: experimental methods, social media, statistics) via in-depth peer review across two rounds, with one reviewer returning to

validate analysis code and methdological accuracy on a final round. Based on the authors' careful responses and revisions, the revised manuscript was judged to meet the Stage 2 criteria and was awarded a positive recommendation.

URL to the preregistered Stage 1 protocol: `https://osf.io/9z6hm` **Level of bias control achieved**: Level 6. *No part of the data or evidence that was used to answer the research question was generated until after IPA.* **List of eligible PCI RR-friendly journals**:

- Collabra: Psychology

- International Review of Social Psychology

- Peer Community Journal

- PeerJ

- Royal Society Open Science

- Social Psychological Bulletin

- Studia Psychologica

- Swiss Psychology Open

References:

1. Bayer, J. B., Triệu, P., & Ellison, N. B. (2020). Social media elements, ecologies, and effects. Annual review of psychology, 71, 471-497. https:// doi.org/10.1146/annurev-psych-010419-050944

2. Masciantonio, A., Heiser, N., & Cherbonnier, A. (2025). Unveiling the Positivity Bias on Social Media: A Registered Experimental Study On Facebook, Instagram, And X [Stage 2]. Acceptance of Version 2 by Peer Community in Registered Reports. https://osf.io/s93yu

Reviews

Evaluation round #2

Reviewed by Marcel Martončik ⓘ, 24 January 2025

I would like to thank the authors for all the manuscript revisions and their responses to my questions. I am satisfied with all the modifications and believe they have significantly contributed to the clarity of the findings and overall transparency. I particularly appreciate the R output, which elegantly explains the data analysis procedure. I have no further comments on the manuscript.

Equivalence testing Comment 5: Regarding equivalence testing for ANCOVA, I found myself that there isn't a readily available or straightforward method. I appreciate the authors' persistence in addressing this challenging issue. The counterintuitive results (such as non-significant TOST results alongside significant NHST results for main effects and contrasts) likely arise from using the TOSTone function, which is primarily designed for one-sample t-tests rather than ANCOVA with covariates. The TOSTER package offers an equ_anova function that can test at least the main effect of time. If implemented, this might yield more comprehensible result with significant TOST result and non-significant NHST result. However, this approach is still imperfect, as it

excludes covariates whose effects were found to be significant. Regrettably, I have no viable solution also for planned contrasts. Given these limitations, I concur with the authors' recommendation to withdraw these specific analyses. This will naturally impact the way results are discussed, as the absence of an effect cannot be conclusively confirmed.

Evaluation round #1

Authors' reply, 13 January 2025

[Download author's reply](#)
[Download tracked changes file](#)

Decision by Veli-Matti Karhulahti 🆔**, posted 07 November 2024, validated 08 November 2024**

Revision

Dear Alexandra Masciantonio and co-authors, Thank you for your carefully written Stage 2. We were lucky to have all 3 reviewers return to assess the full manuscript. In general, everyone is happy with the outcome, with a few identified issues to be revised before recommendation. Please consider the reviewers' feedback carefully. In addition, I list minor points that I noticed myself. · In Stage 1 content p. 12 it says "does not allow us to answer RQ2" but my impression is that the data answers RQ2 negatively. You can correct this possible error in Stage 1 content.

· On pages 8-9, I would also suggest changing "RQ1" (RQ2, RQ3, RQ4) to "Pilot RQ1" etc to help readers separate pilot and main RQs. If you choose to do this, please make the change coherent by editing the terms elsewhere in the ms too.

· For coherence, you may wish to add a short sentence about Pilot RQ4 at the end of 3.3.4. (the only section in 3.3. that doesn't end with summary sentence).

· Please proofread the text (section 4 onwards) to ensure language is consistent (e.g. "The main research will therefore aims to address..." and especially the use of tense later on).

· On page 15, I suggest adding "RQ" at the end of the question: "how does the positivity bias manifest on social media, and does it vary depending on the type of social media platform? (RQ1)"

· For consistency, end of p. 16, perhaps change RQ to RQ2 and perhaps "We therefore additionally explore our pilot research question"

· There are some parts in the discussion that I cannot fully follow. I list some of them here:

- "*It may only reflect the increased control over self-presentation that social media conveys, compared to real-life interactions.*" –> how would the obtained result of no difference reflect increased difference in social media vs real-life? I cannot follow the reasoning, please clarify or remove.

- "*Therefore, instead of reflecting a cognitive bias towards positivity, the results may be indicative of the inherent design of social media platforms, which naturally lead users to present a more positive version of themselves.*" –> I cannot follow this. If people have, as the paper suggested earlier, positive bias in all social contexts, it's the people who have inherent bias (not the design). People would use any kind of social interaction design in ways that show them in a positive light to others, no? Please clarify or remove.

- Unless I've misunderstood something earlier, it might be good to remove this from conclusions p. 26: "*this may still reflect the inherent control that social media gives users over their self-presentation. Rather than demonstrating a distinct positivity bias, our findings suggest that the architecture and affordances of social media platforms encourage*

behaviors aligned with self-enhancement and positive self-presentation, offering users tools to manage their image effectively." –> I don't see any evidence was obtained that would support claims of encouragement by design. Only the unidirectional design was logically connected to the distinct nature of about X, right? (see below)

- What is said on p. 24 sounds logical: "*Twitter/X's unidirectional mode of connection fosters less intimacy and reci-procity, which could increase the need for positive self-presentation.*" The results clearly support this new hypothesis, and it's an important finding. When it is noted later that "*By focusing on these platform-specific characteristics...*" I would encourage you to more explicitly pinpoint those specific features/mechanisms (if there were others beyond unidirectionality) that you believe were related to the finding. It would also be valuable to explain to readers what are the likely reasons for Avalle et al 2024 not finding such differences (there must be several aspects that explain it in their design vs yours).

- Still on p. 24, it is said "*This finding contrasts with prior research suggesting that users are more likely to post messages containing emoji on social media (Daniel & Camp, 2020), and messages featuring emoji are often perceived as more positive than those without (Novak et al., 2015).*" Please clarify what the cited papers claim (e.g. users are more likely to post emoji on social media – more likely compared to what?)

· One reviewer asks for the code so that we could reproduce the analysis. I can see you have data and code in OSF but there's no mention of it in the paper. Please add the links to the data and code in appropriate places.

· That said, I was not able to use the provided code to reproduce the analysis and my colleague wasn't either. Could you please double check the code and, perhaps with a help of a friend or colleague, ensure that the code can be used by someone external to the study. E.g., as the paper preregistered ICC 0.75 and obtained exactly 0.75, post-publication readers might be interested in re-running the analyses [related: if some form of consensus negotiations were applied to improve rater consistency, please report those procedures, as discussed at Stage 1 – recall transparency steps in data rating].

· Two reviewers ask for a clarification about effect sizes. At Stage 1, we discussed this topic in detail. Adding more of this information to the discussion could clarify things to readers. The fact remains that we don't know what ES are meaningful in practice. For me it was clear that your study was powered for .21 and you considered this ES interesting for that reason. That said, especially the main positive effect about X valence should be reported and its ES discussed.

· Please add conflicts of interest information, and a link to your original registered Stage 1 paper.

· Finally, if there are notable changes that were made to Stage 1 content, please list them in the next response. We don't have a tracked changes file to help us review the changes. E.g., comparing versions, I can notice that this section has been removed: "*We set a predefined threshold for agreement at 0.75, indicating good reliability among raters. If the initial coding round does not meet the ICC threshold, will also conduct additional training sessions for new coders to ensure a clearer understanding of the coding criteria and reduce biases in subsequent coding rounds. The same procedure will be used to assess the valence of the description of the image/video associated with the post.*" This is important information and should be brought back. I hope the reviewers' feedback and my comments are helpful in making the final revisions. If you wisht to discuss anything during the revisions, feel free to contact me directly before submitting. All the best wishes and good luck the last steps! Veli-Matti Karhulahti

Reviewed by anonymous reviewer 1, 03 November 2024

Thank you for providing the opportunity to review the stage 2 version of the manuscript "Unveiling the Positivity Bias on Social Media: A Registered Experimental Study On Facebook, Instagram, And X". This manuscript presents an interesting examination of the positivity bias on social media. The authors have effectively maintained consistency with their approved Stage 1 hypotheses and methodologies, and the results are clearly presented and well-supported by the data.

A few suggestions for improvement include providing more descriptive statistics, such as the frequency of

social media use. Additionally, H2a should be also mentioned in the discussion section.

I also noted a minor issue in Table 1 where the statement "H1 is disconfirmed" may actually refer to H2.

Regarding Figure 4, the color choices could be adjusted for better visibility, as the grey and light blue shades may be hard to distinguish.

Congratulations to the authors on their work!

Reviewed by Marcel Martončik ⬤, 25 October 2024

I would like to thank you for the opportunity to review this manuscript after Stage 2. I commend the authors for their research implementation and results presentation – specifically for strictly adhering to the preregistered plan in both presenting and discussing the results. I appreciate that the interpretations are firmly grounded in the data and avoid speculation. I am confident that this manuscript will make a valuable contribution to knowledge in this field.

I have several questions and comments for the authors:

1) Could you please explain what "Did not participate" means in the enrollment section of Figure 3? Does this refer to participants who were excluded during prescreening?

2) While I'm not entirely certain about this point and offer it merely for consideration, I would suggest removing the phrase "As anticipated" from the exploratory section (p. 21), as it might create a misleading impression that this analysis was preregistered.

3) Regarding the exploratory analyses, I would recommend adding effect size measures (presumably Cramer's V) when presenting the chi-square test results. Similarly, an effect size measure is missing for the following analysis: "a significant main effect was found for the second contrast (Twitter/X vs. other platforms)" - I assume this would be Cohen's d or Hedges' g. Additionally, descriptive statistics should be included here, as well as for the subsequent analysis regarding gender differences: "gender showed a significant negative effect $(F(1, 521) = 14.949, p = .001, \eta2 = .026)$, with women reporting lower valence of the image than men".

4) I realized only now, and regret not noticing during Stage 1, that the Methods section lacks a Statistical analysis subsection. Unfortunately, as it cannot be modified after Stage 1, I would suggest (depending on the Recommender's advice) including information about the software used at least in the Results section. Following this, I would recommend adding the analysis script to enhance reproducibility. Additionally, sharing raw data and a codebook would help provide better insight into what was analyzed and how.

5) For analyses that yielded non-significant results, I would recommend conducting equivalence testing to confirm the absence of effects larger than SESOI. This could be included in the exploratory analyses section.

6) For descriptive purposes, when presenting results for H1 (the non-significant interaction), I suggest including results for main effects (not just contrasts). I appreciate how the non-significant result for H1 is discussed, avoiding speculation beyond existing data. However, precisely because the interpretation is based on the absence of an effect, I would strengthen this claim with equivalence testing.

7) The effect size for the number of followers on Instagram and X is smaller than $r = 0.1$. For completeness, it would be appropriate to include the results for Facebook follower numbers in the Results section (despite being non-significant) to allow comparison of effect sizes. Is this effect size substantial enough to warrant the stated interpretation that it "demonstrates that Instagram and Twitter/X's unidirectional mode of connection fosters less intimacy and reciprocity"?

8) Regarding the interpretation on p. 25, you state that a "possible explanation is that emoji usage may depend on the context: emoji might be more commonly used in private, interpersonal exchanges rather than in public social media posts." This interpretation assumes that participants interpreted the instruction "imagine sharing this event on [Facebook][Instagram][Twitter/X]. Write a post below as you would in real life. Please note

that this must be a post, and not a story" as referring to a public post rather than a post for friends. However, this distinction isn't clearly implied by the instruction.

9) The interpretation that follows in the discussion of the same RQ relies on platform norms, specifically emotional display norms. Could you please elaborate on this? Which norms are you referring to? What do existing empirical findings tell us about these norms?

Reviewed by Julius Klingelhoefer ⓘD, 02 November 2024

Thank you for the opportunity to review this paper in its final form. I have read the preprint with great interest and I believe it is well written and well conducted. I only have some very minor comments, otherwise, I recommend the article for publication.

Overall, I believe this is really well done and I especially enjoyed the discussion of self-presentation, as it was clear and theoretically interesting.

As all comments are minor, I have listed them chronologically:

- p.12: $\eta 2 = 0.0001$ has four digits, usually I would expect only three digits.

- Some latin letters are not italicized even though they should be and some greek letters are italicized even though they should not be, e.g. $N =$ on p. 13

- I don't think the description of SESOI is clear. For the .21 effect size, the authors write "smaller effects could be practically meaningful", so .21 does not appear to be the smalles practically relevant effect size. Please also refer to my previous comments on SESOI, as I believe the power analysis is not based on a SESOI but on an expected effecte size (which is also good but it's not a SESOI). I would suggest clarifying this in 1-2 sentences. This seemst to persist both in the Method and Discussion sections.

- I believe there should be a space before and after minus signs (e.g, $t = - 2$)

I wish the authors the best of luck with the publication and the next steps.