Dear Dr. Karhulahti,

We would like to thank you for giving us the opportunity to submit a revision of our article, titled “Unveiling the Positivity Bias on Social Media: A Registered Experimental Study On Facebook, Instagram, And X.” to *PCI Registered Reports*.

We would like to thank the editor and reviewers for the time they have devoted to our article. Based on their feedback, we have made four main changes:

- We have completely rewritten the theoretical introduction to provide more information on the positivity bias and the differences between social media platforms. In particular, we have differentiated these platforms on the basis of three characteristics: the architecture, the affordances and the socio-cultural context (Masciantonio et al., 2024).
- We offered more context on the expected link between the positivity bias and the use of emojis. We also removed the number of words from the primary variables.
- We changed the protocol of the main study by offering a comparison of the valence of the event told to a group of friends (non-social media condition) with the valence of the event shared on social media. We also added the possibility of describing an image to accompany social media posts to take into account the specificities of image-oriented social media such as Instagram. We believe these changes offer more ecological validity to the research.
- We have further defined the analyses to be performed. In particular, we carried out an a-priori power analysis (Lakens, 2014), based on the Smallest Effect Size of Interest (Lakens, 2022). The questionnaire, analyses, data and power analysis script have been deposited on OSF.

Please find below a point-by-point response to all the comments raised. In addition, we highlighted the changes to our manuscript within the document by using yellow colored text.

Sincerely,

The authors

Table of Contents

Editor..2
Reviewer 1 8
Reviewer 2 15
Reviewer 3 21
Reviewer 4 33

Editor

General comment:

Thank you for submitting your interesting Stage 1 to PCI RR. As the first two reviews were mixed, I wanted to ensure your manuscript receives comprehensive assessment and invited two more reviewers. After carefully considering all four reviews, I agree that major changes need to be implemented for the study to be informative, but I am also optimistic that such changes are possible if the reviewers' feedback is carefully taken into consideration. I summarize the key points.

Response:

We sincerely thank the editor for the opportunity to respond to the comments raised by the reviewers, as well as for the time spent and suggestions provided. We have completely revised the manuscript and believe the quality to be improved.

Comment 1:

Construct/theory

The reviewers collectively voice that there is some confusion in 'positivity bias' as a concept as well as in its related theoretical implications. For instance, there's also a well-known effect of 'negativity bias'. It could be explained in more detail how and why is the former associated with social media (i.e., what's the mechanism/theory).

My impression is that two different positive biases are relevant here. The first one is about the positive/negative ratio of communication, i.e. which type is more prevalent. The second one is about exaggeration, i.e. how much positive become more positive. It feels that your study addresses and is most suitable for the latter construct. Here it's important to clarify: if positivity bias is true, this should mean that positive events become more positive and negative events become less negative. If positive events become more positive and negative events become more negative, then the effect is not a positivity bias but an exaggeration bias (i.e., all content is exaggerated, perhaps to maximize attention).

Response:

We revised the theoretical introduction to provide more context for positivity bias based on *Face Theory* (Goffman, 1959):

> ***"This tendency towards positivity on social media reflects a desire for online positive image and social approval. Indeed, this bias is rooted in the face theory, which postulates that individuals strategically manage their self-presentation to maintain their social identity and uphold their reputation in the eyes of others (Goffman, 1959). This impression management on social media is achieved in a variety of ways, through the selection of topics posted, the audience targeted and the way in which***

> ***information is presented (Marwick & boyd, 2011; Merunková & Šlerka, 2019; Vitak & Kim, 2014)." (p. 3)***

We also defined how social media provide a particular context for positivity bias:

> ***"As a result, the positivity bias on social media not only prompts users to highlight the favorable aspects of their lives but also encourages them to frame both positive and negative facets in a positive light. This phenomenon is propelled by several factors. Firstly, social media platforms afford users a level of control over self-presentation that surpasses real-life interactions (Merunková & Šlerka, 2019). Secondly, these platforms provide features that actively promote positivity, such as filters and emoji. Emoji, in particular, have become integral to self-expression, also contributing to the formation of users' identities (Huang et al., 2022). Users are more likely to post a message on social media when it contains an emoji (Daniel & Camp, 2020), and messages with an emoji are perceived as more positive than those without (Novak et al., 2015). Lastly, when users are posting publicly (e.g., Facebook), rather than privately through messaging application (e.g., Facebook Messenger), the potential audience is significantly larger, amplifying the pressure to maintain a positive image (Spottswood & Hancock, 2016)." (p. 4).***

We believe that the conceptualization of the positivity bias is clearer this way.

Comment 2:

Other theoretically relevant elements, as the reviewers note, are the platforms. Because the goal is to study differences between platforms, it would be valuable to explain how the design and mechanics of platforms differ (I believe this was also a central idea of the cited Meier & Reinecke 2021). For example, closed Facebook groups provide protection and safe spaces, whereas open Twitter debate is more riskly (meanwhile, in both users can customize modes of participation). I hope the above examples help you to further clarify the construct of the study and how it may be more explicitly connected to the hypotheses and theory.

Response:

We have revised the theoretical introduction, distinguishing social media according to three characteristics: architecture, affordances and socio-cultural context:

> ***"Firstly, social media architecture comprises several features (Bossetta, 2018). The connection mode is the most essential to consider when studying positivity bias, as it relates to the type of relationships between users. Facebook has a bidirectional connection mode (e.g., friends) whereas Instagram and X have a unidirectional connection mode (e.g., followers). This implies that Facebook users partly know their friends on the platform, which is not necessarily the case for Instagram and X. Secondly, affordances address not the objective features of platforms, but how users perceive them (boyd, 2010). Two affordances are especially relevant in the context of positivity bias. Shareability relates to the content shared on platforms (Masciantonio et al., 2024): Facebook is perceived as suitable for posting text and image, Instagram***

> ***is mainly associated with image content and Twitter with textual content (Pittman & Reich, 2016). Image-oriented social media are associated to the most stylization from users, and thus impression management (Boczkowski et al., 2018). The visibility affordance can also be at play, focusing on the perception of the degree of visibility of the published content (Treem & Leonardi, 2013). For example, it is lower on Facebook due to its bidirectional nature, and higher on Instagram and Twitter. Finally, the last characteristic is the socio-cultural context (Masciantonio et al., 2024). Users are aware that according to specific social media, certain actions are more accepted by others – the injunctive norms – or more done by others – the descriptive norms (Cialdini et al., 1991; Cialdini & Trost, 1998). These social norms guide user behavior and appropriate contents for each platform (Boczkowski et al., 2018; Tandoc et al., 2019). Waterloo et al. (2018) found that positive emotions were perceived as more appropriate on Instagram and Facebook, while negative emotions were perceived as more appropriate on Twitter and Facebook (Waterloo et al., 2018). These results are in line with sentiment analyses studies showing that Twitter posts are mainly related to negative content (Jiménez-Zafra et al., 2021; Naveed et al., 2011; Thelwall et al., 2011)." (pp. 5-6).***

Comment 3:

Methods/materials

I believe the main problem in the current plan is that there's a discrepancy between the dominant modality of social media (images/videos in Instagram, TikTok etc) and the lexical nature of the study. Another main problem is that of controlling measurement. I believe these issues can be solved by dropping out visual-driven media (Instagram) and adding nonsocial media as controls. As the reviews imply, you could consider e.g., comparing 'personal diary' and 'talking to a friend' (f2f) to text-driven social media. Because both Facebook and Twitter can be used in large/open and small/closed groups, I share you an idea of framing these options—instead of Facebook and Twitter—based on *how* they are used: "Imagine posting this event for a closed group of friends in social media such as Facebook or Twitter", "Imagine posting this event publicly for millions of people to read in social media like Facebook and Twitter". In this way, you would get to study the *mechanisms*. It would also allow discussing (albeit not testing) platform differences, as we know that certain mechanisms are more characteristic to certain platforms.

Response:

Regarding the first point (on image-oriented social media), we have elaborated on the peculiarity of image-oriented social media platforms in the theoretical framework:

> ***"Facebook is perceived as suitable for posting text and image, Instagram is mainly associated with image content and Twitter with textual content (Pittman & Reich, 2016). Image-oriented social media are associated with the most stylization from users, and thus impression management (Boczkowski et al., 2018)" (p. 5).***

Additionally, we have modified the protocol to allow participants to describe an image if they choose to include one in their post:

> ***"To reflect the fact that Instagram is an image-oriented social media, they will also be asked an optional question: 'If you plan to use an image or photo to accompany this post, please describe it briefly here'" (p. 16).***

We consider that it is crucial to retain Instagram as its image-based nature can lead to variations in positivity bias, as mentioned in the manuscript:

> ***"While the positivity bias should appear across all social media platforms, its prevalence and manifestations may vary depending on the platform's unique characteristics. For instance, on image-oriented platforms like Instagram, the positivity bias might be more pronounced. The opposite could be true on platforms like Twitter, known for textual concise messages" (p. 6).***

Regarding the second point (real-life settings), we decided to follow the recommendation of the reviewer. We changed the protocol of the main research to compare the valence of the event told to a group of friends (non-social media condition) with the valence of the event shared on social media.

> ***"On the other hand, comparing the valence of an event with that of its expression on social media may not be the most informative. Indeed, to demonstrate the existence of a positivity bias specific to social media, it is necessary to establish that this bias is not equivalent in face-to-face social contexts (Goffman, 1959). For this reason, one solution would be to ask participants to imagine themselves narrating this event to a group of friends, and then ask them to share it on one of the three social media" (p. 13).***

It would have been interesting to compare private and public posting on social media, but some studies have shown that positivity bias manifests mainly publicly (Spottswood & Hancock, 2016). In addition, we wanted to focus on cross-platform differences. This point will be addressed in the discussion of the study as future endeavors.

Comment 4:

Another important methods issue is the effect size. As reviewers note, to test a hypothesis, it's necessary to justify and state the smallest effect size of interest and also explain what will corroborate the null, e.g. by means of equivalence testing (see Section 2.3. Evidence Thresholds in Guidelines). At PCI RR, Cohen's benchmarks are not used so I suggest carefully thinking what would be a meaningful raw effect. Intuitively, I think you're in a good position because you're using human raters to observe differences in posts: we thus already know that one step up in the scale is noticeable and meaningful. Maybe this can help you define a SESOI. As one relevant source on this, I refer to the paper that is part of the PCI RR guidelines: https://doi.org/10.1525/collabra.28202

With the above considered, I encourage the hypotheses to be redesigned and specified and explicitly formulating the current RQ1/RQ2 section as exploratory analysis because it does not involve confirmatory tests. To make the structure coherent, the first study could be just "Pilot".

Response:

We have renamed the exploratory study to pilot study. We also included the analyses for the research questions in the exploratory analyses. Finally, we have redone the power analysis, the script is deposited on OSF:

> ***"To determine the sample size, we carried out an a-priori power analysis (Lakens, 2014), using the package 'WebPower' (Zhang & Yuan, 2018). We set the alpha level to 0.05, and aimed for a power of 80%. Regarding the effect size, we identified the Smallest Effect Size of Interest (Lakens, 2022). We used a subjective justification based on prior meta-analyses (Lakens et al., 2018). As there is no meta-analysis directly comparing positive self-presentation in person and on different social media, we relied on Ruppel et al. (2017) meta-analysis examining the difference between computer-mediated and face-to-face self-disclosure. Their findings indicated an average meta-analytic effect size of r = .211 (equivalent to f = 0.216). For a repeated ANOVA, power analysis indicated a minimum required sample size of 219 for H1 (within-subject), and of 270 for H2 (interaction). We have rounded the required sample size to 300 participants. The R script for the power analysis is available at this link: https://osf.io/akgdj/?view_only=42142acd518a42cf99b33f5ebec1c780." (p. 15).***

Comment 5:

***

Most reviewers wished to see more materials to be able to better assess the design. I very much enjoyed the clear pilot materials with translations, but it would indeed make reviewing easier if the upcoming materials would be accessible too. This would also allow making direct improvement suggestions that can efficiently support the development of materials. E.g., one thing that the reviewers didn't seem to notice is the final sentence in endnote iii: "Please choose an event that is neither very painful nor very positive." Alas, you still plan the following: "Regarding the event that participants are thinking about, they will be asked to what extent this event is positive or negative (-3 = "Very negative"; 3 = "Very positive")." If they are not allowed to think about very positive events, it seems conflicting to have it as an option (unless it's a control question).

Response:

We have deposited the material for the main study on OSF. We have also changed the presentation of the instruction in the manuscript to make it more transparent:

> ***"The text-writing instruction will be almost identical to that of the pilot study (Talarico et al., 2004). However, to prevent participants from reporting traumatic***

> ***experiences, in agreement with the Ethics Committee of the University of Geneva, a sentence will be added at the end: "Please choose an event that is neither very painful nor very positive"." (p. 16)***

As explained, we added this sentence based on the feedback from the ethics committee. However, we believe that we will still get variations in valence of remembered events.

Reviewer 1

General comment:

Thank you for allowing me the opportunity to review this RR Stage 1. Although I appreciate that some further insight is being studied in relation to types of content on social media and the implications of this to perceptions which may come from this, I felt that the methodological design lacks sufficient rigour to test the desired variables of interest. Largely, I would be cautious about assuming that the scenarios that people are asking to imagine and then write is solely attributed to the social platform they are writing on. I would envisage that positive bias in memory/recalled events more generally might be prominent which is not really factored into the design. Additionally, even if differences are found in valence and other measured variables between social platforms, this doesn't explain what might be causing these differences. The authors have assumptions that there are general differences in sentiment on the different platforms, but these seem quite speculative and generalised, without much scrutiny about what is attributable to these. As such, the current research doesn't really contribute much understanding about this. Finally, I am not convinced that writing retrospectively about autobiographic memories on social media is very ecological valid. This doesn't seem like a behaviour which people would do so much and might be more likely to post about immediate situations or similar rather than retrospective memories. Therefore, regrettably I would not recommend this to be progressed further into the editorial process. I have included some specific observations/comments below which I hope might be helpful for the authors.

Response:

Thank you for taking the time to review our manuscript and for providing valuable feedback. We have carefully considered all of the concerns raised by the reviewer. In particular, we have extensively revised the theoretical framework to emphasize the factors that may contribute to the observed phenomenon, highlighting the significance of architecture, affordances, and socio-cultural context on social media. Furthermore, we have modified the protocol so that social media posts are no longer simply compared to the recollection of an event, but rather to how that event is recounted in real-life settings to friends. Additionally, we have made adjustments to enhance the ecological validity of the protocol, such as incorporating the option to describe an image related to the social media post. All these changes will be explained in more detail in the response letter, as well as for each point raised below by the reviewer. We acknowledge the challenges associated with studying social media, particularly within an experimental framework, but we believe we have struck a balance between external and internal validity. Thank you again for the thoughtful feedback.

Comment 1:

#1. Self-expression vs online disclosure – these terms are used interchangeably (both by the authors and often within the literature) but it would be helpful to offer some specificity on what each of these refers to (and how they are different) to help the reader clarify the authors' conceptualisation here

<u>Response:</u>

We have refined the theoretical framework to focus on the phenomena of positivity bias and positive self-presentation. As a result, we have simply removed references to self-expression, which is a broad term encompassing all mentioned phenomena, and online disclosure, which is a narrower term indicating individuals sharing personal information on social media. The theoretical framework now conceptualizes positive self-presentation based on Goffman's theory (1959).

<u>Comment 2:</u>

#2. When studying these issues, one major limitation of the proposed research is that the method doesn't incorporate any measure of non-text based expressions (beyond emoticons/emoji). When considering platforms such as Instagram, I would argue that expressions on where are primarily in relation to image or video sharing rather than text (text tends to be the secondary expression). As such, I am not convinced that just measuring text is entirely helpful to understand expressions as they might naturally occur on these different platforms

<u>Response:</u>

We appreciate the reviewer's comment. Firstly, we have elaborated on the peculiarity of image-oriented social media platforms in the theoretical framework:

> ***"Facebook is perceived as suitable for posting text and image, Instagram is mainly associated with image content and Twitter with textual content (Pittman & Reich, 2016). Image-oriented social media are associated with the most stylization from users, and thus impression management (Boczkowski et al., 2018)" (p. 5).***

Additionally, we have modified the protocol to allow participants to describe an image if they choose to include one in their post:

> ***"To reflect the fact that Instagram is an image-oriented social media, they will also be asked an optional question: 'If you plan to use an image or photo to accompany this post, please describe it briefly here'" (p. 16).***

We consider that it is crucial to retain Instagram as its image-based nature can lead to variations in positivity bias, as mentioned in the manuscript:

> ***"While the positivity bias should appear across all social media platforms, its prevalence and manifestations may vary depending on the platform's unique characteristics. For instance, on image-oriented platforms like Instagram, the positivity bias might be more pronounced. The opposite could be true on platforms like Twitter, known for textual concise messages" (p. 6).***

We therefore believe that this approach is the most appropriate given the complexity of examining social media (Griffioen et al., 2020), however we recognize the particular challenge of comparing different platforms, and will discuss this further in the limitations of the research.

Griffioen, N., Rooij, M. van, Lichtwarck-Aschoff, A., & Granic, I. (2020). Toward Improved Methods in Social Media Research. *Technology, Mind, and Behavior, 1*(1). https://doi.org/10.1037/tmb0000005

Comment 3:

#3. Some of the detail about Self disclosure/self expression in the introduction/literature review could benefit from acknowledging that expressions might vary based on where on each of these platforms things are being shared (on private groups, profile pages etc). There is not much scrutiny about this but in naturally-occurring behaviour, this would be a critical factor determining the nature of people expressions/self disclosures

Response:

We agree with the reviewer, and the theoretical introduction has been completely revised. Here is an example:

> ***"This phenomenon is propelled by several factors. Firstly, social media platforms afford users a level of control over self-presentation that surpasses real-life interactions (Merunková & Šlerka, 2019). Secondly, these platforms provide features that actively promote positivity, such as filters and emoji. […] Lastly, when users are posting publicly (e.g., Facebook), rather than privately through messaging application (e.g., Facebook Messenger), the potential audience is significantly larger, amplifying the pressure to maintain a positive image (Spottswood & Hancock, 2016)." (p. 4).***

Comment 4:

#4. The introduction/literature review starts to conflate two issues a bit. The current research is focused on how positivity bias on senders' use/expressions but this starts to get conflated with discussion about the effects of positively biased content on social comparisons/well-being. This is a somewhat separate issue and not something the current research is testing

Response:

We have removed the literature on the effects of positivity bias from the theoretical introduction, as it is indeed not being tested in this research. Instead, we have expanded on the conceptualization and manifestation of positivity bias on social media.

Comment 5:

#5. P4-5 provides some detail about the directionality of relationships/interactions on the various social platforms but these seem rather generalised. I would argue that this detail should be toned down as relationships on X or IG for example are not always unidirectional and those on FB are not always dyadic as suggested.

Response:

We have provided more details on the architecture of social media platforms, particularly focusing on the directionality of connections:

> ***"The connection mode is the most essential to consider when studying positivity bias, as it relates to the type of relationships between users. Facebook has a bidirectional connection mode (e.g., friends) whereas Instagram and X have a unidirectional connection mode (e.g., followers). This implies that Facebook users partly know their friends on the platform, which is not necessarily the case for Instagram and X." (p. 5).***

Comment 6:

#6. P5- "From a practical point of view, knowing if certain social media favor a negative information presentation has the potential to inform public health recommendations. One example is the phenomenon of cyberbullying, which has become increasingly resource-intensive in recent years (Gumbus & Meglich, 2013)2- I am not fully clear on how the proposed findings will relate directly to public health policy. That is, the current study is focusing on comparing the nature of people's expressions on various social sites, but isn't specifically providing insight which helps understand what is acceptable or how it is interpreted which is perhaps more relevant to examples such as cyberbullying.

Response:

We agree with the reviewer, and therefore, we have removed this reference from the manuscript.

Comment 7:

#7. The issue about emoji is noted in passing at the end of section 1 but there is not much rationale or explanation about the relevance of emoji in respect of positively bias and how this might relate to these social platforms. That is, to what extent can emoji (which might vary in valence and expression of different types of emotions) relate to people's ability to express positively/negatively on these platforms?

Response:

We have completely revised the discussion of emojis in the manuscript:

> ***“Emoji, in particular, have become integral to self-expression, also contributing to the formation of users' identities (Huang et al., 2022). Users are more likely to post a message on social media when it contains an emoji (Daniel & Camp, 2020), and messages with an emoji are perceived as more positive than those without (Novak et al., 2015).” (p. 2)***

We have also changed the analyses and research questions regarding emojis to align with this direction (see section 3.1 Research Questions, and 3.3.3 The Use of Emoji and the Positivity Bias).

Comment 8:

#8. Number of words is a key variable but a confound here is that some platforms (Twitter/X specifically) have character restrictions which presumably will be a factor which affects people’s abilities to self-express (which might not have anything to do with positivity bias)

Response:

The experimental study revealed that users use a fairly similar number of words on social media platforms. However, we agree with the reviewer, and therefore, decided that the number of words should not be one of the main variables. We have modified the manuscript accordingly.

Comment 9:

#9. It is not clear why emoticons are only measured at time 2.

Response:

The reason is that participants at time 1 did not use emoticons, as their usage is more commonly associated with messaging discussions or social media interactions (Huang et al., 2022).

Huang, V., Hu, Y., & Li, Y. (2022). A Systematic Literature Review of New Trends in Self-expression Caused by Emojis and Memes. *2021 International Conference on Social Development and Media Communication*, 75–79. https://doi.org/10.2991/assehr.k.220105.016

Comment 10:

#10. How were emoticon rated in terms of valence? Were these rated independently from the text or in conjunction with? Because emoticons have been found to elicit somewhat diverse interpretations, it is important that detail about the process and the specific inter-agreements here is provided.

Response:

The valence of texts on social media was based on both words and emoticons together. We made this choice because the interpretation of emojis depends on the context of the post: for example, some users used the "😂" emoji in a sad context, while others used it in moments of exaggerated joy.

Comment 11:

#11. Emotional intelligence is measured and applied in respect of the three experimental conditions. My reading of this makes me assume that the scale was used as an informant rating based on the text presented. This seems somewhat problematic to me as this EI scale is a way of helping people themselves rate their own abilities based on their general abilities in situations. The way this has been applied in the current research is an informant rating based on what text is available from which to judge this ability. I don't feel this is an appropriate use of this scale as I don't believe it corresponds validly to the construct it was originally designed to measure

Response:

The scale was used as a covariate to account for individual differences that could explain variations in the manifestation of positivity bias on social media. We provided further explanation in the manuscript:

> ***"Since social media use is highly dependent on individual characteristics (Valkenburg & Peter, 2013), we measured emotional intelligence in an exploratory way (WEIS, Wong et al., 2007). Our assumption was that positivity bias might depend on how users perceive their own emotions and those of others." (p. 9).***

Given the results of the pilot study, which show no effect of emotional intelligence on our dependent variables, it will not be reused in the main study.

> ***"Second, the choice of the variables measured can also be improved. We found no effect of emotional intelligence in any of the analyses. […] The literature highlighted at least two key variables to consider, the number of relations on each social media, and to what extent users know about these relations in real life (H. Lin et al., 2014). These variables could provide further insight into platforms architecture and affordances (Masciantonio et al., 2024)." (p. 13).***

Comment 12:

#12. The inclusion of emoticon/emoji appears to relate to exploring how their use works in ratio to number of words. In relation to these symbols, there might be something more insightful to explore here about what type of emoji they are and their valence rating. In terms of focus on understanding the role of positivity bias, focusing on emoji only seems to be conceptually useful if you know whether the emoji used are indeed valent-relevant to this rather than if they are just used per se.

Response:

As previously explained, we have provided a better explanation of the relationship between emoji use and positivity bias, for example:

> ***"Users are more likely to post a message on social media when it contains an emoji (Daniel & Camp, 2020), and messages with an emoji are perceived as more positive than those without (Novak et al., 2015)." (p. 2).***

Therefore, we hypothesized that the number of emojis used depended on the valence of the event at time 1. We changed the pilot study analyses accordingly:

> ***"We then tested the association between the valence of the text at time 1 and the ratio number of emoji per word. We found a positive association, meaning that the more the text valence at time 1 was positive, the more participants used emoji to write a text on social media at time 2; r(277) = 0.13, p = 0.03." (p. 11).***

We believe these changes offer a better examination of the role of emojis in positivity bias on social media.

Reviewer 2

General comment:

Thank you for the opportunity to be a reviewer for the registered report " Unveiling the Positivity Bias on Social Media: A Registered Experimental Study On Facebook, Instagram, And X". I believe the research explores a very interesting and important topic, which is how the positivity bias differs across different social media platforms. Although I am not an expert in the field of social media research, I believe that the goal and need for the study are clearly explained and that the proposed methods seem appropriate. Below are some comments and suggestions to further improve the proposed research.

Response:

We appreciate the reviewer's time and constructive feedback. We have addressed all of his/her comments below.

Comment 1:

Specific comments

Major comments

- The proposed research is both interesting and relevant, and the study is well explained. However, it would enhance clarity to explicitly state that H1 replicates an established phenomenon (the positivity bias), while H2 introduces a novel perspective by examining this bias across various social media platforms. Although this distinction is mentioned in the general introduction, it could be discussed more explicitly in section 4.1.

Response:

We agree with the reviewer's suggestion, and we have completely revised the manuscript accordingly. Firstly, we have restructured the theoretical introduction, discussing the positivity bias (see section 1.1 The Positivity Bias on Social Media) followed by addressing the limitations of the literature regarding the examination of different platforms (see section 1.2 The Positivity Bias on Various Social Media). Additionally, we have modified the statistical analyses of the pilot study (see section 3.3 Results), as well as the hypotheses and research questions of the main study (see section 4.1 Hypotheses and Research Question).

Comment 2:

- While the different hypotheses are well explained, it would be beneficial to also formulate a specific research question for the confirmatory aspect of the study, not solely for the exploratory part.

Response:

We have added a specific research question before the hypotheses for the main study:

> ***"The main research will therefore aim to address the following fundamental question: how does the positivity bias manifest on social media, and does it vary depending on the type of social media platform?" (p. 14).***

Comment 3:

- Adding additional background information on the relevance and significance of emoticon usage and post length in social media would enrich the relevance of the exploratory aspect of the study.

Response:

We agree and have provided more detailed background information on the relevance of emoticon usage throughout the manuscript, particularly in the theoretical introduction:

> ***"Secondly, these platforms provide features that actively promote positivity, such as filters and emoji. Emoji, in particular, have become integral to self-expression, also contributing to the formation of users' identities (Huang et al., 2022). Users are more likely to post a message on social media when it contains an emoji (Daniel & Camp, 2020), and messages with an emoji are perceived as more positive than those without (Novak et al., 2015)." (p. 2)***

Regarding the word count, in agreement with reviewer 1, we have decided to focus solely on valence and emojis, particularly because the word count is influenced by the platform's architecture (for example, the character limit on X).

Comment 4:

- Clarification is needed regarding the statement "whose text will not be coded by the researcher". Why are some texts not coded by the researcher?

Response:

We thank the reviewer for his/her comment. The texts that were not coded by the reviewers are simply those for which participants did not follow the instructions correctly. For example, some participants wrote responses such as "I will not talk about this on Facebook" or "I don't know how this works" instead of writing a post for one of the social media platforms. This is why the coders did not code these texts. We should have been clearer in the manuscript, and we have modified the sentence accordingly:

> ***"We also removed participants who did not understand the experimental instructions (n = 22)." (p. 7).***

Comment 5:

- The authors do a sample size calculation based on expected effect sizes; however, they do not mention calculating effect sizes for the planned analyses. Incorporating effect size calculations alongside the planned statistical analyses would be beneficial.

Response:

We have recalculated the power analyses, by identifying the Smallest Effect Size of Interest (Lakens, 2022):

> ***"To determine the sample size, we carried out an a-priori power analysis (Lakens, 2014), using the package 'WebPower' (Zhang & Yuan, 2018). We set the alpha level to 0.05, and aimed for a power of 80%. Regarding the effect size, we identified the Smallest Effect Size of Interest (Lakens, 2022). We used a subjective justification based on prior meta-analyses (Lakens et al., 2018). As there is no meta-analysis directly comparing positive self-presentation in person and on different social media, we relied on Ruppel et al. (2017) meta-analysis examining the difference between computer-mediated and face-to-face self-disclosure. Their findings indicated an average meta-analytic effect size of r = .211 (equivalent to f = 0.216). For a repeated ANOVA, power analysis indicated a minimum required sample size of 219 for H1 (within-subject), and of 270 for H2 (interaction). We have rounded the required sample size to 300 participants. The R script for the power analysis is available at this link: https://osf.io/akgdj/?view_only=42142acd518a42cf99b33f5ebec1c780." (p. 15).***

Comment 6:

- Similarly, while the authors adeptly outline their interpretation of significant results, it would be beneficial to explicitly address how this interpretation may be contingent upon the effect size.

Response:

We agree with the reviewer. We will discuss the interpretation of the results in the research discussion. We will notably emphasize that a significant effect is not necessarily meaningful (Lakens et al., 2018). However, we will also discuss mechanisms that can amplify an effect's importance on social media, such as repetition, or counteract it, such as habituation (Anvari et al., 2023).

Anvari, F., Kievit, R., Lakens, D., Pennington, C. R., Przybylski, A. K., Tiokhin, L., ... & Orben, A. (2023). Not all effects are indispensable: Psychological science requires verifiable lines of reasoning for whether an effect matters. *Perspectives on Psychological Science, 18*(2), 503-507.

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence Testing for Psychological Research: A Tutorial. ***Advances in Methods and Practices in Psychological Science, 1*****(2), 259–269. https://doi.org/10.1177/2515245918770963**

Comment 7:

- Clarification for the criteria for data exclusion is needed. Will neutral data be excluded again? Are participants removed if valence cannot be accurately coded? Additionally, are there any other criteria for outlier exclusions, such as excessively short or long posts?

Response:

We have modified the manuscript to provide more details. Specifically, we have decided not to exclude neutral data, as it is an integral part of self-presentation on social media. As mentioned earlier, we have also provided more indications regarding the responses removed. We have not added any other exclusion criteria.

Comment 8:
Minor comments

- The introduction and/or abstract would benefit from an earlier definition of "positive bias," (this sentence may be written earlier "leading users to predominantly share and engage with positive content rather than negative or neutral ones")

Response:

We have defined the positivity bias at the beginning of the theoretical introduction:

> ***"The positivity bias on social media reflects users' tendency to present favorable aspects of themselves rather than negative ones (Schreurs & Vandenbosch, 2021), aligning with the concept of positive self-presentation (Utz, 2011)." (p. 3).***

Comment 9:

- Clarify whether the question regarding participants' usage of Facebook, Instagram, and Twitter at least once a month relates solely to interacting with the platform or includes posting content as well.

Response:

This question originates directly from Prolific, the recruitment platform: "Which of the following social media sites do you use on a regular basis (at least once a month)?". It serves as a filter to ensure that participants are active users of the specified social media platforms, both in terms of interaction and content contribution. However, we do not know what they do most on the different platforms, which is why we added a question to

the main study asking participants whether they would post the publication they've just written on social media.

Comment 10:

- Instagram also requires to add a picture alongside the text (which the other platforms do not). This should be discussed somewhere in the paper.

Response:

We have elaborated on the peculiarity of image-oriented social media platforms in the theoretical framework:

> ***"Facebook is perceived as suitable for posting text and image, Instagram is mainly associated with image content and Twitter with textual content (Pittman & Reich, 2016). Image-oriented social media are associated with the most stylization from users, and thus impression management (Boczkowski et al., 2018)" (p. 5).***

Additionally, we have modified the protocol to allow participants to describe an image if they choose to include one in their post:

> ***"To reflect the fact that Instagram is an image-oriented social media, they will also be asked an optional question: 'If you plan to use an image or photo to accompany this post, please describe it briefly here'" (p. 16).***

We will also discuss this further in the limitations of the research.

Comment 11:

- Including the full questionnaire in the supplement would aid in study replication. Clarify the types of socio-demographic questions asked, such as defining "current situation" mentioned in the exploratory research.

Response:

All the questions from the pilot study and the main study have been deposited on OSF:

> ***"The coding manual, the data and the analyses for the pilot study can be accessed at this link: https://osf.io/akgdj/?view_only=42142acd518a42cf99b33f5ebec1c780. The coding manual for the main study can be accessed at the same link." (p. 6)***

Comment 12:

- Specify what "enough participants" means for the planned sensitivity analyses.

Response:

We have withdrawn this comment since sensitivity analyses will be performed regardless of the number of participants.

Comment 13:

- The authors mentioned having three conditions when doing the power analysis (section 4.2) but did not clarify what these conditions are. While it was understood as the type of social media in the exploratory study, it would be helpful to explain what they are in the registered study.

Response:

We have modified section 4.2 Method by giving more indication on the conditions, we have also specified the contrasts for social media in line with our hypotheses (Table 2).

Comment 14:

- Since the questionnaires will only be answered on smartphones, it would be interesting to include a question about where people typically use each social media platform (could be added as a control variable). I could imagine that especially older adults use Facebook and Twitter also on their computer.

Response:

We have added a question:

> ***"We will ask participants on which devices they most often use social media (computer, tablet or smartphone)." (p. 16).***

Comment 15:

- Another suggestion is to include a question where participants rate the valence of their social media posts themselves and compare these ratings with those from the authors. It would be interesting to know whether those differ and whether conclusions would change. A control question regarding whether people usually share events on several platforms could also be added.

Response:

We thank the reviewer for him/her suggestions. While we appreciate the idea of comparing participants' self-rated valence of social media posts with ratings from the authors, we have decided not to include such a question in this study. Our focus is primarily on objectively coding the valence of social media posts to maintain consistency and reliability in our data analysis. This is also the reason why we changed the protocol to no longer include a subjective assessment by participants of the valence of the event.

Reviewer 3

General comment:

Dear recommender and dear authors,

I appreciate the opportunity to review this manuscript and contribute my comments to this research. I fully agree with the authors that social media has a profound effect on people’s experiences and behaviour. This underlines the importance of exploring this topic. Please find below my comments in the order in which the commented parts of the manuscript appear.

Response:

We thank the reviewer for his time and contribution. We have addressed his comments below.

Comment 1:

Positivity bias is a central construct of the research, yet its definition seems to be missing in the text. I found myself intrigued by how the authors might have defined it. Given its significance, it might be beneficial to introduce its definition early in the Introduction, clarifying what the authors mean by this term. It could be insightful to devote more discussion to this construct, such as elaborating on its various influences or the potential mechanisms of its effect on social media, or even drawing from related fields for its effect mechanism, if applicable. On page 4, the authors state: ‘These findings underline the pivotal role of the positivity bias in understanding the effects of social media on mental health.’ However, the text doesn’t seem to elaborate on what this role entails. Maybe the meta-analysis by Chu et al. 2023 or other studies might provide valuable context in this regard.

Response:

We have defined the positivity bias at the beginning of the theoretical introduction:

> ***“The positivity bias on social media reflects users' tendency to present favorable aspects of themselves rather than negative ones (Schreurs & Vandenbosch, 2021), aligning with the concept of positive self-presentation (Utz, 2011).” (p. 3).***

We have removed the passages on mental health, in agreement with reviewer 1, as this was not the object of the research.

Comment 2:

After reading the first paragraph of the Introduction, I felt especially after the sentence: "This is precisely the case for the positivity bias, which can provide insights into how social media platforms shape our perceptions, emotions, and overall mental health" that the aim of the paper would be to explore the effects of the social media content on users' thinking and experiencing.

But the following section of the text, beginning with the heading Self-expression on Social Media, did not correspond with this. I see 2 distinct effects: 1) the effect of social media content on thinking and experiencing, and 2) self-expression in the space of social media, which may not be related to already existing content on social media at all (the authors also plan to recruit people who use social media infrequently = at least one a month into the sample). That is, I had a feeling that this study is more of an investigation of the effect of enduring characteristics of individuals on the way they self-express in a social media environment. Which of these do the authors intend to investigate? Apologies for my misunderstanding.

Response:

As mentioned earlier, the aim of the paper is not to investigate the effects of positivity bias on mental health but rather to understand how positivity bias manifests across different social media platforms. We have completely revised the theoretical introduction to make this clearer.

Comment 3:

In this context, I found the phrasing of RQ1 (How does the positivity bias affect self-expression on social media?) somewhat unclear, as I didn't find this effect implied or explained in the Introduction. Could you please provide some clarification?

Response:

We have modified RQ1 to make it clearer and more in line with the theoretical framework:

> ***"RQ1: How does the positivity bias manifest on social media?" (p. 7).***

Comment 4:

One of the key premises of the research is that social media platforms differ from each other in various ways, such as the purpose they fulfill, the needs they satisfy, their features, etc. These differences could potentially cause variations in self-expression on these platforms. Therefore, it might be beneficial to describe how these three media differ. The authors state only very generally: 'They also differ in terms of accepted media (e.g., images, text, hyperlinks, etc.) and privacy settings" or "Second, social media platforms offer different features (Bossetta, 2018). " without specifying the differences. It might be helpful to include a table or figure that describes all the differences and features. Without a detailed understanding of the differences among the three social media platforms, it becomes challenging to discuss their potential differential effect and the origins of this effect. I was particularly intrigued by the statement, "Users' relationships are dyadic on Facebook, but unidirectional on Instagram and X." Isn't mutual following and conversation between users on Instagram a form of a dyadic relationship? The same question applies to X.

Response:

We have developed the differences between social media in more detail in the introduction, differentiating them in terms of architecture, affordances and socio-cultural context:

> ***"Although Facebook, Instagram, and X are all classified as social media (Ellison & boyd, 2013), they differ in several aspects that significantly influence users interactions on the platform. The cross-platform approach suggests that social media can be differentiated according to three characteristics: architecture, affordances and social-cultural context (Masciantonio et al., 2024)." (p. 5).***

We described each of these components in relation to the positivity bias. For example, for the passage on social media connections:

> ***"Facebook has a bidirectional connection mode (e.g., friends) whereas Instagram and X have a unidirectional connection mode (e.g., followers). This implies that Facebook users partly know their friends on the platform, which is not necessarily the case for Instagram and X." (p. 5)***

<u>Comment 5:</u>

This sentence doesnt make sense to me: "For example, positive emotions are perceived as more appropriate on Instagram and Facebook, while negative emotions are perceived as more appropriate on Twitter and Facebook (Waterloo et al., 2018)." Is Facebook there twice by mistake?

<u>Response:</u>

Waterloo et al (2018) measured positive and negative emotions differently, and noted differences between platforms for each. But when they did comparative analyses, they found no differences between Instagram and Facebook for positive emotions, nor between Twitter and Facebook for negative emotions, hence the formulation.

<u>Comment 6:</u>

Could the authors kindly clarify the recommendations or insights intended to convey when stating on page 5?: "From a practical point of view, knowing if certain social media favor a negative information presentation has the potential to inform public health recommendations."

<u>Response:</u>

We have removed this statement from the manuscript, in agreement with reviewer 1.

<u>Comment 7:</u>

Exploratory study

It would be greatly appreciated if the authors could elucidate the purpose of conducting the exploratory research (such as testing of instruments, procedures, estimation of effect sizes, etc.).

Response:

We have given more information on the reasons for carrying out the pilot study in the manuscript, for example:

> ***"As no research has directly tested the positivity bias on various social media, we conducted a pilot study to test an original protocol" (p. 7)***
>
> ***"The pilot study provides new empirical insights for the main research." (p. 12)***
>
> ***"The pilot study also provides additional methodological perspectives for the main research." (p. 12).***

Comment 8:

The justification for a sample size of N = 50 appears to be absent. Could the authors provide a rationale for this choice? For instance, accuracy, a-priori power analysis, heuristics, etc..

Response:

Since it is a pilot study, we did not conduct a power analysis. Instead, we aimed to have at least 50 participants per condition, following the recommendation by Simmons et al. (2013).

Comment 9:

It has been noted that 136 participants were excluded from the sample because they "did not give their informed consent or did not fully complete the study". I am curious as to whether it is ethically appropriate to exclude participants who gave consent and have provided at least partial responses (e.g. due to the browser or OS crash, etc.). Could the authors explain why they did not opt for missing data imputation?

Furthermore, a justification seems to be needed for the removal of 22 participants for whom the valence of their text was not coded by the three coders. Could the authors provide some insight into this decision?

Response:

We thank the reviewer for his inquiry. In the pilot study, most of the participants who did not finish the study did not submit any text responses for social media posts, meaning that it was impossible to impute data. Additionally, as this study is a pilot investigation, the focus is on refining the methodology and ensuring data quality for subsequent phases rather than imputing missing data.

In addition, the texts that were not coded by the reviewers are simply those for which participants did not follow the instructions correctly. For example, some participants wrote responses such as "I will not talk about this on Facebook" or "I don't know how this works" instead of writing a post for one of the social media platforms. This is why the coders did not code these texts. We should have been clearer in the manuscript, and we have modified the sentence accordingly:

> ***"We also removed participants who did not understand the experimental instructions (n = 22)." (p. 7).***

Comment 10:

For a more comprehensive understanding of the research methodology, it would be beneficial to have access to the precise instructions for the scales used. In line with this, it would be helpful if the authors could share the survey, complete with questions and instructions. However, if there are copyright concerns, perhaps the completed questions could be removed. Alternatively, the authors could consider sharing at least the specific instructions they developed for the study.

Response:

We deposited on OSF all questions for the pilot study and the main research:

> ***"The coding manual, the data and the analyses for the pilot study can be accessed at this link: https://osf.io/akgdj/?view_only=42142acd518a42cf99b33f5ebec1c780. The coding manual for the main study can be accessed at the same link." (p. 6)***

Comment 11:

I noticed that one of the constructs measured was the number of words. However, it was not immediately clear from the manuscript why this was measured and how it relates to the research question. Could the authors provide some clarification on this?

Response:

We have indeed removed the number of words as a dependent variable in our study. Upon reflection, we realized that there is not a clear theoretical basis for how positivity bias may impact the number of words used in social media posts.

Comment 12:

The valence of texts is identified as one of the main constructs. However, the methodology for evaluating the texts is not clearly outlined. The manuscript mentions that "three researchers qualitatively analyzed all the texts to estimate their valence on a 7-point scale (-3 = 'Very negative'; 3 = 'Very positive')". Could the authors elaborate on the instructions given to these researchers? On what grounds were they supposed to evaluate the valence of the text? What exactly was this valence intended to express? A detailed explanation of the instructions given to the three researchers would be very helpful in this context. I also wondered if the authors had considered using sentiment analysis. It appears to me that it could be well-suited to the task at hand but I admit I have no idea how difficult it is to use.

Response:

Regarding the evaluation of text valence, we provided the detailed instructions given to the three researchers in the coding manual deposited on OSF: "The coders first agreed on a definition of valence as: the positive or negative character of emotions (Brosch & Moors, 2009, p. 401). For example, positive content refers to happiness, satisfaction, calm, pride or serenity. Conversely, negative content refers to anger, depression, distress, sadness or lassitude. Some content can be neutral, i.e. neither positive nor negative.”.

Additionally, we considered using sentiment analysis tools; however, many of them only provide a dichotomous classification of sentiment (positive/negative), whereas we wanted a more nuanced measure for our analyses. Furthermore, finding sentiment analysis tools that support multiple languages, including French, was challenging.

Comment 13:

Without access to the survey, it is challenging to understand how and by what means some constructs were measured. I was particularly interested in examining the wording of the items used to measure descriptive norms, especially considering the ω value for Instagram was .62. However, I was unable to find information on either the number of items or their wording. Could the authors shed some light on this?

Response:

All survey questions have been deposited online on OSF, as mentioned earlier. Additionally, we have included a sample item for each construct in the manuscript:

> ***"Injunctive norms were measured for each platform with three items; for example “The people who influence my behavior expect me to post content on [Facebook][Instagram][Twitter] mainly…” (1 = very negative; 7 = very positive). Descriptive norms were also measured for each platform with three items; for example, “The people who influence my behavior post content on [Facebook][Instagram][Instagram] mainly…” (1 = very negative; 7 = very positive)." (p. 9)***

Comment 14:

Exploratory results. Effect of Social Media on Texts’ Valence

I noticed that the authors provided respondents with the option to choose the social media platform on which they would like to share or write a given text. This differs from a procedure where respondents would write a text for each social media platform and their valence would be compared. Consequently, it's conceivable that the observed differences in valence between social media may not be attributable to the platform itself, but rather to certain characteristics of the respondents that influence their preference for a particular social media platform.

Response:

Thank you for your observation. We have addressed this concern by including covariates in our analyses. Specifically, we measure the frequency of platform usage, which helps account for individuals' preferences for a particular platform.

Comment 15:

I'd like to kindly ask for clarification on the justifications for the covariates, as I was unable to locate this information.

Response:

We have provided more detailed information on the rationale for selecting covariates throughout the manuscript, both for the pilot study and the main study. For example:

> ***"Since social media use is highly dependent on individual characteristics (Valkenburg & Peter, 2013), we measured emotional intelligence in an exploratory way (WEIS, Wong et al., 2007). Our assumption was that positivity bias might depend on how users perceive their own emotions and those of others." (p. 9)***
>
> ***"The literature highlighted at least two key variables to consider, the number of relations on each social media, and to what extent users know about these relations in real life (H. Lin et al., 2014). These variables could provide further insight into platforms architecture and affordances (Masciantonio et al., 2024)." (p. 13)***

Comment 16:

In the results section, along with stating the main outcome for the interaction, it might be beneficial to provide a detailed explanation of what the interaction implies, including differences between groups and effect sizes, among other things.

Response:

We have provided more details on the results of these analyses:

> ***"The valence of texts at time 1 (M = 0.46, SD = 1.58) was less positive than the valence of texts at time 2 (M = 0.82, SD = 1.39), with valence highest for Instagram (M = 1.08, SD = 1.37), followed by Twitter (M = 0.72, SD = 1.39) and Facebook (M = 0.62, SD = 1.44)." (p. 10)***

Comment 17:

Regarding the “Effect of Event’s Valence on the Choice of Social Media”, it would be helpful if the process of creating the dichotomous variable “valence” could be explained, especially considering that three researchers qualitatively analyzed all the texts to estimate their valence on a 7-point scale, with -3 representing “Very negative” and 3 representing “Very positive”. Also a justification for dichotomization is missing.

Response:

We have revised the analyses, including the neutral responses. Regarding the conversion of the continuous variable into a categorical one, we opted for this approach because the other variable (Facebook, Instagram, Twitter) is also categorical, which is more suitable for the statistical test. It is worth noting that this is an exploratory analysis rather than a primary one:

> ***“First, we have created a new variable depending on whether the event at time 1 was positive, negative or neutral. We then performed a chi-square to test the association between the text’s valence at time 1 (positive vs. negative vs. neutral) and the question where participants could choose which of the three social media was most appropriate to share this event (Facebook vs. Instagram vs. Twitter). In this way, we were able to determine whether, depending on the valence of an event, users will turn to one social media platform rather than another to express themselves.” (p. 11)***

Comment 18:

It would be appreciated if the authors could provide a clear explanation of how the “Effect of Social Media on Emoticons/number of words” section of the results is connected to the research questions of this study.

Response:

We have completely revised the research questions of the pilot study (see section 3.1 Research Questions), as well as the analyses (see section 3.3 Results). In doing so, we have better defined the importance of emoticons in the positivity bias, and we have removed the number of words as a variable of interest.

Comment 19:

I’m curious as to why a result with a p-value of .051 is interpreted as statistically significant.

Response:

Thank you for your observation. We have revised the interpretation of the analyses to reflect that a p-value of .051 is not statistically significant.

Comment 20:

In the discussion of the exploratory section, the authors suggest, “It would therefore be intriguing to propose a design where participants are not required to write the event beforehand.” If the authors were to implement this design, I’m interested in understanding how they would ascertain whether the text produced for social media is positively biased compared to text that was not intended for social media. Could you please elaborate on this?

Response:

We have revised the proposed protocol for the main study. Participants will be required to write a text, which we believe will provide more valid results.

Comment 21:

Confirmatory - RR part

It would be beneficial if the hypotheses could include precise estimates of effect sizes or their intervals. Without these, the hypotheses might be unfalsifiable in their current formulation. The chosen effect sizes or their intervals should be justified based on more than just convention, such as Cohen’s guidelines.

Response:

We have recalculated the power analyses, by defining the Smallest Effect Size of Interest (Lakens, 2022):

> ***“To determine the sample size, we carried out an a-priori power analysis (Lakens, 2014), using the package ‘WebPower’ (Zhang & Yuan, 2018). We set the alpha level to 0.05, and aimed for a power of 80%. Regarding the effect size, we identified the Smallest Effect Size of Interest (Lakens, 2022). We used a subjective justification based on prior meta-analyses (Lakens et al., 2018). As there is no meta-analysis directly comparing positive self-presentation in person and on different social media, we relied on Ruppel et al. (2017) meta-analysis examining the difference between computer-mediated and face-to-face self-disclosure. Their findings indicated an average meta-analytic effect size of r = .211 (equivalent to f = 0.216). For a repeated ANOVA, power analysis indicated a minimum required sample size of 219 for H1 (within-subject), and of 270 for H2 (interaction). We have rounded the required sample size to 300 participants.” (p. 15).***

Comment 22:

I’m also curious about whether the positivity bias is specifically generated by social media, or if it would also be present when writing text for another type of medium or purpose, such as a print newspaper, a blog, or a diary. I wonder if the positivity bias results from the need to abbreviate text, among other factors. For this reason, I would recommend considering the use of a control group in this research.

Response:

We have modified the protocol to compare posting an event on social media with recounting the same event to a group of friends. This adjustment will provide more insights into whether the positivity bias is unique to social media platforms:

> ***"On the other hand, comparing the valence of an event with that of its expression on social media may not be the most informative. Indeed, to demonstrate the existence of a positivity bias specific to social media, it is necessary to establish that this bias is not equivalent in face-to-face social contexts (Goffman, 1959). For this reason, one solution would be to ask participants to imagine themselves narrating this event to a group of friends, and then ask them to share it on one of the three social media." (p. 13)***

Comment 23:

One aspect that I find missing in the manuscript is the rationale for measuring emoticons and the number of words in the context of positivity bias. Could you please provide some insight into this?

Response:

As explained earlier, we have provided more context on the use of emojis in the introduction, and we have removed the number of words as primary variable.

Comment 24:

Method

It would be beneficial if the authors shared the analysis script for future analyses, as well as the script used in the power analysis calculation. The authors mention, "For all analysis, we used small effect sizes (r = .3)." I'm curious to know the rationale behind considering r = .3 as a small effect size.

Response:

The power analysis script has been deposited on OSF. As mentioned in response to comment 21, we have also carefully identify the SESOI in the new version of the manuscript:

> ***"The R script for the power analysis is available at this link: https://osf.io/akgdj/?view_only=42142acd518a42cf99b33f5ebec1c780." (p. 15)." (p. 15).***

Comment 25:

In relation to the exploratory part, the authors intend to employ a similar procedure in the confirmatory part: “As with the exploratory study, participants who will not give consent to take part in the study, who will not respond to the entire study, or whose texts will not be coded by the three coders will be removed from the study.” This raises again a question about the ethical implications of using forced choice items and the lack of missing data imputations - that is, excluding participants who, for instance, only complete up to the last question in the survey. Additionally, it would be helpful to understand why all texts should not be coded by three coders.

Response:

For the main study, we anticipate fewer participants being removed because we will use the Prolific platform to recruit participants who will be compensated. Similarly, we have revised the statement to clarify that participants were not excluded because their texts were not analyzed by the coders, but rather because they did not follow the instructions.

Comment 26:

Similar methods as used in exploratory part is outlined for the variable “valence”: “the participants will have to think about an event, but this time they will not be asked to write a text to describe it.” I’m curious about how the authors plan to compare the change in valence between the text intended for the media and the original text. I noticed in the Measures section that the participants themselves will be tasked with evaluating the valence of this event. Wouldn’t this approach introduce a degree of subjectivity and bias, given that each text will be assessed from a unique perspective by a different participant? Wouldn’t it be less biased if the same researchers evaluated all the texts? Alternatively, could software sentiment analysis be used to ensure consistency in the evaluation process?

Response:

We have indeed revised the protocol to no longer rely on participants' estimated valence. In addition, we will use the Intra-Class Correlation Coefficient to verify inter-rater reliability among the coders. Finally, as mentioned before, we encounter similar issues with sentiment analysis software, which is the reasons why we prefer to use qualitative analyses.

Comment 27:

According to the Methods participants should only be able to complete the questionnaire on a smartphone. The authors attribute this to the absence of emoticons on PC. Since I don't use these networks, nor have I ever used them except on X, I consulted GPT4 :D and got this response: “Yes, you can use the same amount of emoticons on a PC as on a smartphone. Using a PC does not limit your ability to use emoticons on these platforms. You can express yourself just as freely and creatively as you would on a smartphone! 😊" So how is it?

<u>Response:</u>

While it is true that you can use emoticons on a PC just as easily as on a smartphone, it is not the typical behavior for most users. However, to enhance the ecological validity of the study, we have decided to add a question, as suggested by Reviewer 2, to determine which device participants use most often for social media usage:

> ***"We will ask participants on which devices they most often use social media (computer, tablet or smartphone)." (p. 16).***

<u>Comment 28:</u>

In the manuscript, I was unable to locate a clear justification for the use of the POMSS tool, or an explanation of how it relates to the research question.

<u>Response:</u>

Thank you for your feedback. We have removed the POMSS tool from the study as it was deemed not directly relevant to the research question.

<u>Comment 29:</u>

I'm also curious as to why the authors didn't consider a more precise measurement of social media usage frequencies. For instance, they could have used the time logs provided by social media platforms, if such a feature is available. This could potentially offer a more accurate assessment.

<u>Response:</u>

Thank you for your question. Using time logs provided by social media platforms could indeed offer a more accurate assessment of social media usage frequencies. However, this feature may not be available on all smartphones, which could introduce biases into the study. Additionally, not all participants may know how to access or interpret this information. Therefore, we opted for a method commonly used in the literature to ensure consistency and accessibility for all participants.

Reviewer 4

<u>General comment:</u>

I would like to thank the authors for their work opportunity to review this manuscript. I view the already existing exploratory work very positively and believe that the proposed study builds nicely upon this. My more detailed suggestions are listed below. Overall, I think the background section highlights the basic question but theory could be expanded and clarified in key areas, as indicated in my comments below. I like the vignette experiment that is suggested here but I also identified some issues with the methodological approach that I believe should be addressed.

While I see some challenges, I believe that it is quite possible to revise these and would encourage the authors to address the suggested changes and further follow this line of investigation.

I hope that my comments can improve this registered report and wish the authors the best of luck.

<u>Response:</u>

We sincerely appreciate the reviewer for dedicating his time to provide feedback on our manuscript. We have carefully addressed each of his comments, which are detailed below.

<u>Comment 1:</u>

Background
p. 4: When discussing self-expression across different platforms, I believe the background would benefit from employing an affordances perspective to ground it in a theoretical approach (for example, Steinert & Dennis, 2022). Some aspects seem to be touched upon but explicating and systematizing theoretical assumptions would be helpful in my opinion.

<u>Response:</u>

In our revised manuscript, we have expanded on the discussion of the positivity bias across different platforms by incorporating an affordances perspective, as suggested. This is an example from the theoretical introduction:

> ***"Secondly, affordances address not the objective features of platforms, but how users perceive them (boyd, 2010). Two affordances are especially relevant in the context of positivity bias. Shareability relates to the content shared on platforms (Masciantonio et al., 2024): Facebook is perceived as suitable for posting text and image, Instagram is mainly associated with image content and Twitter with textual content (Pittman & Reich, 2016). Image-oriented social media are associated to the most stylization from users, and thus impression management (Boczkowski et al., 2018). The visibility affordance can also be at play, focusing on the perception of the degree of visibility***

> ***of the published content (Treem & Leonardi, 2013). For example, it is lower on Facebook due to its bidirectional nature, and higher on Instagram and Twitter." (p. 5)***

Comment 2:

Sections 1.3 and 1.4 should include a more systematic overview of the mechanisms and effects at play within the different social media platforms. As it is now, the theoretical insights remain somewhat shallow. As the authors emphasize the importance of assessing specific platforms, a central contribution lies in the assessment of the platforms. I would suggest reviewing literature more in depth and explicating which mechanisms are at play in which context.

Response:

We thank the reviewer for his comment. We have completely revised the theoretical introduction. As explained above, we decided to compare platforms on the basis of three characteristics: architecture, affordances and socio-cultural context. We also explained how each characteristic might impact the manifestation of positivity bias on different social media (see section 1.2 The positivity bias on Various Social Media).

Comment 3:

In the discussion of the exploratory study, the authors highlight that motivations are central and may even be more relevant than factors assessed in the exploratory study. However, the planned research does not account for this possibility. It seems likely to me that motivations for sharing self-expression posts, yet this is not addressed or measured in the proposed study.

Response:

We decided to remove the motivations scale, as suggested by reviewer 3. Indeed, we do not have any theoretical evidence linking these motivations to the positivity bias.

Comment 4:

Method
In the planned vignette experiment, there is no text as a control/baseline condition to assess a potential positivity bias. As in the exploratory study, I believe there should be a condition to compare the social media conditions to, i.e., a non-social media condition, for example a diary entry, a description like in the initial study, etc. Without this, it is only possible to assess positivity relative to other social networks but not in general. It thus seems impossible to assess H1, as it compares the valence of social media posts vs. event valence. In my opinion, it is not possible to assess positivity bias without a text, as Likert-type questions cannot plausibly be compared to the coded valence of a text.

Response:

We decided to change the protocol based on the reviewer's feedback. On the one hand, in the pilot study, the valence on social media is compared to the valence of the event, but this is not quite comparable. On the other hand, we had proposed in the main study to compare the valence estimated by the participants of the event with the valence of their posts, but again this is not comparable. We therefore decided to compare the valence of the event told to a group of friends (non-social media condition) with the valence of the event shared on social media.

> ***"On the other hand, comparing the valence of an event with that of its expression on social media may not be the most informative. Indeed, to demonstrate the existence of a positivity bias specific to social media, it is necessary to establish that this bias is not equivalent in face-to-face social contexts (Goffman, 1959). For this reason, one solution would be to ask participants to imagine themselves narrating this event to a group of friends, and then ask them to share it on one of the three social media" (p. 13).***

Comment 5:

In the footnote the authors define the difference between emoticons and emojis and argue that both are mostly used as complementary or surrogate to text. From this background, would it not make sense to include both emojis and emoticons in these analyses, as both can be used to express positive or negative feelings?

Response:

We agree with the reviewer. We have included both emoticons and emojis in the analyses and have kept only the term emoji in the manuscript (as it includes both terms).

Comment 6:

Regarding the method of recollecting an event, I am somewhat worried about the external validity of the instructions. The experimental situation of asking to post about an event is already different from natural online behavior. Yet, within the confines of the experiment this seems like an appropriate choice. However, I am wondering how exactly the instructions will specify this. As most social media posts are about recent events, recalling a "early childhood" (p. 7) event may provide an unrealistic scenario. Similarly, writing a text may not be equally appropriate for every platform. For example, would participants post the text as a picture for Instagram, as it is an image-/video-based platform?

Response:

This is indeed one of the main limitations of the study. While social media can be used as memory albums (Pittman & Reich, 2016), it is true that most platforms are used to post about recent events, particularly since the advent of stories. For us, although we sacrifice ecological validity, we believe that it should not affect the manifestation of positivity bias

on the platforms. Furthermore, regarding Instagram being an image-oriented social media, we have modified the protocol to mitigate this issue.

> ***"To reflect the fact that Instagram is an image-oriented social media, they will also be asked an optional question: 'If you plan to use an image or photo to accompany this post, please describe it briefly here'" (p. 16).***

These limitations will be addressed in the general discussion of the manuscript.

Pittman, M., & Reich, B. (2016). Social media and loneliness: Why an Instagram picture may be worth more than a thousand Twitter words. *Computers in Human Behavior*, *62*, 155–167. https://doi.org/10.1016/j.chb.2016.03.084

Comment 7:

Another interesting research question could be whether positive and negative events are affected differently due to a positivity bias. This could be done by introducing another factor by specifying whether participants should recall a negative or positive event. However, I believe this is only something that might be considered for future research, as this would complicate the design and reduce power. It could also be an interesting exploratory question to look at differences between self-selected positive and negative events in the data.

Response:

We thank the reviewer for his suggestion. We had indeed considered this possibility, but decided that it would complicate the study and that the focus should be on the differences between the platforms. We will address this point in the discussion and carry out the exploratory analyses suggested for the main study.

Comment 8:

I appreciate that a power analysis was conducted. However, I see issues with how this was reported or conducted. First, as a pre-study exists for the effects, I would expect to use the specific effect sizes discovered, not a rule of thumb (i.e. small effect size). An effect size of r = .3 would be considered a medium effect based on the classic Cohen's (1977) convention. Please indicate where this claim comes from or if this is a mistake. To alleviate these concerns, I would suggest addressing these points and providing the power analyses scripts as open materials via the OSF folder.

Response:

We have recalculated the power analyses, by identifying the Smallest Effect Size of Interest (Lakens, 2022):

> ***"To determine the sample size, we carried out an a-priori power analysis (Lakens, 2014), using the package 'WebPower' (Zhang & Yuan, 2018). We set the alpha level to 0.05, and aimed for a power of 80%. Regarding the effect size, we identified the***

> ***Smallest Effect Size of Interest (Lakens, 2022). We used a subjective justification based on prior meta-analyses (Lakens et al., 2018). As there is no meta-analysis directly comparing positive self-presentation in person and on different social media, we relied on Ruppel et al. (2017) meta-analysis examining the difference between computer-mediated and face-to-face self-disclosure. Their findings indicated an average meta-analytic effect size of r = .211 (equivalent to f = 0.216). For a repeated ANOVA, power analysis indicated a minimum required sample size of 219 for H1 (within-subject), and of 270 for H2 (interaction). We have rounded the required sample size to 300 participants." (p. 15).***

The power analysis script was also deposited on OSF:

> ***"The coding manual, data and analyses for the pilot study can be accessed at this link: https://osf.io/akgdj/?view_only=42142acd518a42cf99b33f5ebec1c780. The coding manual and power analysis script for the main study can be accessed at the same link." (p. 6)***

Comment 9:

P. 13: Valence: I appreciate the use of a qualitative analysis. However, with a registered report, I think it is vital that the instructions for the raters are described. I was further wondering whether the authors have considered using computational text analysis methods, e.g. dictionary approaches or language models to assess the texts valence or to validate the reviewers' decisions?

Response:

Regarding the evaluation of text valence, we provided the detailed instructions given to the three researchers in the coding manual deposited on OSF: "The coders first agreed on a definition of valence as: the positive or negative character of emotions (Brosch & Moors, 2009, p. 401). For example, positive content refers to happiness, satisfaction, calm, pride or serenity. Conversely, negative content refers to anger, depression, distress, sadness or lassitude. Some content can be neutral, i.e. neither positive nor negative.".

Additionally, we considered using sentiment analysis tools; however, many of them only provide a dichotomous classification of sentiment (positive/negative), whereas we wanted a more nuanced measure for our analyses. Furthermore, finding sentiment analysis tools that support multiple languages, including French, was challenging. Finally, the interpretation of emojis depends on the context of the post: for example, some users used the "😂" emoji in a sad context, while others used it in moments of exaggerated joy.

Comment 10:

P. 15: "For all analyses, results will be considered significant if less than .05". Please specify that this means the p-value.

Response:

We made the modification:

> ***"For all analyses, results will be considered significant if the p-value is less than .05."***

Comment 11:

Results: I would suggest avoiding phrases like "tendential effects" (p. 11) to describe p-values between .05 and .1. If the pre-determined threshold is not met, the test is not significant (e.g., Gibbs & Gibbs, 2015).

Response:

We thank the reviewer for his comment. We removed these sentences from the manuscript.

Comment 12:

Clarity
The authors should clarify which of the claims they make can – with limitations – be interpreted causally and which cannot. Some interpretations of associations imply causality when the direction of the effect is not known. For example, the authors write: "[…] age and social media frequency of use seem to impact self-expression" (p. 12). However, it could be the case that different types of self-expression cause differences in frequency of media use (e.g. from a Uses-and-Gratifications Perspective) and that age is associated with a specific history and cultural impressions that could be the cause of different social media habits. Similarly: "[…] results show that as age increases, the frequency of Instagram use […] and Twitter […] decrease, while the frequency of Facebook use increases […]" (p. 11). I think this sentence implies a false causality. Aging probably does not cause individuals to use Facebook more, it could be generational differences that explain higher Facebook use for older individuals.

Response:

We agree with the reviewer and have removed these comments from the manuscript.

Comment 13:

I would suggest being specific about the constructs that are used and describing them accurately. One example for a lack of clarity about concepts/constructs can be seen in the heading "Effect of Social Media on the Number of Words". Here, this would imply that whether social media is/was used or not or to which degree would be the independent variable. However, the analysis is about the social media platform. Something like "type of social media" etc. could be helpful.

Response:

We have rewritten the entire results section. We have tried to be clearer in the choice of analyses, their tiles, and their interpretation (see section 3.3 Results).

Comment 14:

There are several spelling errors in the script, e.g. “will be carry on”, “analysis” instead of analyses. Please conduct a thorough spell check of the manuscript.

Response:

We have corrected the errors mentioned by the reviewer, and reread the article to correct any other spelling errors.

Comment 15:

Figures: Generally, I don’t think it is necessary to include “chart” in the caption, as this is redundant to the figure label.

Response:

We have withdrawn it.

Comment 16:

Figure 1: I would suggest changing the caption from “Chart of text’s valence at time 1 and 2 according to the social media” to “Text valence at time 1 and 2 by social media” or similar. I would follow a similar naming scheme for the rest of the figures.

Literature
Cohen, J. (1977). Statistical power analysis for the behavioral sciences (Rev. ed). Academic Press.

Gibbs, N. M., & Gibbs, S. V. (2015). Misuse of ‘trend’ to describe ‘almost significant’ differences in anaesthesia research. BJA: British Journal of Anaesthesia, 115(3), 337–339. https://doi.org/10.1093/bja/aev149

Steinert, S., & Dennis, M. J. (2022). Emotions and Digital Well-Being: On Social Media’s Emotional Affordances. Philosophy & Technology, 35(2). https://doi.org/10.1007/s13347-022-00530-6

Response:

We have made the suggested changes.