**Unveiling the Positivity Bias on Social Media: A Registered Experimental Study On Facebook, Instagram, And X**

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.

#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

#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

#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

#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

#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.

#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.

#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?

#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)

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

#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.

#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

#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.