

Recommender's Comments:

I examined the revised report after it was submitted and I believe that the authors have thoroughly address the comments from the reviewers. However, we noticed that the wording of hypothesis 3 in study 3 lacked clarity and there was redundancy between hypothesis 2 and 3 in the same study.

In the words of the lead author "The new Study 3 H2 logically follows from the combination of H1 and H3. If there's no relationship between total playtime and wellbeing (H1 true) + no differences between genres (H3 true) then H2 has to be false (genres do not differ from 0), whereas if there is a relationship between total playtime and wellbeing (H1 false) + no differences between genres (H3 true) then H2 has to be true (all genres differ from 0)." Based on this, the authors prefer to cut H2 for clarity and reduce the number of Study 3 hypotheses back to 2 without losing any information.

I agree with this assessment, hence, the authors are now asked to make this revision and resubmit their report.

We appreciate this being flagged, and apologize that it escaped our notice in the first instance. In the latest revision, we have removed Study 3 Hypothesis 3, condensing and clarifying the Study 3 design table while still allowing us to answer our research questions as before.

The abstract ends with "objective telemetry data". I understand what you mean given how self-reports are highly biased. Telemetry data could also have many errors (for example, inflated playtime because someone paused a game and went on a walk).

We fully agree, the use of "objective" is unwarranted. We've replaced it with "more granular digital trace data".

For study 3, hypothesis 3, is this between subjects or within subject? I think both could be interesting in different ways. A clarification might be nice.

Study 3 H3 does indeed test both within-person (H3a) and between-person (H3b) relationships. We've added that language to the main text and design table to clarify.

- Thank you for disclosing the news with Microsoft and genres. I think the mapping table in the appendix makes sense. It basically naps 1-to-1. However, would it make sense to use platform provided genres for data from all platforms rather than just for Xbox third party for consistency? Because the crowd would have revised the genre classifications coming from IGDB whereas it would not have been revised from Xbox? So, for me the "issue" is

not if the genre maps accurately from Xbox to IGDB but if the genre from Xbox is correct to begin with: i.e., it reflects how humans would largely perceive a game.

We considered using platform-provided labels, but ultimately we'd rather have the consistency and transparency of a community-led categorization than trust potentially opaque categorizations by platforms—especially since each platform might itself use different heuristics. We've added content to the limitations flagging this.

“We acknowledge that IGDB labels are themselves community-driven and prone to subjectivity, but elect to use these labels wherever possible (rather than relying exclusively on platform-provided genre labels) to maximize consistency across platforms, given that platform is likely to use a distinct set of opaque genre heuristics.”

The limitations mention “absence of third-party Nintendo data”. Can you please point me to where this was discussed in-text? I couldn't find it explained earlier.

The lack of 3rd party Nintendo data is first mentioned in Table 1 in the “type of data collected” column, and now also in the limitations. If you think this needs to be highlighted further, we are of course happy to expand.