Dear Philip Newall and co-authors,

I have now received the review of your revised PCI-RR submission from one reviewer who also reviewed your initial submission. Based on their and my own evaluations (below), I am recommending a revision of this paper. Satisfactory responses to the reviewer (and editor) comments is likely to result in a Recommendation.

Reviewer 1 makes two particularly good points which I also endorse. Regarding the reproducibility of the power analyses for H1 and H2, I suggest that you upload supporting information to the OSF showing how you arrived at these numbers -- having ran into stumbling blocks myself with this, I now upload these to remind myself exactly what I did. Regarding the equivalence tests, Reviewer 1 is correct that these are informative (and should be conducted) for both significant and non-significant effects (to discern a significant result from a practically meaningful one).

Here is my own review of the manuscript; please address these comments along with those of Reviewer 1s.

Editor Comments/Review

1.   The Abstract is rather confusing because it first reads as though you will test two conditions of the house edge format with 2000 participants, but then later (after explaining the DVs), states that you will also compare this to an existing suboptimal format with 1000 gamblers. It would be clearer to state the full design within the Abstract: specifically, the three conditions, the total sample size recruited (3000?) and then the DVs.

Response_1: This change to the Abstract has been made. Please see the tracked-changes manuscript for details.

2.   Introduction, Page 4, Line 5: I suggest removing the term "Similarly" to improve this sentence ("Gambling is another public health issue…")

Response_2: We have deleted this word.

3.   Page 4, Line 9-11: You state that "However, by comparison, gambling information can be criticized on the grounds of a lack of prominence, and suboptimalities with which it is

communicated (Newall, Walasek, et al., 2022). This Registered Report contributes to this second issue, by experimentally testing two different phrasings of relevant information about gambling products", but it's not clear that you are doing this in order to get over this suboptimality. Could you end this sentence by clarifying why you are doing this/why it's important?

Response_3: We have rephrased this sentence as follows: "This Registered Report contributes to this second issue, by experimentally comparing two equivalent alternatives to some relevant information that is currently given on many gambling products."

4.   The Introduction provides some examples of the return-to-player and the house edge communication, but these odds seem very high in favour of the gambler ("This game has an average percentage payout of 90%"). Are these factual examples? Are the odds sometimes this high in gambling?

Response_4: Yes these figures are quite representative of electronic gambling machines. The odds can be even higher on other products, such as European roulette which has an RTP of 97.3%. However, these odds are not really "high" in our opinion, as this is more representative of the illusion of high returns that the return-to-player provides. We have added citations to support this claim:

"This example of 90% means that for every £100 bet, an average of £90 will be paid-out as winnings, for a net loss of £10, and this figure of 90% is roughly representative of the average payouts of electronic gambling machines internationally (Harrigan & Dixon, 2009; Schwartz, 2013; Woolley et al., 2013)."

5.   Page 5, Line 6 states "However, seeing as how replication is an important aspect of gambling psychology research (Heirene, 2021), a secondary aim of the present research is to attempt to replicate findings on rates of understanding and perceived chances of winning from the original studies on this topic (Newall, Walasek, & Ludvig, 2020a, 2020b)", but this is quite confusing because you haven't yet outlined the main aim specifically.

Response_5: This sentence has been moved towards the end of the Introduction and is also unpacked more fully into two sentences:

"Furthermore, seeing as how replication is an important aspect of gambling psychology research (Heirene, 2021), a secondary aim of the present research is to attempt to replicate previous findings showing that house edge information results in higher rates of understanding and lower perceived chances of winning than equivalent return-to-player information. As in previous research, this study will do so via a direct replication using the original wording of the house edge (Newall, Walasek, & Ludvig, 2020a, 2020b), and also a conceptual replication using for the first time use the alternative phrasing of the house edge."

6.   As per my comment about the Abstract, Page 6 outlines the present research but does not make it clear that there are actually three conditions, with the two phrasings of 'house edge' also being compared to a 'return-to-player' condition. Can you make this clearer? The reader needs to be able to fully understand the design at this point.

Response_6: We have rephrased this part of the Introduction as follows:

"Furthermore, in following our research aim to replicate previous findings, we make a secondary directional hypothesis comparing each of the two house-edge conditions to a third condition where participants will be given equivalent return-to-player information:"

7.   Please also give the manuscript a thorough proofread – there are some instances of missing words (e.g., Page 7, 'the' in the sentence 'as this is the closest round number which exceeds the required sample size\e in each of the below power analyses".

Response_7: Thank you for spotting this error. We have corrected this and also proof-read the rest of the manuscript.

8.   Please note that at Stage 2, the approved Stage 1 text cannot be revised (unless there is a significant error or you are changing tense). For this reason, ensure that there is no information that you'd wish to remove or change in some way at Stage 2: e.g., the sentence that states "For the reviewers, a link to the experiment is here: https://bristolexppsych.eu.qualtrics.com/jfe/preview/previewId/b380c47d-7c9a-4128-a469-65082 fabdabe/SV_b1TvgkAn3B7Er7U?Q_CHL=preview&Q_SurveyVersionID=current"

Response_8: We have removed this and also checked the rest of the manuscript along these lines.

9.   What is the rationale for a change in H1's outcome from 4.1 to 3.8/4.4 (representing an effect size of 0.188) and change in accuracy of 6% (representing an effect size of d = 0.133. Put simply, why an effect of 0.188/0.133 and not any other (small) effect? You should provide a rationale for your smallest effect size of interest (SESOI).

With best wishes,

Charlotte

Response_9: This decision has now been justified on p.11:

"For H1 and H2, it is impossible to know at this stage what magnitude of change on the dependent measures would lead to meaningful differences in actual gambling environments. Therefore, we were required to proceed heuristically, by powering our study for relatively small effects which were within our budget of resources to run this study."

Reviews
Reviewed by Zhang Chen, 14 Nov 2022 10:38
Thank you for the opportunity to review this revised version of the Stage 1 RR on house edge information in gambling. Most of my previous comments have been addressed satisfactorily. I have some remaining minor comments/questions, which are listed below.

More details on the power analyses will be useful, especially the planned analysis for H1 and H2. For H1, when I used an independent t test as the planned analysis (G*Power version 3.1.9.4), d = 0.1875, alpha err prob = 0.05 and Power = 0.95, I was able to reproduce the reported results - 741 participants per condition. However, for H2, I was not able to get the same number, probably because I was not doing this correctly (I used the logistic regression in the family of z tests, and there seemed to be many extra parameters that could be set).

Response_10: We believe this was because of G-Power's default option selecting a one-tailed test. Nevertheless, we have decided to take the Editor's comment on-board with respect to power analysis reproducibility, and have decided to rerun these in R, and these have now been added to the analysis script found in the OSF repository. The results have not changed significantly, and are now described as follows:

"Given this uncertainty, we chose to explore a change on H1's outcome from 4.1 to either 3.8 or 4.4 (SD = 1.6, d = 0.188), with 95% power and an alpha of 0.05. This was calculated using the 'WebPower' package in R (Zhang et al., 2018). This identified a requirement of 1473, or 737 participants per condition. For H2, we chose to explore a change in accuracy of 6% (accuracy moving from 50% to 56% or 44%, OR = 1.27, d = 0.133), again with 95% power and an alpha of 0.05. Using the WebPower package, this identified a requirement of 933, or 467 participants per condition."

The authors proposed to use equivalence tests only when H1 or H2 is not statistically significant. I wonder if it is possible for both the null hypothesis significance test and the equivalence test to be significant - which may suggest that the effect is statistically significant (i.e. differs from zero), but not large enough to be practically significant based on the SESOI. Equivalence tests may therefore be informative regardless of whether the initial NHST results are significant or not. Related, in the tutorial on equivalence tests by Lakens and colleagues (https://journals.sagepub.com/doi/full/10.1177/2515245918770963), they also mentioned minimal effects test, which are complementary to equivalence tests and entail comparing an effect to the SESOI (d = 0.133 in this case) rather than zero. This may also be an interesting analysis to run. To be clear, I am not suggesting that the authors should include these extra analyses in the RR at this stage - just some exploratory analyses that they may consider after data collection.

I wish the authors good luck with the project and I look forward to seeing the results.

Response_11: In line with the Editor's feedback we have changed the analysis section to run the equivalence tests for H1 and H2 no matter what.