

Revision

We would like to thank the recommender for their time, encouraging comments, and insightful feedback on our first attempt at a registered report. We have responded to and addressed all the points raised, with all corresponding corrections highlighted in bold in the main manuscript.

Suggestion #1

You state one key hypothesis in the report, namely that there is no relationship between PCS and mean accuracy to incongruent stimuli in the illusion task. However, you are collecting much more information from the participants - apart from the response times in the illusion task, you also collect demographic information as well as personality assessment. Do you plan to do anything with these measures? What is the analysis plan there? And if not, why even collect the data at all?

Demographic information is collected to provide a detailed and thorough description of the sample (i.e., to be transparent about the composition of our sample from a descriptive point of view), and to have the same information available as the previous studies using these paradigms. This study won't include any statistical analysis of these variables but aims nonetheless at maximising the reusability of data (e.g., for future potential meta-analysis). The personality questionnaires are included as a strategic mean of providing a break between the two cognitively taxing blocks of a comparable activity and duration. They are also present to ensure consistency with the paradigm used before (Makowski et al., 2023) and again maximize the reusability of data. This information has been clarified in the manuscript as follows:

Lines 133-136:

“Participants will be presented with a consent form followed by demographic questions (gender, education level, age, and ethnicity). **Although these variables are directly not analyzed in the current study, they will be used to provide a detailed and thorough description of the sample and maximizing data reusability.**”

Lines 179-182:

“The two sets of 3 illusion blocks will be separated by 2 short questionnaires, namely the IPIP-6 (Sibley et al., 2011), measuring 6 personality traits with 24 analogue scales items, and the PID-5 (Krueger et al., 2011), measuring 5 maladaptive personality traits with 25 Likert scales items. **These questionnaires are included as a way of providing a break between the two cognitively taxing blocks and maintain paradigmatic consistency with previous studies (Makowski et al., 2023).**”

Suggestion #2

In line with the previous comment, I wonder what data exactly you plan to publish "born-open"? The way I read the approach, none of the data will be stored privately, but all data will immediately be made public. If you include the demographic information, anonymity seems to be out of reach here. And especially if these data are not needed because you do not plan to use it, I think it would probably be advised to remove these parts from the study.

"Born-open" refers to the concept that the experiment's pipeline is built from the start in a fully transparent way (as opposed to uploading materials to open servers a posteriori). An example of pipeline is to host the experiment's script (in our case written in javascript using JsPsych) on GitHub, use the platform to deploy the web page, and have it send the participants' data files directly to an OSF repository. These can be then downloaded, preprocessed, analyzed and saved again, all from reproducible scripts available on GitHub. Unfortunately, data collection often encompasses sensitive information, such as Prolific or Sona IDs, which prevents researchers from making the participants' raw data files available. To accommodate this, we modified the full "born-open" pipeline so that the OSF storage of data files (which contain sensitive information) is private. The preprocessing script that downloads the data also removes this information and ensures anonymization, before saving the anonymized and aggregated data back on GitHub where it is accessible for analysis. In other words, all data modification steps are transparent (the preprocessing script is available), but access to the individual raw data files is restricted to ensure confidentiality. Regarding the data not-analyzed in the current study, we would like to make it available nonetheless to maximize reusability. This information has been clarified in the manuscript as (lines 124 – 131)

"The experiment's setup follows the "born-open" philosophy (De Leeuw, 2023), with adjustments made to ensure data privacy and confidentiality. The experiment, implemented entirely using JsPsych (De Leeuw, 2015), has its code stored on GitHub and will leverage the power of the platform to host the experiment online for free. **Participant's raw data files (containing identifiers) is automatically stored in a private OSF repository. The preprocessing and analysis scripts, as well as the anonymized data, will be available directly on GitHub, ensuring the transparency and reproducibility of all the analysis steps.**"

Suggestion #3

From what I read there is a key difference between your approach and Lush et al.'s approach, which is the gamification of the illusion task. Granted you only provide the participants with aggregated performance, they might start planning to adjust for the illusion effect nevertheless. In Lush's study, there was no reason/motivation to do so. Do you think this might have an effect on your results, and if so, which?

The aim of the "score" feedback at the end of each block is to motivate participants to do their best and remain engaged throughout the blocks. Although participants might in principle

adjust their performance based on the feedback, it would require them to have an intuition of the relative value of the score (if it's a "high" score, or a "low" score) which is not straightforward given the relatively arbitrary scale of the score (in other words, they don't have a frame of reference to interpret the score). Moreover, they would also need to be aware of what strategy to adopt to influence the score, which in our opinion is not obvious: as they don't have a feedback for each trial, they don't know if they have a high proportion of errors, or if they are too slow.

That said, we agree with the recommender that the role of participants' strategy, in particular through the speed/accuracy trade-off, is something to keep in mind. Participants might indeed adopt strategies such as focusing on minimising errors at the expense of speed. One way to try to mitigate for this problem inherent to most decision-making tasks is to maintain a double-constrained instruction (i.e., always mention with equal weight to "be as fast possible without making errors" to avoid big margins of task-demand interpretation). Note that statistical models have been developed to in-principle be able to delineate some of these strategies (such as drift-diffusion models), but we decided in this study to restrict ourselves to the same paradigm previously used and focus on sensitive reproducible scores (such as the error rate).

Lines 160-162:

After each illusion type block, a score is presented (computed as a scaled Inverse Efficiency Score) as a gamification mechanism to increase motivation to perform to the best of one's abilities. **To mitigate for the potential variability in the speed/accuracy trade-off, the instructions emphasize with equal weight to be fast and to avoid errors.**