Revision

We would like to start by apologizing for the delay in submitting our response to the review process. This was due to a combination of personal and professional commitments (including a new baby). We are also grateful to the reviewers for their thoughtful and detailed feedback, which has been invaluable in improving our paper. We have carefully addressed all comments and suggestions, with changes highlighted in bold throughout the manuscript.

Editor and Reviewer comments:

Overall, the reviewers and I agreed that the report is clear and the research question and experimental approach are well justified. The main comments were about clarity of the experimental design and issues with the analysis plan. The format of the registered report allows you to make considerable changes to your analysis plan before collecting the data, and these changes are necessary. Ultimately, I am convinced that the revision will allow you to much better answer your research question.
I encourage you to read the reviewer's comments very carefully and take all of them seriously. It is not often that you get this level of helpful and detailed comments on a paper. Here are my own comments:

We thank the editor for their positive feedback on the clarity and justification of our report. We also acknowledge the importance of improving clarity and refinement in certain sections. Below, we provide our responses to the editor's comments.

One of the main changes involved changing our outcome measures to add a compound index that integrates accuracy and speed (the Inverse Efficiency Score - IES).

1. Reliability: This point is best made by Brysbaert (2924): "You cannot interpret a correlation between two variables if you do not have information about the reliability of the variables. This is especially true if you find a low correlation, because a low correlation between two variables can have two origins: the variables are not related at the population level, or the variables were not measured reliably. A variable cannot correlate with another variable any more than it correlates with itself." Your report does not mention assessing reliabilities of your two tasks even though this is vital if you want to answer your research question.

We agree with the editor on the importance of assessing reliability of our individual-level scores before interpreting correlations. Regarding our measure of phenomenological control, we will calculate Cronbach's alpha (McDonald, 1999). Hence to reflect this, the following lines have been changed (228-229):

"Participants who fail any one of these will be excluded from further analysis. **Finally, the reliability of the PCS will be assessed by computing Cronbach's alpha of its items (Cronbach, 1951)."**

Regarding the scores for illusion sensitivity, we will assess the inter-illusion reliability using the same approach as for the PCS (i.e., computing Cronbach's alpha). These alpha coefficients will be calculated separately for the Strong Illusion Strength and Mild Illusion Strength groups (see the following response for more details). If error rates and IES correlate highly within these groups, we will then recompute them on the whole data (effectively collapsing the illusion strength category) for the rest of the analysis. This approach is necessary because reaction times may not have a linear relationship with illusion strength (Makowski et al., 2023), and treating the data as a whole without this check could distort the results, leading to a crude estimate of the underlying processes.

Lines 244-261:

"To assess whether the illusions functioned as expected, stimuli will be categorized into three groups: Strong Illusion Strength & Incongruent, Mild Illusion Strength & Incongruent, and Congruent. The two outcome measures—error rate and IES—will be computed separately for each illusion and each illusion strength group. To evaluate differences between these conditions, two Bayesian t-tests will be conducted: one comparing the Congruent and Mild conditions, and the other comparing the Mild and Strong conditions. Significant differences in either IES or error rate across these comparisons will be taken as evidence that the illusions operated as intended.

Next, to determine whether the Mild and Strong Illusion Strength groups can be collapsed for further analysis, Bayesian correlations will be computed between them for each illusion and outcome measure. If these correlations are sufficiently high (r > .50, Cohen, 2013), the groups will be collapsed and outcomes recomputed across all relevant trials (i.e., by averaging across these groups). If not, the groups will be analysed separately. This step is necessary because reaction time may not have a linear relationship with illusion strength (Makowski et al., 2023), and collapsing the data without this check may obscure meaningful differences. Finally, reliability analyses will be conducted on all resulting indices, with Cronbach's alpha used to evaluate internal consistency across the three illusion types."

2. Outcome-neutral analyses: A similar point is made by Reviewer 1 about adding manipulation checks. Especially for your illusion task, you should include an analysis to assess whether the illusions work as expected. This includes an analysis of the strength of the illusion (based on your manipulation) as well as a comparison between congruent and incongruent trials.

We agree with the editor and reviewer 1 regarding the need for manipulation checks to assess whether the illusions worked as expected. As a manipulation check, we will divide the stimuli into 3 categories (Strong Illusion Strength & Incongruent; Mild Illusion Strength & Incongruent; Congruent). We expect the outcome variables Inverse efficiency score (IES) and error rate to be to be significantly different between these 3 categories (Strong > Mild > Congruent).

Lines 204-207 were changed as such:

"Visual illusion sensitivity will be measured as the average error rate in the incongruent condition, and separately for the 3 illusion types. Although the error rate is arguably a crude score, which does not take into account the effect of varying illusion strength, the interaction with task difficulty and the possible adjustments in response strategy (speed-accuracy trade off), it is also the most simple and easy to reproduce, hence its usage as our primary outcome for the current **registered report**. **As a secondary exploratory outcome, the Inverse Efficiency Score (IES, Townsend & Ashby, 2014) will also be computed. This metric incorporates both speed and accuracy by dividing the mean reaction time of correct responses by the proportion of correct responses, separately for each illusion."**

Lines 244-251:

"To assess whether the illusions functioned as expected, stimuli will be categorized into three groups: Strong Illusion Strength & Incongruent, Mild Illusion Strength & Incongruent, and

Congruent. The two outcome measures—error rate and IES—will be computed separately for each illusion and each illusion strength group. To evaluate differences between these conditions, two Bayesian t-tests will be conducted: one comparing the Congruent and Mild conditions, and the other comparing the Mild and Strong conditions. Significant differences in either IES or error rate across these comparisons will be taken as evidence that the illusions operated as intended.”

3. Dependent variable: I am not convinced by your justification of the dependent variable which only includes accuracy of the incongruent condition. For starters, you are not controlling for confounds with this variable - people may be more or less accurate in general based on their motivation, perhaps problems with vision, and many other aspects. Since you are experimentally manipulating the effect of your illusion, I suggest you rethink whether there is an alternative measure. Reviewers 2 and 3 also picked up on this requesting analyses of response times. In addition to your instructions, RT analyses may help justify that the effect really is on accuracy and not RT.

As alluded to above, in order to strike a good balance between the simplicity and reproducibility of our analysis (i.e., not relying on models with many additional researcher degrees of freedom, such as mixed models), and incorporating/adjusting for the speed-accuracy tradeoff effect on RT, we are going to add a second main outcome variable: the Inverse Efficiency Score (IES) computed by dividing the average reaction time of correct responses by the proportion of correct responses. [This measure has been used in Kuefner et al., 2010 and Statsenko et al., 2020]. This analysis will be mostly exploratory but will provide an additional index that accounts for both speed and accuracy.

Our analysis on the correlates with PC will be run on our primary outcome (the error rate) as well as on the IES. Possible discrepancies between these two indices will be discussed in terms of speed-accuracy tradeoff strategies.

1. Analysis script not available: Reviewer 2 could not find your analysis script, and I couldn't either. Please make sure it is on github for the next submission.

The analysis scripts can be accessed here:
https://osf.io/da3u6/?view_only=247d4efa1afe456aa07662732946d4e6

2. Missing information: All reviewers pointed to a lack of clarity in the methods section. I think a more detailed explanation of the tasks as well as supporting visual information would help a lot. Please include a table with the stimuli from the PCS and a figure with example stimuli from the illusion task

We agree that incorporating visual information could enhance readers' understanding of the tasks. However, we are concerned that doing so extensively within the manuscript may overcrowd the presentation. To address this, we have included a table that clarifies the task instructions and provides example stimuli for each illusion in the illusion task.

For the PCS, task instructions corresponding to each suggestion are delivered auditorily. The associated audio files and transcripts are available here:
https://osf.io/da3u6/?view_only=247d4efa1afe456aa07662732946d4e6. This link will be replaced

with the GitHub page of the current project upon completion of the review process to ensure continued anonymisation. Ratings for each task are also accessible via the same link.

Additionally, the visual stimuli used in the illusion task—specifically, the three coloured balls presented on screen—are described in the analysis section (lines 221–225).

Minor points:

1. Table 1 is difficult to read in this format. I suggest you switch rows and columns.

We have now changed the order of the columns and rows for readability (refer to manuscript).

2. p. 8, line 135-137: The sentence is a little wonky at many spots: 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.. Should be: Although these variables are not directly analyzed in the current study, they will be used to provide a detailed and thorough description of the sample and maximizing data reusability.

These changes have now been made.

Reviewer # 1

This study is a conceptual replication and extension of the Lush et al 2022 study exploring the (lack) of correlation between phenomenological control and sensitivity to a visual illusion in order to infer about the top-down vs bottom-up nature of the experience during the illusion task. I find the proposed study very interesting and the design is neat. However, I have some issues, particularly the lack of outcome neutral tests, that I think should be addressed

We appreciate the reviewers' positive feedback on the interest and clarity of our design. We have now addressed the raised issues.

I found the responses of the authors to the editor`s suggestions 1 and 2 sufficient (note that the authors repeated "to provide" in the manuscript where they extended the text to address the comment [line 136]).

This has now been addressed.

However, I also have concerns regarding the illusion game, which, I think could be addressed with the application of outcome neutral tests. The authors acknowledge that the participants can adopt strategies re speed/accuracy trade-offs and by using a crude score for illusion sensitivity, discarding all data except performance on the incongruent trials, I believe that there is a potential for bias in the outcome measure. For instance, phenomenological control may be related to individuals` strategies to cope with the illusion in the task (e.g., Palfi et al., 2022, Royal Society Open Science), and these strategies could be implemented via speed/accuracy trade-offs. Hence, if the outcome measure only focuses on accuracy, an additional analysis is needed to check and ensure that main test of interest is not confounded by speed/accuracy trade-offs.

As mentioned above, we added a new outcome (i.e., the IES) to the analysis that takes into account reaction times to also test against a measure that mitigates for speed-accuracy tradeoff.

Other outcome neutral tests could be also applied to demonstrate that the two outcome measures were successfully implemented and the collected data are of good quality to test the proposed hypothesis. For instance, I presume the illusion measure could be tested against zero to demonstrate that the participants experienced the illusion, and the same could be done for the phenomenological control measure. It could also be shown that illusion strength is related to errors. Moreover, since this study focuses on testing the correlation between two continuous measures, I was thinking that an outcome neutral test about the level of variance on the two critical outcome measures could be also applied, to ensure that there is sufficient variance in the data.

Please refer to our response to the editor's comment above regarding the addition of manipulation checks, as well as our reply to Reviewer 2 concerning the inclusion of a positive control test.

Additionally, we will plot the distribution of outcomes (our participant-level scores, namely the PC score, the error rate and the IES for each illusion) to assess the presence of inter-individual variability within the data.

I had two more comments about the analyses:

1. I believe that the Bayes factor analysis should be in line across the original and the replication studies, unless there is a strong reason to divert from the procedure of the Lush paper. Lush used half-normal prior models with informed SDs, whereas the current project aims to use a half-Cauchy prior that lacks a justification for the scale. I advise to avoid using a half-Cauchy prior, which is heavy-tailed and puts a lot of weight on the larger effects sizes compared to a half-normal prior (i.e., the Cauchy prior is biased towards evidence for the null by overestimating the size of an effect that can be realistically expected).

We agree that it is important for the Bayes factor analysis to remain consistent across the original and replication studies unless there is a strong rationale for a methodological change. We understand that Lush used half-normal prior models with informed standard deviations, whereas our current study proposes using a different prior approach. However, we want to clarify that we are not using a half-Cauchy prior in our analysis, as mentioned in your comment. Instead, we are using a shifted beta prior, which is the default prior in the BayesFactor package. This shifted beta prior distribution with a medium prior on the coefficient (i.e., the r-scale parameter set to ⅓: see Morey and Rouder (2018) for a justification) provides a balanced approach to estimating effect sizes, without placing undue weight on larger effect sizes or artificially inflating evidence for the null hypothesis.

Lines 262-273

"Bayesian correlations are then computed between PCS and illusion sensitivity scores - with the resulting IES and error rate indices. Following Lush et al. (2022), we expect to collect evidence against ($BF10 <= 1/3$) a relationship between PCS and VI sensitivity. Additionally, Bayesian correlations will be computed between maladaptive trait facets and illusion sensitivity scores. Based on prior research (Makowski et al., 2023), we expect to find evidence ($BF10 \geq 3$) supporting a relationship between the psychoticism facet of the PID-5 and illusion sensitivity. **All Bayesian analyses will be conducted using the BayesFactor package (Morey and Rouder, 2024). For correlations, a medium shifted beta prior will be applied (r-scale parameter set to 1/3), as recommended by Morey and Rouder (2018), providing a balanced approach to estimating effect sizes, without placing undue weight on larger effect sizes or artificially inflating evidence for the null hypothesis."**

We understand the rationale behind Lush's use of a half-normal prior, particularly in the context of their argument that illusion sensitivity is not influenced by top-down processes such as phenomenological control. However, as the half-normal prior assumes a positive effect (and biases the result towards it), we prefer to remain agnostic as to the direction of the effect. Specifically, if the relationship between PCS and illusion sensitivity were negative, a half-normal prior would constrain estimates toward zero or small positive effects, distorting the inference. Our study, in contrast, investigates whether a correlation exists between PCS and illusion sensitivity, rather than whether PCS predicts illusion sensitivity. Since correlations can be either positive or negative, a prior that captures both possibilities, such as the shifted beta prior, is more appropriate for our analysis.

2. Outlier selection is not completely clear regarding the manipulation checks: there will be two extra tasks to check if participants pay attention, but it is not clear what is the criteria for exclusion (i.e., one or two mistakes?)

We have now clarified the manipulation checks section (see lines 228-229):

"The PCS will contain several manipulation check indices to identify problematic participants. The phenomenological control task consists of various auditory and visual exercises. At the start of the task, participants first hear someone say "hello." They are then asked to choose from several options, including "Hello," "Goodbye," "How are you," and "Thank you." Any participant who selects an option other than the correct one will be considered inattentive. In another exercise, participants receive the following instruction: "Open your eyes. You will see only two balls on the screen…just two balls". However, three differently coloured balls are actually displayed. If participants select the option "no balls were shown", it indicates they failed to pay attention to both the auditory instructions and the visual stimuli. Lastly, in another exercise, participants are asked to press the spacebar six times. If they press it fewer than five times within the allotted time, it suggests a lack of attentiveness to the auditory instructions. **Participants will be excluded if they fail at least one of the attention checks."**

Reviewer #2

I would like to thank the authors for their timely and interesting research proposal. I have a personal interest in seeing empirical work using the Pyllusion toolbox, which I have also started using in studies this year, incidentally. My review addresses each of the Stage 1 criteria, in turn (bold statements). In summary, while I think the research question is generally well motivated, I believe the hypothesis does not fully address the question, and the planned analysis does not adequately address the hypothesis. More methodological details are also necessary to clarify the procedure for transparency and replicability purposes. Specific comments below.

We sincerely thank the reviewer for their thoughtful feedback and for highlighting their interest in our research proposal. We're especially grateful for their enthusiasm regarding the Pyllusion toolbox. We have carefully addressed each of the reviewer's comments below.

1A. The scientific validity of the research question(s).

The authors propose that computational models of Bayesian inference may explain visual illusions: theoretically, illusions present 'confusing' evidence to the senses, increasing reliance on priors to interpret the stimulus. 'Sensitivity' to illusions may depend on individual variability in prior precision, but it may also be influenced by 'phenomenological control', or the (implicit or explicit) suggestion that one will be 'fooled' or not by an illusion. Previous measures of illusion sensitivity are confounded with metacognitive judgments. The authors will use an 'Illusion game' measuring perceptual decisions via response times and error rates, allowing for a more objective measure of illusion sensitivity, to disentangle the influences of prior precision and phenomenological control on illusory sensitivity. Specifically, they ask the question:

Is the variability in illusion sensitivity related to low-level perceptual processes (e.g., baseline precision of perceptual priors), or rather to the ability to actively control and "resist" the illusion in a top-town fashion in order to achieve the task at hand (higher-level modulation of the perceptual inference parameters)?

This question is valid, but I think the motivation would be clearer to the reader if it is presented only after introducing the concept of phenomenological control and its potential influences on performance (e.g., at the end of the introduction before Table 1).

We thank the reviewer for the suggestion but the reason we present phenomenological control last is to introduce it as a possible explanation for inter-individual differences within the context of illusion sensitivity.

Furthermore, in the paragraphs describing PC, the authors mention that previous research demonstrated "the cognitive impenetrability of illusions, implying that the effect is driven by low-level processes and therefore not influenced by top-down mechanisms such as PC", but prior knowledge is top-down, as mentioned on line 55, and the authors also mention individual variability in prior precision lines 56-59. So it is important to clarify what is meant by "low-level" top-down information versus "top-down … like PC".

We agree with the reviewer that the distinction between low-level perceptual processes and top-down influences needs further clarification. For clarity we have discussed how these processes might occur at different levels. We have now changed the following lines:

Lines 114-117:

" This finding was interpreted as indicative of the cognitive impenetrability of illusions, implying that the effect is driven by low-level processes and therefore not influenced by top-down mechanisms such as PC. **Note that both prior-knowledge and phenomenological control are considered top-down processes, but the cognitive impenetrability hypothesis suggests that the processes at stake for the illusions happen at a lower- encapsulated- level (e.g., in the form of *perceptual* priors)."**

Also, where does the PC fit in a Bayesian model of illusory perception? The introduction currently describes how priors and sensory evidence interact in a Bayesian model, but it is unclear where PC fits in this.

While both prior knowledge and phenomenological control are considered top-down processes, we argue that they operate at different levels, with phenomenological control occurring at a higher cognitive level. Prior knowledge influences the perception of the illusion at the perceptual level, whereas higher-order cognitive processes, such as phenomenological control, do not. Our reasoning is based on the hypothesis that these perceptual illusions are cognitively impenetrable (as suggested by Lush et al.), meaning that the Bayesian inference process at this level remains distinct from the influence of phenomenological control.

1B. The logic, rationale, and plausibility of the proposed hypotheses, as applicable.

The authors hypothesise a null effect of PC on illusory sensitivity. To test this, they aim to recruit 500 participants and conduct Bayesian inference tests to determine the strength of evidence for or against a null result. This is an adequate measure of the strength of null effects. However, an absence of an effect of PC on illusory sensitivity would not be confirming evidence for low-level effects on illusory sensitivity (the first part of the research question). There would need to be another hypothesis/test that could provide confirming evidence for this. In the introduction, it is mentioned that individuals with schizophrenia have weak perceptual priors, so, for example, the authors could add a test to see if schizotypal traits correlate negatively with illusory sensitivity. But there needs to be some measure of low-level priors disentangled from the ‘illusory sensitivity’ measure, which could be otherwise influenced by a number of other factors not measured here (e.g., sensory processing sensitivity).

We agree with the reviewer that it would be valuable to assess whether schizotypal traits correlate negatively with illusory sensitivity. To address this, we will examine the relationship between maladaptive traits and illusion sensitivity. Specifically, we will compute Bayesian correlations between the psychoticism facet of the PID-5 and illusion-related scores (see answer to reviewer 1 and editor clarifying the study outcomes) using the *BayesFactor* package (Morey and Rouder, 2024). The design table as been updated, as well as the manuscript:

Lines 121-126:

The goal of this study is thus to replicate the results from Lush et al. (2022) pointing to an absence of a relationship between phenomenological control and illusion sensitivity, by generalising them to a different illusion paradigm that encompasses other illusion types. **Additionally, we will explore the relationship between psychoticism, as a proxy for schizophrenia, and illusion sensitivity to assess the potential impact of lower-level effects—such as weak priors observed in individuals with schizophrenia (Costa et al., 2023)—on sensitivity to illusions. These analyses may offer evidence clarifying whether inter-individual variability in illusion sensitivity is driven by lower-level perceptual mechanisms or higher-level cognitive processes (see Table 1).** “

Lines 263-268:

Following Lush et al. (2022), we expect to collect evidence against (BF10 <= 1/3) a relationship between PCS and VI sensitivity. **Additionally, Bayesian correlations will be computed using the *BayesFactor* package, employing a medium prior on the coefficient (r-scale parameter set to 1/3) to assess relationships between maladaptive trait facets and illusion sensitivity scores. Based on**

prior research (Makowski et al., 2023), we expect to find evidence (BF10 ≥ 3) supporting a relationship between the psychoticism facet of the PID-5 and illusion sensitivity."

The *Illusion Game* is designed so that perceptual difficulty and illusion strength are orthogonal, meaning that illusion sensitivity should not be conflated with general perceptual difficulty. Because of this, we argue that illusion sensitivity itself serves as an adequate proxy for assessing priors without requiring an additional independent measure. However, we acknowledge that other unmeasured factors, such as sensory processing sensitivity, could play a role and appreciate the suggestion for further disentangling these influences in future research.

1C. The soundness and feasibility of the methodology and analysis pipeline (including statistical power analysis or alternative sampling plans where applicable). The PCS currently occurs before the Illusion Game in the procedure – perhaps it should come after, as participants may become aware of what the study is measuring and perform differently in the Illusion Game because of it?

The order between the tasks is counterbalanced, to account for order effects. This is specified in lines 147-148:

"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. **Participants will then be administered the PCS and the Illusion Game task (IG) in a counterbalanced order."**

Paragraph starting line 179: I do not understand the motivation for including these questionnaires. Are these questionnaires being used to collect data for another paper? Or for an exploratory analysis? Wouldn't these measures of personality traits be good to include to potentially reveal individual differences in prior precision, as motivated in the introduction?

The questionnaires serve two primary purposes. First, they provide a break following the Illusion Game, which is expected to be cognitively demanding since participants are instructed to respond as quickly as possible without making errors. Including this break maintains consistency with the previous paradigm that employed the same illusion game (see lines 213-215). Additionally, we added the following point:

"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). **Additionally, the psychoticism subscale of the PID-5 will be used to examine the correlation between maladaptive traits and illusion sensitivity, evaluating the existence of the link proposed in previous studies (Costa et al., 2023)."**

The introduction suggests a computational explanation for the influence of prior precision over sensory evidence in illusory perception, but there is no computational model provided in the methods/analysis information. If the authors are using a Bayesian model to motivate and test their hypotheses, it seems important to include this model explicitly here, and perhaps even present predictions that would provide evidence for or against this model.

While predictive coding & Bayesian inference processes have been used to frame the effect of visual illusions (Notredame et al., 2014; Schürmann et al., 2019; Nour & Nour, 2015) , they mostly - we agree with the reviewer - remain vague about the exact inner workings of the mechanisms at stake.

However, explicitly proposing and testing a specific computational model of illusions is a large endeavour that appears to be beyond the scope of this study, which focuses on the potential link with PC.

> 1D. Whether the clarity and degree of methodological detail is sufficient to closely replicate the proposed study procedures and analysis pipeline and to prevent undisclosed flexibility in .the procedures and analyses.

In the description of the Illusion Game starting on line 150, the authors indicate that the task will be adapted to make it 'shorter and more reliable'. Please specify how this was done (e.g., shortened from 100 to 80 trials per condition? From 3 sets to 2 sets? Some illusions removed entirely?), and provide evidence that shortening the task a) makes it more reliable, and b) does not fundamentally change results. A power analysis should also be performed to justify b), perhaps on data that is already available to the authors from the 2023 paper cited.

Makowski et al. (2023) used an illusion game that included 10 different visual illusions and a total of 1,340 trials. In contrast, the current study focuses on three illusions that were found to most strongly contribute to illusion sensitivity, reducing the total number of trials to 384. In both versions of the illusion game, participants first completed a practice trial, followed by two sets of illusion blocks separated by the questionnaires. We have removed the term 'reliable' from our description of the shorter task, as we do not have a previous measure of its reliability. However, this will be clarified through the reliability analysis conducted in the current study. See lines 164-168 for this clarification.

"The task is an adaptation of the one used in Makowski et al. (2023) to make it shorter, in which participants must make perceptual judgments (e.g., "which red line is the longer") as quickly and accurately as possible. **In the original Illusion Game, 10 visual illusions were presented in two sets, following a practice trial, and separated by two short questionnaires. Participants completed a total of 1,340 trials, with the experiment lasting approximately 55 minutes. In the current procedure, only three illusions are used, selected based on the original study's findings that these illusions most strongly contribute to illusion sensitivity.** The procedure encompasses 2 sets of 80 trials for each illusion type, **preceded by a practice trial for each illusion.**

Regarding the power analysis, it is unfortunately not clear to us what to model as a function of the number of participants and trials, thus we prefer to focus on a posteriori assessments of sensibility and reliability, in particular by reporting the reliability (inter-illusion and internal consistency) of the paradigm.

Lines 162-163: It would be helpful to get a summary of the rationale and execution here without having to refer to another paper for this.

We have now provided a summary of the rationale and execution from Makowski et al. (2021).

Lines 186-188:

"For each illusion type, two continuous dimensions are orthogonally manipulated, namely task difficulty and illusion strength, so that each trial corresponds to a unique combination, **providing an objectively correct answer for each trial. The use of these manipulations allows concise, standardised reporting of illusion parameters and ensures our stimuli are fully reproducible (see Makowski et al., 2021)."**

Lines 169-170, describing incongruent and congruent trials as ‘biasing perception in the direction of the incorrect’ or ‘correct’ response, is confusing – what is the ‘correct’ response in the case of an illusion? Is ‘seeing’ the illusion correct or incorrect? And also, ‘biasing perception’ is probably more accurately phrased as ‘biasing perceptual decisions’. Perhaps ‘biasing decisions toward or against the illusory percept’?

By "congruent," we mean that the illusion manipulation makes the task easier. For example, in the Müller-Lyer illusion, when the longer line has outward-facing arrowheads (and the shorter line has inward-facing ones), the difference between the lines becomes more apparent, facilitating the judgment. Conversely, if the longer line has inward-facing arrowheads, the distinction is harder to perceive, increasing task difficulty. In this context, a "correct response" refers to accurately identifying the truly longer line. We have made the following revisions for clarity:

Lines 193-198:

“Note that the illusion effect can be either “incongruent”, **making the task more difficult by biasing perceptual decisions toward the incorrect response** or “congruent”, **making the task easier by** biasing decisions toward the correct response **(e.g., in the Müller-Lyer illusion, if the outwards-facing arrowheads are placed on the longer line, identifying which line is the longest becomes easier).** Participants respond using an arrow key (left vs. right or up vs. down), while their reaction time (RT) and accuracy are recorded.”

Lines 173-174: The measure of illusion sensitivity is error rates on ‘incongruent’ trials – I would further clarify this to the reader – does that mean the number of times a participant’s decisions are biased by the illusion?

We have clarified the meaning of “incongruent” above.

If response times are also being collected, and participants are instructed to respond ‘as fast and accurately as possible’, why is RT also not a measure of interest here? Using RTs is also motivated earlier in the introduction (lines 74-75). Its removal from analysis needs to be motivated specifically.

RTs will be used to compute the IES score which has now been proposed as another outcome measure (check response to editor for more detail).

Paragraph starting line 186: Please describe these manipulation/attention checks in a little more detail. What is the ‘negative visual hallucination suggestion’? And is the ‘5 space presses’ instruction an attention check that appears without warning? And if any participant answers either of these questions incorrectly, will they be removed from analysis?

We have now clarified the manipulation checks for the PC task (see response to reviewer 1 above)

Line 193: ‘if only the second block is bad’ – do the authors mean ‘set’ where they say ‘block’ throughout this paragraph? (I believe the authors described 2 sets of 3 blocks for each illusion type?). Is the idea that participants may have paid attention in the first set, but lost interest in the second set, leading to poor performance in the second set? Please clarify.

Indeed we meant sets instead of blocks. We have made the changes accordingly (see manuscript).

Lines 203-204: I cannot find the analysis script at the provided github page to adequately judge whether this script follows the proposed analysis in the manuscript. It is unclear what the packages tidyverse and easystats will be used for if the correlation will already be performed in the BayesFactor package. I would also suggest that any 'additional information' about the analysis be presented in the manuscript, as well.

We have now added the link to the analysis script and we clarified the correlation approach above.

1E. Whether the authors have considered sufficient outcome-neutral conditions (e.g. absence of floor or ceiling effects; positive controls; other quality checks) for ensuring that the obtained results are able to test the stated hypotheses or answer the stated research question(s).

Quality checks: I think this will be adequate once my related comments are addressed, concerning clarification about attention/manipulation checks and RT-based exclusions.
Positive controls: Positive controls may be needed for this study, as currently it seems to be proposed that the absence of one relationship (PC and illusory sensitivity) will be taken as evidence for the presence of another relationship (prior precision and illusory sensitivity).

We have now proposed (see answer elaborating on the inclusion of this analysis) that evidence for a relationship between illusion sensitivity and schizotypal traits (i.e., the psychoticism dimension of the PID-5 questionnaire) would work as a positive control, based on previous research suggesting a link between lower illusion sensitivity due to lower influence of top-down processes and more emphasis on lower-level processes such as precision of sensory information (Notredame et al., 2014; Costa et al., 2023).

Minor:

Line 38: I would not say that a 2014 reference is 'recent'

This has now been addressed (lines xx-xx).

Line 68: 'sensibility' should be 'sensitivity'

This has now been addressed

Line 82-83: 'the fact that inter-individual variability in illusion sensitivity seems to persist in this task' needs a reference

A reference has now been added.

Line 178: this is now a registered report, bumped up from preregistration status

This has now been addressed

Line 193: Typo: ‘did not properly do the text’ – ‘text’ should be ‘task’

This has now been addressed

Reviewer # 3

Thank you for the invitation to review this interesting Stage I registered report. The manuscript is generally well written and clear. My two main queries are regarding the novelty and the importance of the work, which I believe are currently insufficiently justified, and a more thorough description of the illusion types used. I am also suggesting some minor amendments to improve the overall clarity, reproducibility and readability of this report.

We appreciate the reviewer's positive feedback on the clarity and overall quality of our report. Below, we address the provided suggestions and comments in detail.

1A. The scientific validity of the research question(s)

The research question seeks to ascertain whether illusion sensitivity is related to low-level perceptual processes as opposed to high-level priors, expanding and replicating the findings of Lush et al. (2021) and Lush et al. (2022). This seems sensible and scientifically sound. However, I am currently missing a justification for the expansion of the work - in particular, why you predict that this effect would replicate across the aforementioned illusion types, and how the relevance of the predicted findings add to the current literature, above and beyond what is currently out there. My main takeaway is that the implications of adding the additional illusory tasks should be discussed in more detail.

Our study extends Lush et al. (2021, 2022) by testing whether the absence of a relationship between illusion sensitivity and top-down influences observed in the Müller-Lyer illusion holds across other visual illusions. While Lush et al. argued that the Müller-Lyer illusion is driven by low-level perceptual processes, the inclusion of the Ebbinghaus and Vertical-Horizontal illusions allows us to assess whether this effect generalizes to other illusions. Additionally, we add a positive control test by assessing the relationship between the psychoticism dimension of the PID-5 - as a proxy of schizotypal characteristics - and illusion sensitivity.

These illusions differ in their geometric and contextual properties, providing a broader test of whether illusion sensitivity is consistently influenced by low-level perceptual mechanisms rather than phenomenological control. This expansion adds value by testing the robustness of Lush et al.'s findings and offering a clearer understanding of the mechanisms underlying visual illusions.

1B. The logic, rationale, and plausibility of the proposed hypotheses (where a submission proposes hypotheses).

If I have understood correctly that you are looking for a null result, the hypothesis might be better phrased as ‘we expect to find evidence to support the absence of a relationship between VI and PCS’.

In the section ‘Rationale for Deciding the Sensitivity of the Test’ in Table 1, please could you clarify if the hypothesis is supported if this finding is for all the illusion types, or two out of three, or an alternative?

The hypothesis is supported only if evidence is found to support the absence of a relationship for all three illusions.

I also assume that the test will be based off the VI error rate based on later discussion in the methods section. Please could you clarify if this is the case, and if so, amend appropriately?

See our answers above clarifying the outcomes used.

1C. The soundness and feasibility of the methodology and analysis pipeline (including statistical power analysis or alternative sampling plans where applicable). There seems to be some appropriate justification for the sample size – however, I wonder whether further justification can be added with regards to effect sizes and sample sizes used in similar studies for the other illusion types?

There are no existing studies examining the relationship between other illusion tasks and phenomenological control, so we cannot justify this aspect based on prior research. Furthermore, the sample size of 500 was determined by practical considerations, including financial constraints.

1D. Whether the clarity and degree of methodological detail is sufficient to closely replicate the proposed study procedures and analysis pipeline and to prevent undisclosed flexibility in the procedures and analyses

On line 130, *is automatically* should be *are automatically*

This typo has now been addressed.

Are there any inclusion or exclusion criteria for your participants other than just English speakers, for example participants who have previously participated in the Lush 2021/Lush 2022 studies?

Regarding inclusion/exclusion criteria, the only specific requirement for participants is that they must be English speakers. Participants will be recruited via Prolific, enabling international recruitment and thereby reducing the likelihood that individuals who studied psychology at the University of Sussex between 2019 and 2020 will take part in the study

PCS

There is a typo in Line 137 (two full stops after the word reusability).

This typo has now been addressed.

It would be useful to see all the questions and instructions utilised in the Lush PCS in a table, as there are only ten items.

While we believe including all items directly in the manuscript may disrupt its flow and lead to overcrowding, we recognise the importance of making these materials easily accessible. Therefore,

the ten suggestions (both audio and script), along with the corresponding subjective ratings, are available for review at the following link:
https://osf.io/da3u6/?view_only=247d4efa1afe456aa07662732946d4e6

Could you clarify whether the 6-point Likert scale is from 0-5 or 1-6?

This has now been clarified in the report (see line 158):

Once the 10 suggestions are completed, participants will be asked to rate their subjective experiences and response to each suggestion on a 6-points Likert scale **(from 0-5).**

Illusion Game

How has your adaptation to the task used in Makowski et al. (2023) made the task more reliable? I would recommend describing the task by Makowski et al in detail, and then describing any changes you have made to the task and your rationale for the changes.

We have now added more information for the task used in Makowski et al., (2023) and the differences between the shorter version in our study. Notably, we do not know whether the current version of the task is more reliable, this will be clarified through the reliability analysis conducted in the current study. For this reason we have removed the term 'reliable' from our description of the shorter version of the Illusion Game. More detail has now been provided in the manuscript -and in our response to reviewer 2 - about the original task (see lines 164-168).

Please describe in more detail how the scaled Inverse Efficiency Score is calculated. Does task difficulty change based on this score?

The Inverse Efficiency Score displayed after each illusion is calculated using the formula: *100 - (IES / 35)*. This transformation is arbitrary and was designed to align with a gamified version of the task, aiming to maintain participant engagement. Lastly, this score does not change illusion difficulty.

Are there any exploratory analyses planned with reaction time?

We are planning to conduct exploratory analysis of IES as an outcome variable to examine speed-accuracy trade-offs in our study, calculated by dividing the average reaction time of correct responses by the proportion of correct responses.

There is a typo on line 193 (*text --> test*)

This typo has now been addressed - changed to task.

A major point of concern is that there needs to be a brief description of what these illusion types in the illusion game entail, perhaps in a table or figure (Ebbinghaus, Müller-Lyer, and Vertical-Horizontal), alongside rationale as to why you think that the findings of this study will replicate across all these illusion types.

We have now added a visual aid (Table 3) showing the 3 illusion types used in the illusion game, which provides an example image and brief description for each illusion. Additionally, we have added rationale as to why we think the findings of this study will replicate across these illusion types (see lines 167- 168):

"In the original Illusion Game, 10 visual illusions were presented in two sets, following a practice trial, and separated by two short questionnaires. Participants completed a total of 1,340 trials, with the experiment lasting approximately 55 minutes. **In the current procedure, only three illusions are used, selected based on the original study's findings that these illusions most strongly contribute to illusion sensitivity.** The procedure encompasses 2 sets of 80 trials for each illusion type.

There are two questionnaires mentioned – the IPIP-6 and the PID-5. Are there any analyses planned with these questionnaires?

The PID-5 will be used to compute correlations between schizotypal traits and illusion sensitivity as a positive control test. No exploratory analysis is planned for the IPIP-6 but we will keep it in the current study to remain consistent with previous studies using the Illusion Game (Makowski et al., 2023). As the data of this study will be open-access, future studies will be able to re-use this data for other purposes.

There are many outcome measures collected during this study. Please could you summarise these and outline which will be used for pre-registered hypothesis driven analyses, which will be used for exploratory analyses, and which will not be used?

Yes, there are several outcome measures in this study, and we appreciate the opportunity to clarify their use. While we acknowledge the relatively large number of outcomes, all interpretations will adopt a holistic approach grounded in both theoretical rationale and the consistency of findings across measures.

For the PCS, a single composite score will be calculated: the mean of all subjective suggestion tasks. Two items—Taste Hallucination and Post-Session Experience—require clarification, as they each involve two responses. For Taste Hallucination, the final score is the average of the sweet and sour ratings. For Post-Session Experience, the geometric mean of the urge and amnesia responses is used, following the PCS authors' guidelines.

For visual illusion sensitivity, two outcome measures will be computed: error rate (the primary outcome) and inverse efficiency score (IES), which will be explored as a secondary measure. These will be calculated separately for each illusion (Müller-Lyer, Ebbinghaus, and Vertical-Horizontal) and for each illusion strength condition: Congruent, Incongruent & Mild, and Incongruent & Strong.

To determine whether the Mild and Strong illusion conditions can be combined, we will assess the correlation between their outcome scores. If highly correlated ($r > .50$), they will be collapsed into a single "Incongruent" condition, resulting in 12 total outcomes (2 outcome types × 3 illusions × 2 conditions). If they are not sufficiently correlated, we will retain them as separate conditions, resulting in 18 outcomes (2 outcomes × 3 illusions × 3 conditions).

1E. Whether the authors have considered sufficient outcome-neutral conditions (e.g. absence of floor or ceiling effects; positive controls; other quality checks) for ensuring that the obtained results are able to test the stated hypotheses or answer the stated research question(s). I believe these are adequately covered throughout the paper. You may also want to check that the PCI and Illusion Sensitivity Game results are similar to those carried out in previous papers, in order for the results to be comparable.

We have now proposed 1 positive control analysis and one manipulation check, see response to editor.

Some general comments about the introduction
It would be useful to see a diagrammatic illustration of the different types of illusion that encompass the 'illusion game'.

We have now included two visual aids to clarify the illusion game paradigm. Figure 1 illustrates the parametric framework, detailing how difficulty and strength are manipulated. Additionally, Table 3 presents the three visual illusions used in the paradigm.

I am also not sure what the authors mean by the 'interference effect' of the task, but this might be resolved by adding a more thorough description of the task in the introduction.

We acknowledge that clarification is necessary. Traditional illusion sensitivity paradigms typically rely on participants' self-reports of their experiences with visual illusions. In contrast, the current paradigm does not specifically assess the experience of illusions themselves, but rather examines how the illusion effect impacts perceptual discrimination tasks. This approach offers a more objective measure of the illusion's influence, as it directly observes the behavioral impact of the effect of illusions on perceptual decision-making.

Lines 80-84:

"This paradigm, inspired by psychophysics, lends itself to the computational modelling of illusion sensitivity through its interference effect —**an effect that disrupts an individual's ability to accurately discriminate between perceptual stimuli. This approach aims to bypass some of the metacognitive processes involved in other paradigms, offering a more direct and objective measure of how illusions influence perceptual judgment."**

The following paragraph (line 79 onwards) is unclear to me. For instance., I am not too sure what is meant by the following statement: "resist" the illusion in a top-town fashion in order to achieve the task at hand".

See elaboration about top-down effects above to reviewer 2.

I understand that you have a description of the game in the Methods section, but I believe adding some more detail to the as well as a diagram will substantially improve the clarity of this manuscript.

We have incorporated visual information to illustrate how the illusion game is manipulated across two dimensions: strength and difficulty. Additionally, we have clarified the definitions of congruent and incongruent conditions. We hope these enhancements provide readers with a clearer understanding of both the manuscript and our analysis.

I am also unsure what is meant by the following statement: ‘Interestingly, the fact that inter-individual variability in illusion sensitivity seems to persist in this task suggests that it is not solely explained by metacognitive abilities difference’, as I would assume there is individual differences in metacognitive abilities also. I also think the wording of this should be metacognitive ability differences (as opposed to metacognitive abilities difference).

Most other illusion sensitivity paradigms involve some level of metacognitive judgments, i.e., judgments about participants’ own conscious perceptual experience. The illusion sensitivity variability observed with these paradigms might be related to variability in metacognitive abilities. The fact that illusion sensitivity interindividuality persists with another - less metacognitive - paradigm suggests that illusion sensitivity is not an exclusively metacognitive characteristic. Although out of the scope of this study, we agree with the reviewer that a study comparing the illusion sensitivity indices using various paradigms would be highly beneficial to the field.