Dear Dr. Marta Topor,

Please see below our replies to your helpful comments.

1. *I need to know whether you have already received an **ethics approval** from a relevant review board. If not, then I can only do a provisional IPA. The provisional IPA will be promoted to a full IPA as soon as you inform me about an obtained ethical approval. For more information, please refer to point 3.4. on the policies and procedures page.*

Ethics approval is still pending from our institutional review board, as we are continuing to address a few of their remaining (minor) comments. We have just submitted another reply to them, and we anticipate obtaining final approval upon their next review. However, since we are submitting this round of revisions to the Registered Report without yet having their final approval, we understand that we may receive the provisional IPA until we hear back from them and are able to relay that to you.

2. *See the letter from reviewer Marcin Koculak. The letter is very positive and provides some useful points for consideration. The reviewer states that these points are not necessary to be addressed prior to IPA, however, I would like you to consider if you would like to change anything in your Stage 1 report prior to its finalisation based on these comments. I leave this decision up to you except for one specific point I would like you to address: Please address the last point raised by the reviewer asking for **clarifications of the rationale for your hypotheses**.*

We have added clarifications of the reasoning behind the hypotheses to each of the aims subsections in section 1.4. For our detailed responses, please see our individual replies to Dr. Marcin Koculak.

3. *Regarding your sampling plan - thank you for adding more information here. It is good to have a more conservative estimate.*

    a. *However, in your report text, it is not clearly stated **what is your target sample size** - I understand that the maximum sample size is 51 because it's clearly stated in Tables 1-3. **This is not quite the case for the text**. So this needs to be edited for clarity.*

Please note that page 23 indicates both the maximum and minimum target sample sizes, but we have added the following sentence to further clarify it: “The target sample size is thus 25-51 subjects.”

    b. *In addition to the above, I don't see **a justification for why 36 is the number from which you'd start sequential analyses** to check if sufficient power has been reached in your study. Some justification is needed.*

Please note that page 23 also indicates that the reason for starting sequential analysis at 36 is in order to be the largest study of its kind to date, and as long as this justification is sound, we have clarified that our minimum sample size target is actually 25 (since the largest study of this kind currently to date is 24 subjects) (Orlowski & Bola, 2023). As such, we will aim to recruit a total minimum of 30 participants based on our estimate of 15% attrition. We have clarified these details in tables 1-3.

c. *Plus you need to clarify what sequential analyses you will conduct. So far you state that you will check BFs for the alternative and null hypotheses (BF>3, BF<⅓ respectively), but are you going to obtain BFs for all of your 17 hypotheses after each participant? What if the results provide sufficient BFs only for PCI and not LZ? Or only for some of the hypotheses and not all.* ***There needs to be a clearer statement of when exactly data collection will stop based on the different possible outcomes from the sequential analyses****.*

We plan to compute BFs for all 17 hypotheses after each participant and stop data collection only when the BFs are conclusive for all 17 hypotheses. We have added these clarifications to the main text and tables 1-3. Consistent with our overall sample size justification, we will base these analyses only on PCI and acknowledge that the resulting sample size might be underpowered for LZc. Our maximum sample size also serves as a limit to how extreme data collection based on all 17 hypotheses can get.

4. *Regarding your analysis plans - I see that you have decided to add BFs to also aid your interpretation of findings. In the last paragraph of your report, you write: "we will compute Bayes factors (to compare model evidence) and probabilities of direction (to evaluate the likelihood of positive or negative effects". You should also state* ***what criteria will be used for BFs and probabilities of direction*** *(e.g. BF>3, BF<⅓) and* ***add this information both in text and in Tables 1-3*** *in the column "Hypothesis Test Sensitivity Rationale".*

For probabilities of direction, consistent with our primary aims being non-directional, we will interpret values greater than 97.5% as moderate evidence for an effect and have added this information to the main text and tables 1-3.

5. *Finally, I would like to signpost you to the policy for changes between Stage 1 and Stage 2. I know that I have made this point before, but just want to remind you before Stage 1 is finalised that* ***no major changes will be accepted to the introduction and method sections after IPA*** *- point 3.10 in policies and procedures.*

We have taken the opportunity to make other minor refinements, particularly to the introduction, with all edits/additions indicated in green text (and any deletions indicated in ~~strikethrough~~ red text), with a clean version also uploaded to our repository.

Thank you again for your helpful comments, and we look forward to any further feedback you may have.

References

Orłowski, P., & Bola, M. (2023). Sensory modality defines the relation between EEG Lempel–Ziv diversity and meaningfulness of a stimulus. *Scientific Reports, 13*(1), 3453.

Dear Dr. Marcin Koculak,

Please see below our replies to your helpful comments.

1. *I have not seen an explicit statement in the manuscript, but I highly encourage authors to consider [making raw data publicly available] before data collection, so proper consents can be acquired from participants.*

Our data will indeed be made available per the TOP guidelines of PCI-RR.

2. *Relation between complexity and consciousness stems from theory*

    a. *Manuscript seems to be written from a perspective of treating relation between complexity and consciousness as purely empirical that requires more precise mapping to discover its nature. However as authors themselves describe in the introductory part, this relation is theory-driven and as such has inherent directionality –* ***more consciousness is connected to more complexity*** *in relevant signals. Especially since one of the motivating factors of the study is to employ the most robust measure of state to the other aspect of consciousness,* ***one should either expect the tools to perform similarly or provide a theoretical argument for observing e.g. a relation in the opposite direction****. I have seen authors response to one of the reviewers about this matter that pointed to discrepancies in data and motivation to maximizing sensitivity of the statistical models. However, I am not sure that current formulation helps to achieve this goal. From the theoretical perspective, it makes interpretation of the results less clear (especially when this relation will not be consistent throughout the planned tests). From methodological perspective, directional hypothesis testing offers more statistical power (e.g. one-sided vs two-sided tests).*

This is a very insightful observation that merits a nuanced response. Although we acknowledge a theoretical (and empirical) directionality between complexity and conscious level (primarily for PCI), we chose not to impose that directionality on our primary hypotheses, because we do not have sufficient theoretical justification that these complexity measures will behave similarly in the context of conscious *content*. In particular, this is the case for PCI, because it has not yet been investigated in this context, and this is the case for LZc, because many discrepancies have been reported. Therefore, our primary aims are focused first on establishing more robustly whether any differences in these complexity measures exist between the comprehensive range of conditions in our single, large sample-size study rather than presuming their direction.

Furthermore, even if theoretical directionality could be argued, our primary hypotheses do not incorporate any ground-truth information about the *experiences* that subjects are having to the various sensory stimuli, precisely because it is not known if/which aspects of experience these complexity measures may index. Relatedly, determining which dimensions of subjectivity may correspond to observed complexity differences is the focus of our first *exploratory* aim. For

example, it could plausibly be the case that a subject has a “more complex” *experience* to what seems like a less complex sensory stimulus, and we prefer to capture such phenomena in our exploratory analyses rather than risk biasing the primary results. We have added a comment on page 13 to reflect these rationales.

b. *Additionally, the experimental manipulation itself is constructed and coded in statistical models as having a direction of increasing complexity. If one would want to commit to agnostic approach,* ***shouldn’t the stimulus types be treated as categorical variables*** *(e.g. allowing the changes in complexity between blurred and unblurred images be independent of changes between blurred images and noise). The coding scheme in the statistical model seems to also be imposing linear changes in complexity measures between stimulus types which might obscure the actual relations between those levels if the are not actually equally separated.*

Although some studies support a linear structure across stimulus classes, we do acknowledge that other studies have indicated variability in the ordering across conditions, measures, and modalities. Therefore, in addition to our choice to use linear coding during the data collection phase, which is partially based on our proof-of-concept analyses and a pragmatic modeling choice to increase statistical efficiency, we will also explore categorical coding to assess whether other relationships exist between the conditions in the data. We have added exploratory aim 7 on page 16 to reflect these points.

3. *Quantification of conscious content*

a. *Authors follow the previous literature in manipulating the informational content of stimulation to evoke differences in complexity of the processing those stimuli by the brain (reflected then in EEG). They also focus on simultaneously controlling the perceptual complexity, so only non-sensory aspect is changing. This a valid strategy, however it is worth considering that one could argue* ***there are many other dimensions along which complexity measures could track subjective experience*** *(number of consciously experienced objects, apparentness of their features e.g. vividness, relations to each other).*

Indeed, while complexity measures may track subjective experiences on many dimensions, because very little systematic work has been done in this regard, we have chosen to start with five dimensions motivated both by theoretical considerations and prior empirical findings (see Table 6 in the manuscript). While additional dimensions could be valuable to explore, we consider these five to strike a reasonable balance between representativeness and scope.

b. *However the manuscript proposes to search for these changes in the same way it is used to quantify the global level of consciousness. While this approach might be successful,* ***there is a good chance that these conscious-content changes***

are reflected only in parts of the brain and ignoring this *can highly impact signal-to-noise ratio of the phenomenon of interest.*

There could be many ways to implement these more focused investigations, I will just mention two that stood out to me from the manuscript. Authors briefly mention in the introduction ERP correlates of conscious perception, but even looking at more general category of EEG markers, they tend to be rather local and involve only parts of the cortex. ***Maybe more granular approach would be more informative****, especially since authors want to test two modalities?*

With regards to the point about granularity, in response to a similar suggestion from a previous reviewer, we agree that this could be important, so we plan to compute the complexity measures on an electrode-by-electrode basis in our pilot data so that we can examine topographies and groups electrodes accordingly, so in the meantime, we have added this as an exploratory aim (on page 15).

c. *Another aspect is the aforementioned manipulation of non-sensory information. Whether one would call in categorical or semantic, there is a vast literature showing differences in timing of relevant processes. One could argue that presenting familiar or meaningful images vs blurred or scrambled should invoke some processes typically related to N400 component. However, authors limit their analysis to the first 400ms after stimulus onset, potentially missing relevant brain activity. On the other hand, images are to be presented for a full second.* ***Wouldn't it make sense to search for changes that are present for the whole duration of the stimuli*** *(since we can safely assume participants will be conscious of them for the whole time)?*

Our justification for limiting the analyses to the first 350 ms for auditory stimuli and 400 ms for visual stimuli is based firstly on the fact that the default response duration for PCI is 300 ms, but since PCI was developed with TMS in mind, we have added 50 ms and 100 ms to account for auditory and visual processing, respectively. Secondly, the relatively short duration is consistent with empirical findings that indicate that upon controlling for pre- and post-perceptual (cognitive) processes, the time course of sensory experiences themselves appear localizable in time to durations as early as ~120-200 ms (Dembski et al., 2021). Nevertheless, we do agree that it would be interesting to compute these complexity measures at later durations, such as [400 ms, 700 ms], and even to decompose the window [0 ms, 400 ms] into [0 ms, 200 ms] and [200 ms, 400 ms], so we have added this as an exploratory aim (on page 15).

d. *This brings up also the notion of time-resolution of the proposed measures. While they are well established and tested, they were used in substantially different manner which might strongly influence their reliability (there is a fairly recent paper by Mediano and colleagues, 2023; that discusses these issues). It seems that* ***including some additional measures that are geared toward short***

signals and more rapid changes would be beneficial *for testing the proposed hypotheses.*

The measure introduced in Mediano et al. (2023) does seem quite applicable, although since the scope of the current preregistered study is already quite substantial (including multiple versions of LZc to address previously reported discrepancies), we can investigate this in our pilot data and consider adding it to our study post hoc.

3. Finally, I think the manuscript would benefit from a **more detailed reasoning behind the hypotheses**.

We have added clarifications of the reasoning behind the hypotheses to each of the aims subsections in section 1.4.

    a. Authors want to check differences **between eyes opened and closed**, but does that mean they **treat them as manipulation of conscious content**? There are some hypotheses related to testing interactions between opening and closing the eyes and meaningfulness of stimulation, but there is **no rationale why we should expect their to be any difference (and in what direction**).

With regards to the eyes-open vs. eyes-closed aim (our primary aim 2), we are not necessarily treating this as a manipulation of conscious content, keeping in mind that our primary aims seek first to establish if/how sensitive sPCIst and sLZc are to sensory induced changes in brain activity, while our exploratory aim 1 seeks to determine if/which dimensions of subjectivity any corresponding changes in sPCIst or sLZc may be indexing. As such, similar to our response in 2a, we do not necessarily have a strong theoretical justification for a direction (nor even a difference) between eyes-open and eyes-closed conditions; however, we seek to provide new evidence to hopefully shed light on discrepancies reported in previous studies (Farnes et al., 2020; Mediano et al., 2020). If we do find a difference in complexity one way or the other, for example, that sPCIst and/or sLZc are higher in the eyes-closed condition, and these differences either do or do not correspond to differences in subjective ratings, then we can suggest potential explanations in the discussion (with caution). We have added a model specification for the latter analysis. The aim in general also gives us an independent aim from aim 1 in case the aim 1 findings are not conclusive.

    b. It is also **not clear for me if some of the hypotheses relates to differences in complexity measures between visual and auditory stimulation? Or interaction of modality and stimulation types?** Some more clarifications would benefit be useful here.

Indeed, aim 3 targets this exact question, i.e., if sPCIst and sLZc discriminate between brain responses to visual vs. auditory stimuli. We include interaction terms in our statistical models, because any differences in complexity may exist at only certain levels of stimulus granularity, which our interaction terms will allow us to determine. We have added a minor clarification in the

first column of Table 3 (indicating that these research questions address both aims 1 and 3), as well as a statement clarifying this purpose of the interaction terms on page 34 where we present our statistical models.

Please note that all edits/additions to the main manuscript are indicated in green text (and any deletions are indicated in ~~strikethrough~~ red text), and we have also uploaded a clean version to our repository.

Thank you again for your helpful comments, and we look forward to any further feedback you may have.

References

Boly, M., Sasai, S., Gosseries, O., Oizumi, M., Casali, A., Massimini, M., & Tononi, G. (2015). Stimulus set meaningfulness and neurophysiological differentiation: a functional magnetic resonance imaging study. *PloS one, 10*(5), e0125337.

Dembski, C., Koch, C., & Pitts, M. (2021). Perceptual awareness negativity: a physiological correlate of sensory consciousness. *Trends in Cognitive Sciences*, 25(8), 660-670.

Farnes, N., Juel, B. E., Nilsen, A. S., Romundstad, L. G., & Storm, J. F. (2020). Increased signal diversity/complexity of spontaneous EEG, but not evoked EEG responses, in ketamine-induced psychedelic state in humans. *Plos one, 15*(11), e0242056.

Mediano, P. A., Rosas, F. E., Luppi, A. I., Noreika, V., Seth, A. K., & Carhart-Harris, R. L. (2023). Spectrally and temporally resolved estimation of neural signal diversity. bioRxiv 2023-03. *View Article*, 2023-03.

Mediano, P. A., Rosas, F. E., Timmermann, C., Roseman, L., Nutt, D. J., Feilding, A., ... & Carhart-Harris, R. L. (2024). Effects of external stimulation on psychedelic state neurodynamics. *ACS Chemical Neuroscience, 15*(3), 462-471.