Thank You for the opportunity to review this very interesting manuscript.

The proposed study tackles an important notion of generalizability of most promising measures of state of consciousness – complexity estimates – to content that constitutes one’s subjective experience while being conscious. I am fully convinced that it will improve our understanding of this matter simply by virtue of providing new properly-powered empirical dataset (especially, if raw data will be publicly available – I have not seen an explicit statement in the manuscript, but I highly encourage authors to consider it before data collection, so proper consents can be acquired from participants). It also aims at replicating results from existing literature, which is also valuable as they are contradictory.

My main concern is related to the ability of this experimental design to provide more conclusive insights then just strengthening one side of the available contradictory results, especially when only a subset of the numerous hypotheses will reach statistical significance (which is a fairly probable scenario). I am discussing the most important aspects from my perspective in more detail below, but I think there still might be ways to either introduce some improvements or to make the theory and methodology sections more focused on resolving these discrepancies in the literature. However, I do not treat these as requirements for manuscript acceptance, but more as potential improvements of the impact of the study.

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

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). 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.

**Quantification of conscious content**

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). 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?

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)?

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.

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

Marcin Koculak

Mediano, P. A. M., Rosas, F. E., Luppi, A. I., Noreika, V., Seth, A. K., Carhart-Harris, R. L., Barnett, L., & Bor, D. (2023). *Spectrally and temporally resolved estimation of neural signal diversity*. <https://doi.org/10.7554/eLife.88683.1>