We would like to thank the reviewer for his helpful and constructive comments. Below are our point-by-point responses (in blue font) to the reviewers' comments (in black font). The line numbers refer to the manuscript with tracked changes.

Reviewer (Samuel Schwarzkopf)

Informed consent procedure

You manuscript states that written consent will be obtained after the EEG cap has been placed on the participant's head. This seems highly unusual and ethically questionable to me. I am sure that it is sensible to ask for further assent before beginning the actual experimental procedure, but from an ethical perspective the entire setup is part of the procedure and surely participants should consent to that before being subjected to it? (I have personal views of this from the few times I have suffered through that process myself...). Perhaps this is standard operating procedure there and normally approved by the local ethics committee, but even then it deserves some clarification.

This is a mistake in the manuscript, thanks for catching it! We will of course obtain written constent before the application of the EEG cap. Changes to the manuscript have been made accordingly (l. 351).

Power analysis

- You estimate power based on a one-tailed t-test of an effect size of dz=0.5. Please include a justification why this effect size is relevant for your study and the specific measures.
- Is this effect size appropriate for RTs as well as error rates? Especially when it comes to ERP amplitudes this seems questionable.
- You also propose to use a two-tailed test for the ERP results, but your power analysis is based on a one-tailed test. This is not appropriate because these two-tailed test will require a larger sample.
- So either use the sample estimated for the two-tailed test or relegate the ERP aspect to exploratory analysis - in which case *no hypothesis* about the ERPs can be mentioned at Stage 1 at all.
- There is also an error in the alpha level which is stated as 0.5 both in the text and the Design Table.

These are very valid points and helpful suggestions.
We assumed a moderate effect size of dz=0.5, thus a smaller effect size than reported by Biafora & Schmidt (2022) for the comparison dual task versus single task.
We decided to focus our analyses solely on RTs and removed error rates therefore from our hypotheses. We also decided to relegate the ERP aspect to exploratory analysis (l. 454) „Exploratory Analyses
Regarding the ERPs, we are cautious making any predictions, since, to our knowledge, the influence of task modality and task complexity on P3b amplitude and latency has not been studied so far. However, we expect that P3b amplitude and latency will be affected by both task manipulations in some way."
The alpha level was corrected to 0.05.

Hypotheses & analysis plan

- Please ensure that the statistical tests match the hypothesis they are testing. You mention ANOVAs in your analysis plan, but your power analysis is based on t-test contrasts (which seem to be the correct way to test your specific hypotheses).
- You can run the ANOVAs but each hypotheses should be accompanied with a clear description of the statistical contrast used to test it.

Thank you for drawing our attention to this.
We will be using paired samples t-tests to test the main effects of modality and complexity and changed the manuscript accordingly (l. 439) „We will conduct paired samples t-tests to test for significant differences in RTs as a measure of dual-task costs, and in priming effects between the levels of the two factors modality and complexity, as well as between single and dual-task."
We also decided to relegate the ANOVAs to exploratory analysis, to test for possible interactions of modality and complexity (l. 473) „We will also be calculating a 2 (Modality: vocal vs. manual) x2 (Complexity: high vs. low) repeated measure ANOVAs with RTs as the dependent variable to check for interactions between the factors."

No counterbalancing?

The order of your experimental conditions appears to be fixed, which would suggest a considerable risk of order effects. I would suggest counterbalancing the order across participants. If not, please provide a compelling justification why not. I can understand why the single-task condition and prime discrimination task should come at the end, but it seems wise to me to vary the order of the main dual-task conditions.

Thanks for catching this one! We will change block order for each participant randomly. Changes have been made to the manuscript. (l. 383) „The order of blocks will be randomized for each participant as to avoid order effects."

Visual inspection of ERP data

I fully appreciate that formalising some analyses a priori can be extremely difficult. However, a RR should minimise such methodological flexibility as much as possible. Choosing appropriate time windows could be done through pilot experiments where visual inspection would be fine, and then set in stone for the actual study. Without such a pilot, there needs to be much more detail on the criteria used to chose the time windows. This could also be a reason to remove the ERP hypotheses from the Stage 1 manuscript. You can still present those results as exploratory in Stage 2, provided this is flagged explicitly under a separate heading.

We agree with the objection and decided therefore to remove the ERP hypotheses from the manuscript and relegate their anaylsis to the section of exploratory analyses. (see our response to the topic „Power Analysis").

Typographic errors

Obviously, this is a minor issue but you may wish to address the following points:

- p9: "When more resources are needed than *are* available..."
- p10 "...*number* of options..."
- p12: Typos in "*monotonical* decline", *Isreal,* and "*therefore*"
- p13: "...procedures were approved..." and alpha level should be 0.05 (see above)
- p15: "...experimental manipulations will only *affect* task 2..."
- p20: Since you will analyse error rates, I suggest rephrasing the statement that only correct trials will be analysed.

Done.
We decided not to include error rates in our analyses and will therefore only include correct trials. The author is actually written Isreal.

Isreal, J. B., Wickens, C. D., Chesney, G. L., & Donchin, E. (1980). The Event-Related Brain Potential as an Index of Display-Monitoring Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *22*(2), 211–224. https://doi.org/10.1177/001872088002200210