In this proposed study, the authors plan to investigate whether not having higher order thoughts of intending facilitates not having higher thoughts of intending. They have collected pilot data with a protocol that was highly similar to the one they propose for their study. The authors plan to collect data from two groups of participants, a control and an intervention group. Both groups will be trained to imagine certain behaviors from verbal suggestion. In the intervention group, these behaviors will be actually performed by the participants. Afterwards, participants will be presented with more test suggestions and asked to imagine these behaviors as well. They will rate the strength of the effect and how involuntary the effect felt. The authors will use Bayesian statistics to evaluate the results. While the general idea of the study is interesting, I am at the moment not convinced that the authors will be able to answer their research question with the presented experiment. My largest concerns are (1) the fact that participants in the control group are never actually performing the imagined behaviors and (2) the fact that participants in the intervention group are aware that they should achieve high involuntariness. Please find a list of my major and minor suggestions below.

**Major points**

Introduction

-The introduction is made up of rather short paragraphs, each listing some previous findings/theories. I found it complicated to follow the authors’ line of thought and suggest they make stronger connections between individual aspects relevant for their study.

-Related to my previous point, I feel like the wealth of information provided in the introduction is overwhelming and often appears incoherent. Restructuring the introduction (see above) might solve this issue, but I also suggest the authors to critically evaluate which information is crucial to understand their research question. (For example, I found the short insertion on trance on page 4 more confusing than helpful)

- “Participants score their subjective responses to each item on a Likert scale from 0 to 5, with 0

indicating not feeling the effect at all and 5 indicating feeling the effect as if it were completely real.” (page 5) – what effect are you referring to here? The exact nature of the rating does not become clear to me throughout the manuscript.

-From the information presented in the introduction, I do not entirely understand the motivation for the planned research and the research question per se. This is probably due to the incoherent structure of the Introduction, as mentioned above. Similarly, I do not understand what the motivation for the pilot study was. I think the manuscript would profit from a short section where the authors summarize the most important information and make clear what is not known, i.e., needs further research, and clearly formulate the research question.

-I do not find the design of the (pilot) study convincing. Participants in the control group did not physically produce any behavior. Therefore, differences between groups could simply arise due to the lack of any motor response in the control group. One would need an additional control group that (1) imagines a behavior actively, carries out that behavior but (2) imagines it was voluntary.

Methods

-I think it is not a convincing argument to limit the number of participants based on “the number of undergraduate psychology students available and the number of experiments competing for their participation”, even though the analysis shows that > 300 participants would be optimal. Another solution would be to stretch the experiment over multiple terms or to recruit also non-students, as mentioned below.

-In the section “Pre-registered Experiment”: why will the music hallucination be swapped with the hand lowering suggestion? I see the authors’ point that they do not only want motor suggestion in the practice phase, but like this, it is still not balanced (i.e., still more motor suggestions in practice than hallucinations). Similarly, there will be more motor suggestions in the test phase as well. I think according the pilot findings, it would be highly important to keep the types of suggestion balanced.

-The authors note that they will not place any constraints for the selection of participants. I think it would make sense to carefully think about this again. E.g., it might be difficult to recruit deaf participants or such with a history of auditory hallucinations due to for example psychiatric diseases. I also wonder whether all students are by default over 18?

-It does not become clear why the changes to the experimental protocol were made after the pilot study.

-In the intervention group, participants are always informed in the training that the aim is to make the behavior feel involuntary. This likely influences their ratings of Involuntariness (for training and test) since they know involuntariness is expected from them.

**Minor points**

Title

-I found the title quite confusing and had to read it multiple times to get an idea of what the manuscript is about. The authors might consider revising the title.

Introduction

-The Introduction starts with a paragraph about the early history of hypnosis. This seems unnecessary (and even a bit confusing) since none of these historical facts are relevant to understand the present research.

-What exactly are “imaginative suggestions”? (first introduced on page 3)

-What are “successful participants”? (page 5)

-Please avoid colloquial language (e.g., “doesn’t” on page 6, “none were still highs but that almost half were mediums” on page 8)

-I do not understand what is meant by “Given the theory that experienced involuntariness is the basis of the experience feeling real” – please clarify

Methods

-Why is the availability of participants strictly connected to academic terms? Did the authors consider also recruiting non-students?

-In my opinion, one could find better words to say that one participant “failed to state their gender”. Did they actually try but did not succeed? Or rather prefer not to the state their gender? I would consider expressing this in a more neutral way, e.g. “one participant did not declare their gender”

-Were the groups in the pilot study really assigned randomly? Genders seem oddly balanced for a random attribution to groups.

-In the section “Pre-registered Experiment” I do not understand the paragraph starting “Based on the theory that…”. Please consider rephrasing, this sentence is very long.

-What about ethics approval?

-Why will the experiment be conducted in Zoom? Meeting the participants in person might guarantee a more controlled environment to do the experiment. If it needs to be done in Zoom for important reasons, this might even be another argument to also recruit non-students, since they do not even have to be physically present at the authors’ institution.

Results (pilot study)

-Aside from giving information about evidence for H1, I think readers would profit from a short summary of what exactly this means (about differences between groups/conditions).

With best wishes

Sophie Siestrup