**Revision**

We thank the editor, as well as the reviewer for the time taken to review our work, and for the encouraging comments and insightful remarks. We have answered, and addressed, all points below. All corresponding corrections are underlined or with strikethrough in the main manuscript document. Major restructuring of text and changes are highlighted in yellow. The additions/modifications (underlined) in the tables attached in the manuscript are also reflected in the individual editable tables uploaded.

**REVIEWER #1**

1. Literature.

The introduction mentions the evidence against a general factor. This part is critical for the aim of the paper, and needs to be expanded. The evidence should be described, as then the reader can compare to the new evidence. Differences in methodology are important. A few more recent papers are missing:

Cretenoud, et al. (2021) Individual differences in the perception of visual illusions are stable across eyes, time, and measurement methods. Journal of Vision 21(5):26

Cretenoud, et al. (2021). How do visual skills relate to action video game performance?. Journal of Vision, 21(7), 10.

Cretenoud, et al. (2020). Individual Differences in the Muller-Lyer and Ponzo Illusions Are Stable Across Different Contexts. Journal of Vision, 20(6):4, 1-14.

TODO. Anshu

2. Stimuli and procedure.

The description of the stimuli is not sufficient, there is too much reliance on linking to external sources. In the paper the illusions should be listed and described carefully one by one. From the list it seems that there are at least two types, some are old optical-geometrical illusions (Delboeuf, Ebbinghaus, Vertical-Horizontal, Zöllner, Müller-Lyer, Ponzo, Poggendorff) while at least two have to do with brightness (White, Contrast). I assume contrast refers to simultaneous brightness contrast.

TODO. Anshu

Brightness illusions are highly depended on the display used, and not ideal for online studies. It may be worth analysing these two sets separately.

TODO. Dom

3. Terminology

Terminology is important, and especially in the case of illusions. I will take a specific sentence to discuss a problem:

"illusions are conceptualized as ambiguous percepts (noisy sensory evidence) giving ample weight to prior knowledge to minimize prediction error and provide a coherent perceptual experience"

In most illusions the percept is not ambiguous at all, as the percept refers to the experience of the observer. What can be described as ambiguous is the stimulus.

TODO. Anshu?

4. Results

The presentation of the results also needs a complete restructuring. Most of the results are summarised in Figure 3. This Figure is very busy, confusing, and hard to understand. There are too many panels, and they are not labelled. Some are summary of raw data, some are results from model fitting. Many axes are not labelled.

There is enough material here for at least three separate figures. This way basic summary statistics (including scatter graphs) can be presented first, and the model introduced only later.

TODO. Dom. Perhaps split figure. Drop demographic data.

5. Large number of trials

Each participant makes 1340 responses (trials). I am worried that over such a long experiment (online) some different strategies may be adopted. Some people will remain motivated while others will "cheat". The problem is that these different strategies are hard to control online and they are likely to correlate with personality traits. The authors should at the very least consider this possibility.

TODO. Answer. Dom.

**REVIEWER #2**

This paper examines whether illusion susceptibility can be parametrically varied in a range of illusions and in a large sample. Furthermore it examines if generalized illusion susceptibility is influenced by demographics and personality traits. This paper has several strengths, including preregistration, a number of illusions, a large sample, open materials and very nice data visualization. The main area for improvement I believe are in the analysis section, where the use of such a large number of tests could be more clearly justified and explained in a way that makes the analyses easier to navigate given the hypotheses and aims.

Signed: Rebecca Hirst

TODO. General answer.

There are a lot of different statistical models (GAMS, Bayesian Logistic models, General linear mixed models, EFA, SEM) and it could be more clearly justified why each is needed

TODO.

and what set of transformations were used and why (e.g. line 196 – 203 indicate that different scores were transformed in different ways, log sqrt, cbrt – please can the authors outline the approach taken and models that were compared).

TODO.

The pre-registration only mentions Bayesian models and, whilst straying from pre-registration is fine, a clear justification in the text would be useful and help the reader to navigate through all of these tests and what each is doing.

TODO.

There also needs to be clearer explanation of the EFA an SEM models to allow replication. Indeed, is the EFA needed since the final SEM selected holds all illusions as independent, loading onto a single factor?

TODO. Dom.

Line 196: what is “log(diff)”? Please define “diff” here.

TODO. Anshu.

Figure 2 – it looks as though the interaction with task difficulty loads positively for some illusions and negatively for others – might the authors comment on this difference between tasks?

TODO. Dom. Different speed-accuracy tradeoffs.

Line 137 “After a brief demographic survey and a practice series of illusions” please clarify the phrasing used for each question.

TODO. Anshu.

How did the authors encourage/motivate over 1000 trials per participant in an online study? This in itself is impressive and it would be useful to report how long the task took, were participants allowed to complete over several sessions? Did all participants complete all trials? Did all participants complete all illusions?

TODO. Anshu.

Did all participants take part on laptops? Or was it possible on phones and tablets too It is mentioned that screen size was measured (Line 188) – how was screen size measured? Was a screen scaling method i.e. credit card scaling technique (Li et al 2020) included?

TODO. Anshu.

Figure 1 – It could be clearer what is meant by “direction” in these figures. I would suggest adding that to the description where task difficulty and strength are defined (top panel of figure 1 paragraph 2) . Perhaps placing boxes around the stimuli would also make it clearer which stimuli are paired together.

TODO. Dom.

Line 150 “The task was implemented using jsPsych” please can the authors also share how the study was hosted.

TODO. Anshu.

Line 156 “about £7.50” – please clarify what “about” means – was payment different for different participants?

TODO. Anshu.

Line 160 “implausibly fast” please define.

TODO. Anshu.

Figure 2 – For someone not familiar with each illusion it isn’t immediately clear how each image is “stronger” than the other. Perhaps also add to each what the question was i.e. “which is longer” to make it clear what each task entails.

TODO. Dom.

Figure 3 legend – please clarify what the x and y axes of the bottom plots correspond to i.e. they all look 0 centred?

TODO. Dom.

Figure 3 age distribution plot and personality train plots – need y labels.

TODO. Dom.