Reply to the Recommender's / Reviewers' Comments

We would like to thank the recommender and the reviewers for their comments, and have addressed each point in italics below, with additional text indicated with underline (please note: all excerpts of text are from the final version of the document, and therefore will include revisions relating to all reviewers' comments, not just the recommended revisions from a single reviewer):

(Page and line numbers are correct as per the clean manuscript.)

Recommender:

- **In section 2.2 *Experimental Procedure* you abbreviate the unimodal-visual stretching condition with the acronym UVS; however, in the rest of the manuscript you refer to this as UV. Please use a consistent abbreviation throughout.**

The UVS acronym has been removed from page 6, line 236, the unimodal-visual condition is now referred to as UV consistently throughout.

- **Moreover, you also describe the non-illusion with tactile input (NIT). However, this is the only time this condition is discussed. What role does this play in the study? Is it necessary to include for a practical reason? If so please clarify. If not, then you might consider removing this condition to save time?**

This non-illusion tactile condition is needed to see if the effects of pain reduction are due to the illusion or simply due to tactile input, as previous research has shown that tactile input alone can reduce pain ratings (Mancini et al., 2014; Nahra & Palghki, 2009) – this is mentioned in the experimental procedure section, page 6, line 243, but has also been added as an exploratory aspect of hypothesis 3 relating to pain reduction, as can be seen below:

Page 10, lines 425 – 434:

"2.4.2.3Hypothesis 3

(3) We expect to find a subjective reduction in pain, measured via a 21-point numeric rating scale, comparing before and after scores for multisensory and unimodal-visual conditions.

Pain data will also be analysed using JASP (JASP Team, 2022). Since the data will be ordinal, non-parametric Wilcoxon signed rank tests will be used to compare the dependent variable of mean pain scores before and after each independent condition. Comparisons of the visuotactile and the non-illusion tactile conditions will be exploratory and will assess whether a reduction in pain is due to the illusory manipulations or rather, due to the addition of tactile input."

- **As one reviewer points out, the EEG methods seem sparse, potentially resulting in large flexibility. Please describe in more detail how you will get from the notch-filtered data to**

the frequency power amplitudes used in the analysis, plus any other filtering or denoising steps you will apply.

The analysis section for hypothesis 2 has been updated to add more specificity to the EEG methods, as can be seen below:

Page 10, lines 408 – 420: "After pre-processing steps as mentioned in section 2.4.1 are taken, analysis of EEG data will first involve importing the waveforms from MATLAB into R, and then using R to take a Fourier transform for each waveform across all remaining electrodes, to obtain individual results per participant. These will then be averaged across all participants to give overall results, before running a dependent samples t test (two-sided) comparing MS to NI and one comparing UV to NI in the healthy group, along with a dependent samples t test (two-sided) comparing MS to NI and one comparing UV to NI in the chronic pain group. We will finally run a dependent samples t test (two-sided) comparing the healthy group baseline NI data to the chronic pain group's baseline NI data. The dependent variable will be SSSEP amplitude in μV, whilst the independent variable will be the different manipulations given in each comparison condition. No additional filtering or denoising steps will be applied to the EEG data, in line with Figueira et al.'s (2022) report that only a Fourier transform is typically needed for this type of EEG data. If differences are seen in these analyses, then there would be scope to run exploratory analyses between the healthy and chronic pain groups. Based on the pilot data in Figure 3, we would expect to see activation most pronounced over mid-frontal distributions, covering F1 and FC1 electrodes."

Harry Farmer:

1. **Page 3 end of paragraph 2 - Is the refernce to the experiencing of illusion in the unimodal condition from the same in press paper as in the sentence above? If so can this be made more clear.**

The reference to experiencing the illusion in the unimodal-visual condition is indeed from the same paper (now available as a preprint: https://doi.org/10.1101/2023.01.18.524558), therefore additional clarity has been added to show this, which can be seen in the passage below:

Page 2, lines 91 – 96: "When comparing multisensory visuotactile resizing illusions to unimodal visual resizing illusions, our recent work (Hansford et al., 2022) shows that multisensory illusions elicit significantly greater illusory experience in healthy participants, whilst also showing that a subset of participants who experienced an illusion in the unimodal visual condition reported a stronger illusory experience in this condition than in an asynchronous control condition."

2. **Page 4 paragraph 2. This section discussing the Nozaradan et al., 2012 study was somewhat unclear as to the meaning so I suggest changing the phrase "other sensory manipulations" to "other sensory modalities". In addition the phrase "consistent with previous research" used later in the paragraph seems irrelevant and should be cut.**

The mentioned changes have been made as can be seen in the text below:

Page 3, lines 148 – 157: "This paradigm has been used with other sensory modalities to better understand the neural mechanisms underlying multisensory integration, with findings showing that presentation of temporally congruent auditory and visual stimuli significantly enhances the magnitude and inter-trial phase coherence of auditory and visual steady-state responses (Nozaradan et al., 2012). However, research has also found evidence of enhanced steady-state responses for

within-modality stimulation (Giani et al., 2012), in contrast to previous findings. Research using vibrotactile stimulation has found greater increases in steady-state response magnitude when this corresponds with the amplitude modulation rate of stimulation (Colon et al., 2012; Rees et al., 1986) suggesting an entrainment of oscillatory activity to temporal features of sensory stimulation (Timora & Budd, 2018)."

3. **Page 5 paragraph 1 - It would help reader comprehension to give a short outline of the study design at the end of the intro prior to the hypotheses being listed. That way it will be clearer what the difference between the non-illiusion and unimodal illusion is. In addition, given the design the references to no illusion conditions should be pluralised to conditions**.

A short outline of the study has been added before the mention of the hypotheses as can be seen below:

Page 4, lines 180 – 188: "Using different sensory manipulations of finger resizing illusions, in addition to using an electromagnetic solenoid stimulator, this study aims to investigate subjective illusory experience and SSEP responses in both healthy and chronic pain patients, to better understand the relationship between body ownership illusions and experiences of chronic pain from subjective experience and cortical representation perspectives. To test this, different resizing illusions consisting of multisensory (visuotactile) stretching (MS), unimodal-visual stretching (UV), a non-illusion control condition without tactile input (NI), and a non-illusion control condition with tactile input (NIT) will be used to assess alternate aspects of illusory resizing manipulations and their related effects on SSEP response."

The mention of a non-illusion condition in the hypotheses has also been changed to reflect both non-illusion conditions.

4. **Page 5 paragraph 2 - Could the authors provide more information on how exclusion criteria will be determined. Will this be based on a simple self report response to each of the conditions or will you ask more detailed questions eg. regarding sleep disorders or drug and alcohol abuse?**

More information regarding how to determine inclusion and exclusion criteria has been added within the Sample Characteristics (now "Participants") section (2.2) and can be seen below:

Page 5, lines 227 – 228: "Inclusion and exclusion criteria will be determined using self-report responses relating to each item listed below:"

5. **Page 7 paragraph 1 - In the procedure section can the authors clarify whether the vibrations delivered to the finger by the solemoid will be present in all conditions. I assume from the design that this is the case but it would useful for that to be made clear.**

Clarity has been added regarding the solenoid conductor, which is to be present in all conditions:

Page 8, lines 330 – 333: "Vibrations will be delivered to the participant's finger in all conditions using a miniature electromagnetic solenoid stimulator/bone conductor (Dancer Design Tactor; diameter 1.8mm) emitting vibrations produced by sending amplified 26Hz sine wave sound files, with stimulus intensity controlled by an amplifier (Dancer Design TactAmp)."

Assuming it is the case I do wonder if this presents any issues for the design as a whole given the authors note the analgesic affects of tactile input and the fact that some tactile input is constant throughout the study. I wonder if this might be addressed by an additional condition in which there is multimodal stiumulation but not vibrotactile

stimulation, while you would lose the neural data this could at least be used for the chronic pain group to see whether the additional of vibrortactile stimulation modulated the resizing illusions affect on pain response.

We appreciate the comment about a potential issue with the study design if vibrotactile stimulation is present in all conditions, however since we specifically want to look at the steady state response in all conditions, we therefore need the vibrotactile input for all conditions. Regarding the comment about needing a condition to assess if the addition of vibrotactile stimulation modulates the resizing illusion for chronic pain participants, we included the non-illusion tactile condition to assess whether it is the addition of vibrotactile input alone, or in conjunction with the resizing illusions, that leads to a potential reduction in pain. Additional pilot data added shows that there is still an effect of the illusion with the addition of the vibrotactile input in healthy controls, and therefore we do not believe there is the need for an additional condition:

Page 12, lines 501 – 509: "*As can be seen, there is a greater subjective experience of the resizing illusion, indexed by participant's illusion score, in both experimental conditions (UV average = 64.25; MS average = 67.88) compared to both control conditions (NI average = 32.38; NIT average = 24.13). Scores below 50 are indicative of disagreement of experience of the illusion, whilst a score of 50 is a neutral option regarding the illusion experience, and scores above 50 are indicative of agreement of experiencing the illusion. This therefore shows that the addition of the vibrotactile stimulation does not remove the experience of the resizing illusion and can therefore be used in the proposed study to elicit SSEPs without affecting the subjective illusory experience of the resizing illusion.*

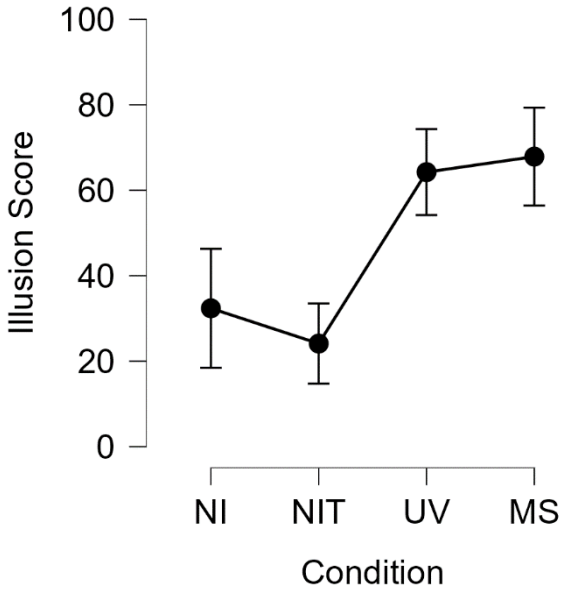

Figure 4. Averaged Illusion score for each condition. Error bars represent standard errors."

6. **Page 9 paragraph 1 - For H2 is there any expectation of SSEP differences between the two groups in any of the individual conditions and should this be tested as a separate (sub hypothesis)?**

Based on the additional text added to the introduction (see below) we can make a prediction about baseline differences in SSEP between the healthy and chronic pain participants, which is addresses in a new sub hypothesis, hypothesis 2e, as can be seen below. However, we have no expectations of SSEP differences between the two groups regarding any of the experimental conditions, and we have

no expectations regarding the direction of these. Therefore, text has been added to section 2.4.2.2 Hypothesis 2, mentioning potential exploratory analyses regarding this, which can also be seen below:

Page 2, lines 112 – 120: "There are two main theories underlying the analgesia seen during resizing illusions, firstly the somatosensory blurring hypothesis, which posits that the cortical representation of a painful body part is blurred, and that viewing the body part sharpens this representation. This is supported through findings in healthy participants, where visual analgesia has been found following experimentally induced pain (Haggard et al., 2013). The other theory is from Gilpin et al. (2015), which showed that participants with arthritis make smaller hand judgements compared to healthy participants, and posited that this could be influencing pain, as when stretching the hands, the pain was reduced. Therefore, it could be that increasing the cortical representation through magnifying the affected body part, reduced their pain.

…

Page 3, lines 159 – 165: These SSEPs can therefore be used as a measure of the somatosensory blurring hypothesis (Haggard et al., 2013) and the magnifying hypothesis (Gilpin et al., 2015), as an increased SSEP response could indicated evidence supporting the magnification hypothesis, as there is more cortical area being used to represent the body part, whereas a smaller SSEP response could indicate evidence supporting the somatosensory blurring hypothesis, as the cortical representation of the body part has become sharpened."

Page 4, lines 200 – 201: "…(2e) there will be a significant difference when comparing healthy to chronic pain participant's baseline NI SSEP responses."

Page 10, lines 408 – 420: "After pre-processing steps as mentioned in section 2.4.1 are taken, analysis of EEG data will first involve importing the waveforms from MATLAB into R, and then using R to take a Fourier transform for each waveform across all remaining electrodes, to obtain individual results per participant. These will then be averaged across all participants to give overall results, before running a dependent samples t test (two-sided) comparing MS to NI and one comparing UV to NI in the healthy group, along with a dependent samples t test (two-sided) comparing MS to NI and one comparing UV to NI in the chronic pain group. We will finally run a dependent samples t test (two-sided) comparing the healthy group baseline NI data to the chronic pain group's baseline NI data. The dependent variable will be SSSEP amplitude in µV, whilst the independent variable will be the different manipulations given in each comparison condition. No additional filtering or denoising steps will be applied to the EEG data, in line with Figueira et al.'s (2022) report that only a Fourier transform is typically needed for this type of EEG data. If differences are seen in these analyses, then there would be scope to run exploratory analyses between the healthy and chronic pain groups. Based on the pilot data in Figure 3, we would expect to see activation most pronounced over mid-frontal distributions, covering F1 and FC1 electrodes."

The additional hypothesis (2e) has also been added to relevant sections throughout the report.

Alexandra Mitchell:

> **"This study is complex and addresses many critical points, but it is worth considering whether this registration would be better off as two studies. Study one on multisensory illusions and EEG in healthy participants. Study two focusing on the pain reducing element**

of the illusion. This may simplify things somewhat. I have, however, reviewed this under the impression that it is one standalone study."

We appreciate the comment that this study might be better as two separate studies, however the rationale for maintaining this as one study is that we are interested in the group comparisons between healthy and chronic pain participants, addressed in hypothesis 2e, to see if the suggested differences in cortical representations of the affected body part leads to differences in SSEPs between the two groups, and to give the potential for further exploratory analysis between the groups, as can be seen below:

Page 4, lines 200 – 201: "…(2e) there will be a significant difference when comparing healthy to chronic pain participant's baseline NI SSEP responses."

Page 10, lines 408 – 420: "After pre-processing steps as mentioned in section 2.4.1 are taken, analysis of EEG data will first involve importing the waveforms from MATLAB into R, and then using R to take a Fourier transform for each waveform across all remaining electrodes, to obtain individual results per participant. These will then be averaged across all participants to give overall results, before running a dependent samples t test (two-sided) comparing MS to NI and one comparing UV to NI in the healthy group, along with a dependent samples t test (two-sided) comparing MS to NI and one comparing UV to NI in the chronic pain group. We will finally run a dependent samples t test (two-sided) comparing the healthy group baseline NI data to the chronic pain group's baseline NI data. The dependent variable will be SSSEP amplitude in µV, whilst the independent variable will be the different manipulations given in each comparison condition. No additional filtering or denoising steps will be applied to the EEG data, in line with Figueira et al.'s (2022) report that only a Fourier transform is typically needed for this type of EEG data. If differences are seen in these analyses, then there would be scope to run exploratory analyses between the healthy and chronic pain groups. Based on the pilot data in Figure 3, we would expect to see activation most pronounced over mid-frontal distributions, covering F1 and FC1 electrodes."

The additional hypothesis (2e) has also been added to relevant sections throughout the report.

Major Comments

1. **The most pressing issue is that the study premise appears to depend upon the participants experiencing the finger stretching illusion, in at least one of the conditions. This is especially important for the NRS after the illusion. I believe that, as well as pilot data showing the relevant EEG frequency, the authors should also pilot the illusion and present these data, in at least 4 participants, to show whether individuals can be susceptible to your illusory manipulations. This is a critical step before continuing with this stage 1 RR. Confirming that participants are likely to experience the illusion, allows you to address the question of interest with more confidence.**

We have previously tested the resizing illusion and have shown that participants do indeed experience the illusion as measured using a subjective questionnaire (Hansford et al., 2022: doi: https://doi.org/10.1101/2023.01.18.524558), although we understand that in the proposed study, the addition of the vibrotactile stimulator could affect the experience of the illusion, so we have followed your suggestion and run a pilot study with 4 healthy participants undergoing all resizing conditions with the vibrotactile stimulator as would happen in the proposed study and found the following which has been added to section 3 Pilot Data:

Page 12, lines 501 – 509: "*As can be seen, there is a greater subjective experience of the resizing illusion, indexed by participant's illusion score, in both experimental conditions (UV average = 64.25; MS average = 67.88) compared to both control conditions (NI average = 32.38; NIT average = 24.13). Scores below 50 are indicative of disagreement of experience of the illusion, whilst a score of 50 is a neutral option regarding the illusion experience, and scores above 50 are indicative of agreement of experiencing the illusion. This therefore shows that the addition of the vibrotactile stimulation does not remove the experience of the resizing illusion and can therefore be used in the proposed study to elicit SSEPs without affecting the subjective illusory experience of the resizing illusion.*

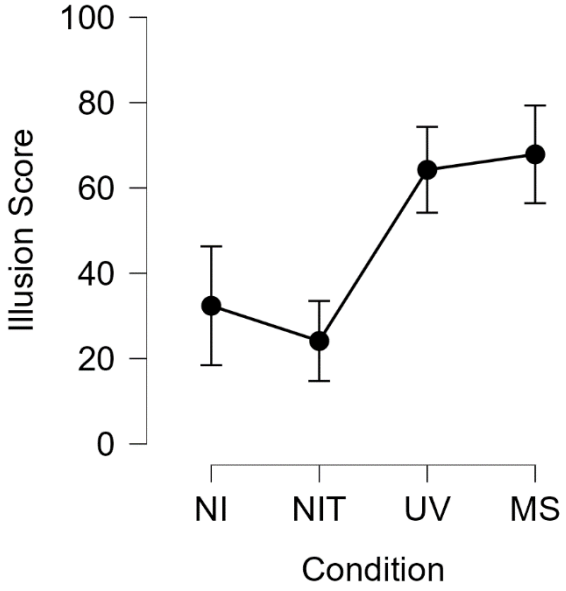

Figure 4. Averaged Illusion score for each condition. Error bars represent standard errors."

2. **The study aims are somewhat unclear. The abstract and the introduction are clearly directed at the relationship between body ownership illusions and pain (a link which is quite tenuous at best) but the focus of most of your hypotheses is, instead on whether multisensory illusion will show the strongest effect. The aims and purpose of the study need to be clear, especially with respect to the use of EEG and multisensory illusion conditions.**

Additional text has been added to the introduction to add rationale for the relationship between body ownership illusions and pain, which leads into the hypotheses relating to SSEP differences in a clearer fashion. Additional text has also been added to the introduction prior to the details of the hypotheses to clarify the aims of the study as can be seen below:

Page 2, lines 112 – 120: "<u>There are two main theories underlying the analgesia seen during resizing illusions, firstly the somatosensory blurring hypothesis, which posits that the cortical representation of a painful body part is blurred, and that viewing the body part sharpens this representation. This is supported through findings in healthy participants, where visual analgesia has been found following lab induced pain (Haggard et al., 2013). The other theory is from Gilpin et al. (2015), which showed that participants with arthritis make smaller hand judgements compared to healthy participants, and posited that this could be influencing pain, as when stretching the hands, the pain was reduced. Therefore, it could be that increasing the cortical representation through magnifying the affected body part, reduced their pain.</u>

…

Page 3, lines 159 – 165: These SSEPs can therefore be used as a measure of the somatosensory blurring hypothesis (Haggard et al., 2013) and the magnifying hypothesis (Gilpin et al., 2015), as an increased SSEP response could indicated evidence supporting the magnification hypothesis, as there is more cortical area being used to represent the body part, whereas a smaller SSEP response could indicate evidence supporting the somatosensory blurring hypothesis, as the cortical representation of the body part has become sharpened."

Page 4, lines 180 – 188: "Using different sensory manipulations of finger resizing illusions, in addition to using an electromagnetic solenoid stimulator, this study aims to investigate subjective illusory experience and SSEP responses in both healthy and chronic pain patients, to better understand the relationship between body ownership illusions and experiences of chronic pain from subjective experience and cortical representation perspectives. To test this, different resizing illusions consisting of multisensory (visuotactile) stretching (MS), unimodal-visual stretching (UV), a non-illusion control condition without tactile input (NI), and a non-illusion control condition with tactile input (NIT) will be used to assess alternate aspects of illusory resizing manipulations and their related effects on SSEP response."

3. **Although the authors have argued their reasoning for a 21-point pain scale I would argue that a 10-point scale is still preferable. One of their reasons is that using a more unusual scale will make the participants think about their pain levels more and be 'less automatic'. I disagree with this take as I believe presenting the 21-point scale will make it less comparable and therefore make it more challenging for participants to rate their pain. The 10-point scale is also more translatable and comparable with different studies, and therefore makes the work more reproducible and replicable. I am happy to hear a rebuttal to this point.**

When 21-point scales have been used in previous work to measure pain levels (Preston & Newport, 2011; Preston, Gilpin & Newport, 2020), there have been no reported issues for the participants to rate their pain on this less commonly used scale, and we believe that whilst the use of an 11-point scale would potentially be more translatable and comparable to research from medical settings where this scale is most commonly used, the use of the 21-point scale in the proposed study will allow for more appropriate comparisons with those previous research studies mentioned. Additionally, from working with chronic pain participants, it is apparent that some individuals can get very used to increasing the level of pain that they report on the typically used 11-point scale to try and get help from medical professionals for historically misdiagnosed and misunderstood chronic pain conditions, such as fibromyalgia. Therefore, using a 21-point scale will hopefully remove the likelihood of a participant using a heightened value that they are used to using in medical settings, and hopefully give a more valid measurement of their pain.

4. **It would add weight if you were able to predict a direction of the change in SSSEPs between the illusory and non-illusory conditions**

Previous literature (Gilpin et al., 2015; Haggard et al., 2013) indicate the possibility for different directions of SSEP response, with Gilpin et al's work suggesting a heightened response, whilst Haggard et al.'s work suggests a reduced response (This narrative has been added to the manuscript as can be seen from the additional text added to the introduction as can be seen in Major Comment point 2's response). Therefore, we are not able to predict a direction of change, but instead are interested to see a change in either direction, as this will expand knowledge of SSEP responses in finger resizing illusions and provide support for either theory from previous literature. Given

directional findings from the proposed study, further research / replications should be run to consolidate the existence of these directional effects.

5. **My final major comment is somewhat related to the second. Although you are recruiting from two groups, none of the pre-registered hypotheses address comparisons between the two groups. Why is this? Between group hypotheses appear, to me at least, to be a natural continuation of the argument shaped in the introduction. Comparisons in illusory strength and susceptibility, as well as SSSEPs, between the two groups may be interesting, as we do not know whether chronic pain may alter the experience of this illusion or interact specifically with the tactile elements of the experiment and how this presents at the neuronal level.**

We have no expectations of illusory strength or SSEP differences between the two groups regarding any of the experimental conditions, however we have added a hypothesis (2e) regarding expected differences seen between the 2 groups at baseline. Text has also been added to section 2.4.2 for Hypotheses 1 and 2, mentioning potential exploratory analyses, which can be seen below:

Page 9, lines 386 – 398: "The subjective illusory experience questionnaire will be used as a positive control for the current study. Previous research has shown significantly greater illusion strength for multisensory conditions compared to non-illusion conditions, which we will attempt to replicate. Questionnaire data will be analysed using JASP (JASP Team, 2022). An ANOVA will be run to compare the dependent variable of mean illusion score from each independent condition. Given significant findings, post-hoc tests will be run, with Bonferroni correction for 3 comparisons at an initial alpha of 0.05. To identify participants who effectively experience the unimodal visual condition for hypothesis 1b, participants will be included in analysis if their averaged illusion scores on the subjective illusory questionnaire scale for the unimodal-visual condition are greater than 1, in line with previous research using mean subjective embodiment scales (Carey et al., 2019), which will indicate experience of the illusion. If the positive control is successful in these analyses, then there would be scope to run exploratory analyses between the healthy and chronic pain groups regarding their subjective illusory experiences."

Page 4, lines 200 – 201: "…(2e) there will be a significant difference when comparing healthy to chronic pain participant's baseline NI SSEP responses."

Page 10, lines 408 – 420: "After pre-processing steps as mentioned in section 2.4.1 are taken, analysis of EEG data will first involve importing the waveforms from MATLAB into R, and then using R to take a Fourier transform for each waveform across all remaining electrodes, to obtain individual results per participant. These will then be averaged across all participants to give overall results, before running a dependent samples t test (two-sided) comparing MS to NI and one comparing UV to NI in the healthy group, along with a dependent samples t test (two-sided) comparing MS to NI and one comparing UV to NI in the chronic pain group. We will finally run a dependent samples t test (two-sided) comparing the healthy group baseline NI data to the chronic pain group's baseline NI data. The dependent variable will be SSSEP amplitude in µV, whilst the independent variable will be the different manipulations given in each comparison condition. No additional filtering or denoising steps will be applied to the EEG data, in line with Figueira et al.'s (2022) report that only a Fourier transform is typically needed for this type of EEG data. If differences are seen in these analyses, then there would be scope to run exploratory analyses between the healthy and chronic pain groups. Based on the pilot data in Figure 3, we would expect to see activation most pronounced over mid-frontal distributions, covering F1 and FC1 electrodes."

Minor/in text comments

As there are no line or page numbers (please add these for resub), I made my own. Page 1 is the page after the title page with the abstract and start of the introduction on it.

Line and page numbers have now been added to the manuscript

1. **The introduction very long and can be cut down substantially. It is not clear why there is so much detail about specific concepts (e.g. predictive coding accounts of pain, the RHI)**

The text regarding the predictive coding and central sensitisation accounts of pain have been cut down as can be seen in the resulting text below:

Page 1, lines 55 – 63: "Theories underlying this cortical misrepresentation are the predictive coding account (Friston, 2008) and the central sensitisation theory (Arendt-Nielsen & Graven-Nielsen, 2003; Arendt-Nielsen et al., 2010). Predictive coding posits that any mismatch between predicted and actual sensory inputs, such as the difference between peripheral signals and symptomatic pain, generates prediction errors. A possible lack of updating of top-down expectations in chronic pain individuals, could lead to constant mismatches between symptomatic and radiographic painful sensory inputs. Central sensitisation theory, however, refers to the central nervous system changing, distorting, or amplifying pain in a way that no longer reflects the peripheral input from the body, leading to pain becoming an illusory perception (Woolf, 2011)."

The text referring to the RHI has also been cut down as can be seen in the resulting text below:

Page 2, lines 70 – 74: "This illusion is based on the rubber hand illusion, in which touch is delivered to a visible fake hand at the same time and in the same place that touch is delivered to the hidden real hand. This manipulation elicits feelings of ownership over the fake hand, through the integration of multisensory (tactile and visual) inputs highlighting the apparent malleability of bodily self (Botvinick & Cohen, 1998)."

2. **On a similar note, the relevant illusion (finger stretching illusion) needs to be much more clearly described either in the introduction or methods**

The specifics of the finger stretching illusion have been detailed more at the end of the introduction as can be seen in the text below:

Page 4, lines 180 – 188: "Using different sensory manipulations of finger resizing illusions, in addition to using an electromagnetic solenoid stimulator, this study aims to investigate subjective illusory experience and SSEP responses in both healthy and chronic pain patients, to better understand the relationship between body ownership illusions and experiences of chronic pain from subjective experience and cortical representation perspectives. To test this, different resizing illusions consisting of multisensory (visuotactile) stretching (MS), unimodal-visual stretching (UV), a non-illusion control condition without tactile input (NI), and a non-illusion control condition with tactile input (NIT) will be used to assess alternate aspects of illusory resizing manipulations and their related effects on SSEP response."

More specifics about the finger stretching illusion can be seen in the experimental procedure section (2.3) within the methods, where wording has been adjusted and additional details added as can be seen below:

Page 6, lines 278 – 322: "Participants will then be seated behind the augmented reality system (Figure 1) and instructed to place their hand onto the black felt fabric within the augmented reality system. Within the self-built system there is a 1920 x 1080 camera situated in the middle of the area, away from the participant's view. Above this area, there is a mirror placed below a 1920 x 1200 resolution screen. Chronic pain participants will be asked which digit is in the most pain and will be asked to place this digit outstretched onto the felt. If multiple digits are equally painful, the digit that the participant chooses as their preference will be used. Healthy participants will be asked to outstretch a digit that has been matched to that of a chronic pain participant. There will be two white dots for each hand on the felt and participants will be instructed to place their hand between these two dots. Participants will be instructed to view their hand's image in the mirror (the real hand will be hidden from view) throughout the experiment. The camera placed underneath the mirror on the felt base will be used to deliver a live feed video of the participants hands to the computer screen at the top of the augmented reality system, which will show in the mirror reflection to the participants. Participants will undergo 4 conditions: multisensory stretching (MS), unimodal-visual stretching (UV), a non-illusion control condition without tactile input (NI), and a non-illusion control condition with tactile input (NIT). There will be vibrotactile stimulation to the finger in all conditions, but only tactile input of the researcher touching the participants hand in the conditions where this is mentioned. Each trial will last 2.4 seconds for the manipulation phase, where the finger will be stretched by 60 pixels in UV and MS conditions, followed by a further 2.4 second habituation phase in which participants can view and move their augmented finger before the screen goes dark, indicating that the next trial can start. MS conditions consist of the researcher touching and pulling the participant's finger as the participant views their finger stretching in a congruent manner. UV conditions consist of the participants viewing their finger stretch without any experimenter manipulation. The NI condition provides no visual or touching tactile manipulations to the finger. The second NI control condition will involve no visual input of the finger stretching, instead the image of their finger will be visible but unchanged. Additionally, this condition will include tactile input of the experimenter's hand touching the participants finger, but without pulling. Previous research has suggested that tactile input alone can reduce pain ratings (Mancini et al., 2014; Nahra & Palghki, 2009), therefore this second control condition will demonstrate if it is the illusion itself that is driving any changes in analgesia rather than the tactile or combined sensory inputs. The experimenter will be seated opposite the participant, the other side of the augmented reality machine and will pull the digit by holding onto the distal interphalangeal joint and gently pulling the finger whilst the participant keeps their hand in place. Conditions will be delivered across 4 blocks, with each block consisting of 24 trials of the same experimental condition, totalling 96 trials over all 4 blocks. The ordering of the blocks will be randomised for each participant to prevent ordering effects."

3. **The hypotheses paragraph in the introduction is very difficult to follow and I suggest saving most of the detail for Section 2.4 and summarising clearly in the intro.**

Further clarity has been added to the hypotheses paragraph, as mentioned in the previous point.

4. **Matching groups is hard to achieve, I assume the authors will recruit chronic pain patients first and then healthy controls. But details on how this will be achieved should be present in the manuscript.**

There have been added details regarding the matched pairs design as can be seen from the text below:

Page 5, lines 218 – 221: "Chronic pain and healthy participants will be matched based on sex, age and handedness, creating a matched pairs design experiment. Due to increased difficulties recruiting

from clinical populations, chronic pain participants will be recruited first, and then healthy participants matched on gender, age and handedness will be recruited."

5. **Recruitment – is there a recruitment age limit?**

An age limit of 75 years has been added to the healthy and chronic pain participants inclusion / exclusion criteria as can be seen underlined below:

Page 5, lines 221 – 246: "…An upper age limit of 75 years is used based on data from the NHS (2019) showing rates of chronic pain conditions increasing from 16% among people aged 16-24, to 53% for those 75 years and older. All participants will take part in all illusory conditions and will complete the subjective illusory experience questionnaire, with chronic pain participants also completing the pain rating scale.

Sample inclusion / exclusion criteria:

Inclusion and exclusion criteria will be determined using self-report responses relating to each item listed below:

- *Inclusion Criteria: Right-handed, over 18 years of age, no older than 75 years of age.*

**Chronic pain participant specific inclusion criteria: must have a diagnosed chronic pain condition involving current hand-based pain in the right hand, hand-based pain present on day of testing.*

- *Exclusion Criteria: Prior knowledge or expectations about the research, a history of developmental, neurological or psychiatric disorders, history of drug or alcohol abuse, history of sleep disorders, history of epilepsy, having visual abnormalities that cannot be corrected optically (i.e. with glasses), or being under 18 years of age, or over 75 years of age.*

**Healthy participant specific exclusion criteria: a history of chronic pain conditions, operations or procedures that could damage peripheral nerve pathways in the hands, current experiences of pain or more than 4 hours of consistent pain experienced in the preceding week.*

**Chronic pain participant specific exclusion criteria: Diagnosed with Complex Regional Pain Syndrome, no restrictions apply regarding any medication the participant might be taking. (Complex Regional Pain Syndrome is excluded as a chronic pain condition here, due to research showing increasing pain after stretching illusions (Moseley et al., 2006)."*

Reference - NHS Digital (2019) Health survey for England 2017. NHS Digital. https://digital.nhs.uk

6. **Recruitment – healthy participants who are currently experiencing pain or have experienced more than 4 hours of consistent pain the last week should also be excluded**

This is a worthwhile point to make, and has been added to the healthy participants exclusion criteria section as can be seen below:

*Page 5, lines 238 – 241: "*Healthy participant specific exclusion criteria: a history of chronic pain conditions, operations or procedures that could damage peripheral nerve pathways in the hands, current experiences of pain or more than 4 hours of consistent pain experienced in the preceding week."*

7. **Recruitment – why are operations that could damage peripheral nerve pathways an exclusion criterion in the chronic pain group? Some of these procedures may result in chronic pain**

This is a valid point and as a result this has been removed from the exclusion criteria for chronic pain participants. The section now reads as follows:

*Page 5, lines 242 – 246: "*Chronic pain participant specific exclusion criteria: Diagnosed with Complex Regional Pain Syndrome, no restrictions apply regarding any medication the participant might be taking. (Complex Regional Pain Syndrome is excluded as a chronic pain condition here, due to research showing increasing pain after stretching illusions (Moseley et al., 2006)."*

8. **How will pain levels to recruit participants in chronic pain group be assessed before and on the day?**

Clarity has been added to the Experimental Procedure (2.3) section to show that the 21-point NRS will be administered on the day of recruitment and then again on the day of testing, as can be seen below:

Page 6, lines 260 – 265: "Clinical participants will then be asked questions regarding what chronic pain condition they are diagnosed with, how long it has been since diagnosis, what medications (if any) they are taking, and their pain score on that day using a 21-point numeric rating scale (NRS) (0 = no pain at all; 20 = most severe pain imaginable). This 21-point NRS will also be administered on the morning of the day of testing, to check that the participant has pain in their hand on that day."

9. **Analysis pipeline page 7 – questionnaire data should be averaged to give median scores as questionnaires are ranked**

This has been amended as can be seen in the text below:

Page 9, lines 371 – 373: "Regarding questionnaire data, scores for both illusion experience questions will be averaged to give median scores, along with both disownership questions and both control questions, resulting in 3 median scores per trial per participant."

10. **Planned analysis pages 7 and 8 – provide specific predictors for ANOVAs and t-tests so reader knows exactly what is being compared. For all hypotheses – hypothesis 3 is particularly lacking in detail**

Further clarity has been added regarding the independent (predictor) variables for the statistical tests, as can be seen from the text below from each hypothesis:

Page 9, lines 389 – 390: Hypothesis 1: "A one-way ANOVA will be run to compare the dependent variable of median illusion score from each independent condition."

Page 10, lines 416 – 417: Hypothesis 2: "The dependant variable will be SSSEP amplitude in µV, whilst the independent variable will be the different manipulations given in each comparison condition."

Page 10, lines 431 – 432: Hypothesis 3: "Wilcoxon signed rank tests (one-sided) will be used to compare the dependent variable of median pain scores before and after each independent condition."

11. **Tables or appendices should be indexed correctly in text]**

References to the design table in text have now been correctly indexed as a reference to appendix B.

12. **Analyses hypotheses 2: why are two t-tests chosen? Would be more appropriate to use an ANOVA or a linear regression to compare results from these conditions on SSSEPs.**

Whilst an ANOVA might be more appropriate, a priori power analyses for an ANOVA suggest that given the effect size of d= .5, alpha of .05 and power at 80%, we would need an unreasonable number of participants in our sample size (128 participants), which is not possible to achieve. Therefore, we have decided to go with matched pairs one sided t tests, and to account for multiple comparisons within our analysis.

13. **Power analyses – Hypothesis 1: It is not clear where the first two effect sizes used to calculate the power came from**

This has been made clearer to the reader as can be seen in the text below, the preprint of Hansford et al (2022) can be found here **doi:** https://doi.org/10.1101/2023.01.18.524558*:*

Page 10, lines 439 -443: "Effect sizes are determined by research from Hansford et al (2022) using the subjective illusory experience questionnaire and hand-based resizing illusions which show an effect size of n^2 = .33 (converted to a Cohen's f = .73) when comparing all participants, and an effect size of n^2 = .35 (converted to a Cohen's f = .74) when looking at participants who experience an effective uni-modal visual illusion."

14. **Power analyses – hypotheses 2: the minimum effect size of interest quoted by Lakens is lower than this. From memory, d = .23, I believe. Also, is there a way of showing whether the chosen effect size is also relevant for EEG studies?**

The d = 0.5 comes from Cohen's small, medium and large effect sizes, with d = 0.5 being the medium effect size. We are using Cohen's d= 0.5 as this is the smallest effect size that we are interested in for a clinical sample. The rationale for using the smallest effect size of interest rather than effect sizes from previous literature comes from the Lakens 2014 paper cited, which advocates for using the smallest effect size of interest for your study rather than a specific effect size from previous literature, and we are using Cohens medium effect size in this instance.

15. **Does the power analysis for hypothesis 3 only relate to the chronic pain group?**

Yes, this is correct, and has been further clarified in the text, as can be seen below:

*Page 11, lines 466 – 468: "A priori power analysis using G*Power shows that for a Wilcoxon signed-rank test (matched pairs), with an effect size (dz) of 1, alpha of 0.02, and power at 90%, for a one tailed test with normal parent distribution, 15 chronic pain participants are needed in total."*

16. **Sample size (2.5) should be at the start of methods**

This has been moved to the start of the method section and is now section 2.1, with all subsequent sections renumbered as needed.

17. **When referencing a section, also include the number e.g. (see Power analysis in section 2.4)**

Numbers for referenced sections have been included and updated to match the new numbering order.

18. **Finally, there is room for much more concise writing throughout**

Overall writing has been made more concise with the help of comments made by yourself and other reviewers and can be seen in the manuscript as a whole.

Susanne Stoll

1. **Clarity of rationale/theory**

• Predictive coding and central sensitization theory are described as accounts explaining the cortical misrepresentation of the body that might occur in chronic pain patients. I interpret this misrepresentation as a distorted homunculus (altered topographic map). Yet, based on the provided info, it is not clear to me how either of these accounts can explain such a cortical misrepresentation. Moreover, w.r.t. predictive coding, some descriptions seem a little unclear, such as what is meant by a mismatch between experienced and actual painful sensory inputs.

These accounts describe the misrepresentation of the incoming pain signals within the body, with predictive coding mentioning that a lack of updating of expectations about incoming painful sensory inputs results in mismatches between experienced and painful sensory inputs (this has been clarified in text referring back to the symptomatic and radiographic pain mentioned previously in the paragraph, as can be seen below). Whereas central sensitisation theory refers to the amplification and distortion of pain within the body. This has been clarified in the text below to show that these theories relate to the misrepresentations of incoming pain signals, not specifically the cortical misrepresentations of the affected limbs. Text has also been minimised relating to the points raised by reviewers about the length of the introduction and more clarity needed:

Page 1, lines 49 – 63: "It has been suggested that in individuals with chronic pain there may be a cortical misrepresentation of the body and its incoming pain signals, along with perceptual size dysfunctions of affected limbs, which underpin their persistent pain (Boesch et al., 2016). There is often reported a lack of concordance between radiographic (physical damage) and symptomatic pain (Szebenyi et al., 2006; Felson, 2005). This highlights the likelihood of a cortical misrepresentation driving pain rather than structural damage, explaining why surgical interventions to treat structural elements of pain could be ineffective. Theories underlying this cortical misrepresentation are the predictive coding account (Friston, 2008) and the central sensitisation theory (Arendt-Nielsen & Graven-Nielsen, 2003; Arendt-Nielsen et al., 2010). Predictive coding posits that any mismatch between predicted and actual sensory inputs, such as the difference between peripheral signals and symptomatic pain, generates prediction errors. A possible lack of updating of top-down expectations in chronic pain individuals, could lead to constant mismatches between symptomatic and radiographic painful sensory inputs. Central sensitisation theory, however, refers to the central nervous system changing, distorting, or amplifying pain in a way that no longer reflects the peripheral input from the body, leading to pain becoming an illusory perception (Woolf, 2011)."

Additional text has also been added in the introduction regarding the somatosensory blurring and magnifying theories, with a clear link back to the predictive coding and central sensitisation theories mentioned above:

Page 2, lines 112 – 126: "There are two main theories underlying the analgesia seen during resizing illusions, firstly the somatosensory blurring hypothesis, which posits that the cortical representation of a painful body part is blurred, and that viewing the body part sharpens this representation. This is supported through findings in healthy participants, where visual analgesia has been found following experimentally induced pain (Haggard et al., 2013). The other theory is from Gilpin et al. (2015),

which showed that participants with arthritis make smaller hand judgements compared to healthy participants, and posited that this could be influencing pain, as when stretching the hands, the pain was reduced. Therefore, it could be that increasing the cortical representation through magnifying the affected body part, reduced their pain. Both theories predict that the cortical misrepresentations mentioned previously through the predictive coding and central sensitisation accounts, therefore occur at the somatosensory cortex, with both theories predicting different neural changes regarding the experience of pain. Specifically, somatosensory blurring hypothesis predicts a larger, more diffuse representation of the painful body part that would be reduced (sharpened) with the illusions, whereas the magnification theory would predict a shrunken representation of the painful body part that would be enlarged following illusory stretching."

- **I have difficulty appreciating the rationale underlying the proposed study. 1. Based on the provided info, it is not entirely clear to me what the "benefit" of inducing SSEPs is, as they seem to be primarily sensitive to temporal features of (multi)sensory stimulation (e.g. synchrony). 2. The intro repeatedly talks about a cortical misrepresentation of the body in chronic pain patients, that resizing illusions might affect this misrepresentation and that this might be related to a reduction in pain observed in these individuals. It is not clear to me how the proposed study can shed light onto these aspects (not least because the authors do not seem to be interested in group comparisons).**

Regarding the benefit of using SSEPS, further clarity has been added to the introduction regarding the SSEPs and how these are an index of the cortical response as can be seen below:

Page 3, lines 144 – 148: "Looking specifically at research into somatosensory cortex modulation using steady-state evoked potentials, low-level somatosensory responses have been induced directly using vibrations of a known frequency applied to a body part. These generate a frequency-locked steady-state evoked potential detectable at the scalp using EEG (Snyder, 1992; Tobimatsu et al., 1999), and are an index of the cortical response to a stimulus."

Further clarification has also been added to the introduction regarding the cortical misrepresentation as can be seen below:

Page 1, lines 49 – 50: "It has been suggested that in individuals with chronic pain there may be a cortical misrepresentation of the body and its incoming somatosensory signals, including pain."

We have also added that given significant findings in the proposed analyses, exploratory analysis regarding group comparisons could be merited, as can be seen in the text below for hypothesis 1 and 2, respectively:

Page 9, lines 396 – 398: "If the positive control is successful in these analyses, then there would be scope to run exploratory analyses between the healthy and chronic pain groups regarding their subjective illusory experiences."

Page 10, lines 419 – 420: "If differences are seen in these analyses, then there would be scope to run exploratory analyses between the healthy and chronic pain groups."

- **Given that the intro does not explicitly state what experimental conditions will be included, the paragraph outlining the hypotheses at the end of the intro is not easy to understand. It also refers to "subjective embodiment theories". It is unclear to me what exactly is meant here and how these theories are related to the hypotheses.**

Additional text has been added to the final paragraph of the introduction to explicitly mention the experimental conditions and add clarity to the hypotheses. The text about the subjective embodiment theories has also been removed as it was unclear what it was referring to, to make for a more streamlined concluding paragraph as can be seen below:

Page 4, lines 180 – 205: "Using different sensory manipulations of finger resizing illusions, in addition to using an electromagnetic solenoid stimulator, this study aims to investigate subjective illusory experience and SSEP responses in both healthy and chronic pain patients, to better understand the relationship between body ownership illusions and experiences of chronic pain from subjective experience and cortical representation perspectives. To test this, different resizing illusions consisting of multisensory (visuotactile) stretching (MS), unimodal-visual stretching (UV), a non-illusion control condition without tactile input (NI), and a non-illusion control condition with tactile input (NIT) will be used to assess alternate aspects of illusory resizing manipulations and their related effects on SSEP response. In line with previous findings regarding effective UV conditions (Hansford et al., 2022), subjective questionnaire data will be used to identify individuals who experience an effective UV condition, and these participant's SSEP data will then be analysed. The first hypothesis, acting as a positive control (1), is that (1a) there will be a greater illusory experience, measured via a subjective illusory experience questionnaire, in the MS condition compared to the non-illusion conditions in the healthy group. There will also be (1b) a greater illusory experience, measured via a subjective illusory experience questionnaire, in the MS condition compared to the non-illusion conditions in the chronic pain group. The main experimental hypothesis for this study is that (2) there will be a significant difference in SSEP response when comparing (2a) MS visuotactile illusory resizing to non-illusions, and when comparing (2b) effective UV illusory resizing to non-illusions in the healthy group. There will also be a significant difference in SSEP response when comparing (2c) MS visuotactile illusory resizing to non-illusions, and when comparing (2d) effective UV illusory resizing to non-illusions in the chronic pain group. Also, (2e) there will be a significant difference when comparing healthy to chronic pain participant's baseline NI SSEP responses. The final hypothesis is that (3) we expect to find a subjective reduction in pain, measured via a 21-point numeric rating scale, comparing before and after scores for (3a) MS and (3b) UV conditions, whilst we expect (3c) no reduction of pain following the NI condition, nor (3d) a reduction of pain following the NIT condition."

2. Presentation of pilot data/accessibility and status of prior work

- **In the intro (+ section on power analysis), pilot data are mentioned, which are not presented in 3.3 Pilot data. Similarly, prior submitted work (Hansford et al., 2022) that is mentioned in the intro and used to calculate effect sizes does not seem to be available as a preprint (at least no preprint including doi has been referenced), making it hard to evaluate these aspects. Moreover, Hansford et al. (2022) is cited as "in prep" in the text, but listed as "submitted" in the reference section, creating confusion about the status of this article.**

Pilot data regarding the illusion strength in synchronous and asynchronous conditions has now been added as appendix C, as can be seen below, and text has been updated in the introduction to reflect this. The DOI for Hansford et al. has now been added to the reference list and has been cited as the preprint throughout. The corresponding reference with DOI is below:

Page 15: "Hansford, K. J., Baker, D. H., McKenzie, K. J., & Preston, C. E. (2023). Distinct Neural Signatures of Multimodal Resizing Illusions: Implications for Chronic Pain Treatment. bioRxiv, 2023-01. https://doi.org/10.1101/2023.01.18.524558."

Page 2, lines 88 – 91: "…previous pilot data using the same experimental set up as the current study, has shown trends towards greater illusory experience in healthy and chronic pain participants during synchronous visuotactile manipulations compared to asynchronous (mismatching visuotactile manipulation) control conditions (Appendix C)."

Page 24: ***"Appendix C:***

Pilot data regarding the efficacy of the illusion for both healthy and chronic pain patients undergoing synchronous and asynchronous illusory resizing of the index finger can be seen in figure C1. 16 participants (7 chronic pain, 9 healthy) had either synchronous or asynchronous multimodal manipulations delivered first in a random order, and were then given the other condition, after which all participants were given an illusion scale. Findings showed that across all participants, no significant difference in illusion experience between the synchronous and asynchronous conditions, t(30) = -0.40, p = 0.69, however as can be seen in figure C1, despite the small sample size, illusion strength was seen to be greater in the synchronous condition compared to the asynchronous condition.

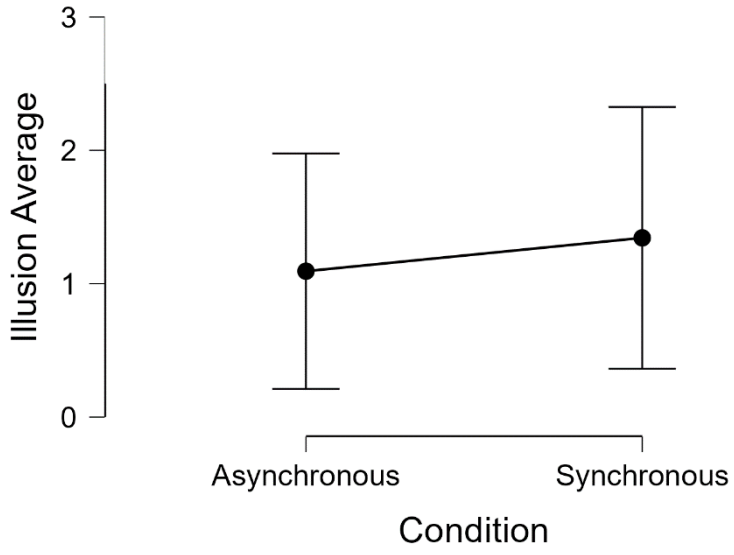

Figure C1. Pilot data from Chronic Pain and Healthy Participants Undergoing Synchronous and Asynchronous Illusory Finger Resizing."

3. Comprehensiveness/clarity/appropriateness/redundancy of inclusion/exclusion criteria

- **Although mentioned nowhere, I think the authors intend to pull the finger of the right hand (?). "Chronic pain in (one of) the fingers of the right hand" thus seems to be an unspecified inclusion criterion.**

You are correct that the intention is to use the right hand to remove lateralisation effects in the SSEPs, therefore the inclusion criteria have been updated as can be seen below to reflect this:

Page 5, lines 229 – 232: "Inclusion Criteria: Right-handed, over 18 years of age, no older than 75 years of age.

**Chronic pain participant specific inclusion criteria: must have a diagnosed chronic pain condition involving current hand-based pain in the right hand, hand-based pain present on day of testing."*

- **It is mentioned that data will be excluded if less than 50% of the experiment has been completed or more than 50% of the electrodes need removal. It is unclear what exactly this refers to (e.g. data set of a single participant and trials?) and why these criteria are sensible. For instance, if only 50% have been completed, the number of trials in the different experimental conditions per participant might be vastly different.**

This is a valid point made about the number of trials being potentially different, and therefore the raw data exclusion criteria have been updated as can be seen below:

Page 5, lines 249 – 250: "Less than 100% of the experiment completed by a participant, more than 50% of electrodes needing removal from EEG data."

- **It is mentioned that participants will be matched based on gender, but that the demographic survey will assess sex.**

This has been updated in the sample characteristics section (2.2) to say that we will be matching based on sex as can be seen below:

Page 5, lines 218 – 219: "Chronic pain and healthy participants will be matched based on sex, age and handedness, creating a matched pairs design experiment."

- **In light of the interindividual differences reported in the intro, I do not understand why the authors intend to include chronic pain patients irrespective of chronic pain condition.**

We will now exclude participants who have been diagnosed with complex regional pain syndrome, as this is the only chronic pain condition where we have alternate predictions about the nature of the effect of resizing illusions on pain. This has been updated in the exclusion criteria as can be seen below:

*Page 5, lines 242 – 246: "*Chronic pain participant specific exclusion criteria: Diagnosed with Complex Regional Pain Syndrome, no restrictions apply regarding any medication the participant might be taking. (Complex Regional Pain Syndrome is excluded as a chronic pain condition here, due to research showing increasing pain after stretching illusions (Moseley et al., 2006)."*

- **Healthy and chronic pain participants share many inclusion/exclusion criteria, which are outlined twice (once for each group). This makes it a little hard to appreciate the differences.**

The inclusion / exclusion criteria section has been reworked to better show the differences between the groups, as can be seen below:

Page 5, lines 229 – 246:

"Inclusion Criteria: Right-handed, over 18 years of age, no older than 75 years of age.

> **Chronic pain participant specific inclusion criteria: must have a diagnosed chronic pain condition involving current hand-based pain in the right hand, hand-based pain present on day of testing.*

Exclusion Criteria: Prior knowledge or expectations about the research, a history of developmental, neurological or psychiatric disorders, history of drug or alcohol abuse, history of sleep disorders,

history of epilepsy, having visual abnormalities that cannot be corrected optically (i.e. with glasses), or being under 18 years of age, or over 75 years of age.

> **Healthy participant specific exclusion criteria: a history of chronic pain conditions, operations or procedures that could damage peripheral nerve pathways in the hands, current experiences of pain or more than 4 hours of consistent pain experienced in the preceding week.*
> **Chronic pain participant specific exclusion criteria: Diagnosed with Complex Regional Pain Syndrome, no restrictions apply regarding any medication the participant might be taking. (Complex Regional Pain Syndrome is excluded as a chronic pain condition here, due to research showing increasing pain after stretching illusions (Moseley et al., 2006)."*

- **To transform the sampling characteristics into a "Participants" section, I think further points need to be covered, such as consent, ethics, and Declaration of Helsinki.**

An additional paragraph has been added to section 2.2, now termed the "Participants" section, as can be seen below:

Page 5, lines 213 – 217: **"***Ethical approval for this research was gained from the Department of Psychology, University of York (ethics application code 950), in line with the Declaration of Helsinki. Informed consent from each participant will be gained prior to the start of any experimental set up, and participants will be instructed that they can withdraw their participation at any time during or after completion of the experiment*."

4. Clarity/appropriateness/detail of experimental procedure

Questionnaires – handedness and pain

- **The Waterloo Handedness Questionnaire has not been described and it is unclear how exactly right-handedness (inclusion criterion) will be determined.**

Description of the revised Waterloo Handedness Questionnaire has been added, along with the criteria for determining right-handedness as can be seen below:

Page 6, lines 253 – 260: "All participants will fill out a demographic survey, asking their age and sex, and will be asked to complete the revised Waterloo Handedness Questionnaire (WHQr) (Elias et al., 1998). The WHQr self-reported handedness questionnaire consists of 36 questions. The questions are answered on a 5-level, Likert scale to determine the degree of preferred hand use, with left always being -2, left usually being -1, equal use being 0, right usually being 1 and right always being 2. The sum of the total WHQr score can then be used to categorise a respondent as left-handed (score of -24 or less), mixed handed (score of -23 to +23), or right-handed (score of +24 or higher). Only participants who are categorised as right-handed will continue participation."

- **It is unclear what participants will be asked when they have to indicate their pain score. Given that they will also be asked which finger is most painful, I would assume they have to rate the level of pain for this finger?**

Clarity has been added regarding how the participants will indicate their pain score as can be seen below:

Page 6, lines 260 – 264: "Clinical participants will then be asked questions regarding what chronic pain condition they are diagnosed with, how long it has been since diagnosis, what medications (if

any) they are taking, and their pain score on that day for their digit in the most pain using a 21-point numeric rating scale (NRS) (0 = no pain at all; 20 = most severe pain imaginable)."

Digit manipulation

- **It is stated that if multiple fingers are equally painful, the one that is easier to manipulate will be chosen and that more than one digit can be manipulated if needed. It is not clear to me why some fingers will be easier to manipulate, or multiple fingers need to be manipulated and how the experimenter figures this out. I think this level of flexibility has the potential to create fundamental differences between some participants.**

The option for more than one digit being manipulated has been removed, as although this is possible and could be useful for longer term treatment, for the purposes of the proposed study, we agree that one digit being manipulated is preferable for consistency between participants. The commentary about choosing the easiest to manipulate digit has also been amended to whichever digit the participant chooses, to remove any researcher bias. All changes can be seen in the text below:

Page 6, lines 278 – 286: "Participants will then be seated behind the augmented reality system (Figure 1) and instructed to place their hand onto the black felt fabric within the augmented reality system. Chronic pain participants will be asked which digit is in the most pain and will be asked to place this digit outstretched onto the felt. If multiple digits are equally painful, the digit that the participant chooses as their preference will be used."

Augmented reality system

- **The text refers to 2 white dots on the felt. In Figure 1, however, I see 4 white dots. It is also stated that the 2 dots will guide where the hand will be placed. How?**

The text has been updated to reflect that it is 2 white dots for each hand, and further clarity has been added to the guide for where their hands should be placed, as can be seen in the text below:

Page 7, lines 286 – 289: "Healthy participants will be asked to outstretch a digit that has been matched to that of a chronic pain participant. There will be two white dots for each hand on the felt and participants will be instructed to place their hand between these two dots."

- **When describing the augmented reality system, I think the camera needs to be mentioned too, so that it is 100% clear where the video/image comes from.**

Further information about the camera has been added to the text about the augmented reality system, see the new text below:

Page 7, lines 289 – 293: "Participants will be instructed to view their hand's image in the mirror (the real hand will be hidden from view) throughout the experiment. A camera placed underneath the mirror on the felt base will be used to deliver a live feed video of the participants hands to the computer screen at the top of the augmented reality system, which will show in the mirror reflection to the participants."

Experimental conditions

- **The experimental conditions are hard to picture. Adding static visualizations/videos might be very helpful.**

Figure 2 has been added to show visualisation of the different conditions as can be seen below:

Page 7, lines 312 – 316: "Visualisation of all conditions can be seen in Figure 2.

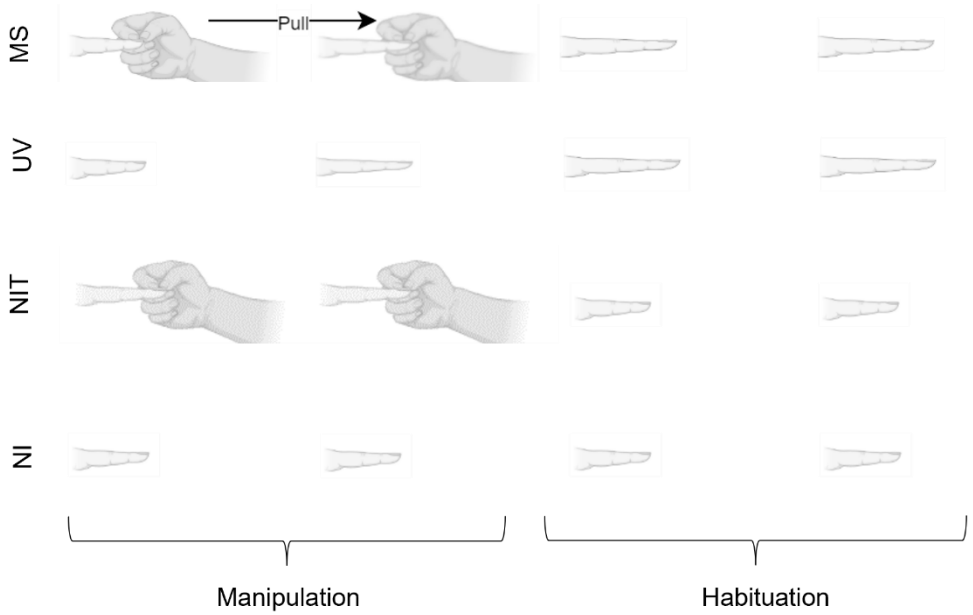

Figure 2. Infographic of Experimental Conditions. MS = Multisensory, UV = Unimodal Visual, NIT = Non-Illusion Tactile, NI = Non-Illusion. Manipulation phase (2.4 seconds) is where experimenter creates illusion, habituation phase (2.4 seconds) is where participants are free to move their finger. Arrow denotes the direction of the experimenter's action."

- **The way the text (including design table) refers to the experimental conditions is confusing. In some places, it refers to multiple multisensory or unisensory illusory conditions, although there is just one of each. Similarly, there are 2 non-illusory control conditions and in several places the text refers to a single non-illusory control condition.**

Text throughout the manuscript has been updated to provide clarity regarding the experimental conditions, such as in the hypotheses paragraph at the end of the introduction where the non-illusion conditions are now pluralised, see the text below for an example:

Page 4, lines 190 – 200: "The first hypothesis, acting as a positive control (1), is that (1a) there will be a greater illusory experience, measured via a subjective illusory experience questionnaire, in the multisensory condition compared to the non-illusion conditions in both groups. Also, regarding subjective illusory experience, we hypothesise that (1b) there will be a greater illusory experience in the unimodal visual condition compared to the non-illusion conditions for those who experience the unimodal illusion. The main experimental hypothesis for this study is that (2) there will be a significant difference in SSEP response when comparing (2a) multisensory visuotactile illusory resizing to non-illusions, and when comparing (2b) unimodal visual illusory resizing to non-illusions in the Healthy Group. There will also be a significant difference in SSEP response when comparing (2c) multisensory visuotactile illusory resizing to non-illusions, and when comparing (2d) unimodal visual illusory resizing to non-illusions in the Chronic Pain Group."

- **For the multisensory illusory condition, the text states that the hand will be stretched and for the unisensory illusory condition that the finger will be stretched. For the non-illusion control condition with tactile input, it is not specified what happens to the visual input.**

Condition NI is described as a non-illusion control condition without tactile input, which seems somewhat incorrect as there will always be tactile input delivered via a tactile stimulator.

The text relating to the description of the conditions has been updated to add clarity as can be seen below:

Page 7, lines 295 – 309: "Participants will undergo 4 conditions: multisensory stretching (MS), unimodal-visual stretching (UV), a non-illusion control condition without tactile input (NI), and a non-illusion control condition with tactile input (NIT). There will be vibrotactile stimulation to the finger in all conditions, but only tactile input of the researcher touching the participants hand in the conditions where this is mentioned. Each trial will last 2.4 seconds for the manipulation phase, where the finger will be stretched by 60 pixels in UV and MS conditions, followed by a further 2.4 second habituation phase in which participants can view and move their augmented finger before the screen goes dark, indicating that the next trial can start. MS conditions consist of the researcher touching and pulling the participant's finger as the participant views their finger stretching in a congruent manner. UV conditions consist of the participants viewing their finger stretch without any experimenter manipulation. The NI condition provides no visual or touching tactile manipulations to the finger. The second NI control condition will involve no visual input of the finger stretching but will include tactile input of the experimenter's hand touching the participant's finger, but without pulling."

- **It is stated that an indicative box tells the experimenter whether to pull the finger or apply no manipulation. This description seems incomplete, as the experimenter also needs to be informed when they have to touch the finger without pulling (condition NIT). This in turn means that in most cases the experimenter knows which condition will be presented. As such, it seems a little odd that the text states the experimenter will be "blinded".**

This is a valid point, and the box will be blue to indicate the researcher to pull the finger, and white to indicate just touching the finger. This will mean that the experimenter is indeed no longer blinded to which condition is being applied, so this has been updated. Updated text can be seen below:

Page 8, lines 322 - "The experiment will be programmed in, and the conditions randomised using MATLAB R2017a and the experimenter will be informed of whether to pull the finger or to apply no manipulation via an indicative box displayed on the screen out of the participant's view. If the box is blue, this will indicate a need to pull the finger, if it is white it will indicate a need to touch the finger."

- **In the section on the experimental procedure, it is stated that the level of pain will be assessed before and after each illusory condition (in chronic pain patients), whereas the section on preprocessing steps refers to pain data that have been assessed for 3 conditions including one of the control conditions (MS, UV, and NI).**

The text in the preprocessing section has been updated to refer to all conditions for pain measures, as can be seen below:

Page 9, lines 376 – 377: "Pain data will be averaged for pre and post all experimental conditions, resulting in 8 averages per participant."

Questionnaire – illusory experience

- **It is stated that at the end of the experiment, all conditions will be presented again in an ordered fashion. It is unclear what the exact order will be and why "ordered" is to be preferred over "randomized".**

This has been updated to be a randomised fashion, to remove order effects for the subjective illusory questionnaire scales, as can be seen below:

*Page 8, lines 344 – 345: "*Finally, each condition will be presented once in *a randomised fashion*, after which, the participant will be asked to complete the subjective illusory experience questionnaire for each trial."

- **The illusory experience questionnaire includes 6 questions, some of which refer to the finger and some of which to the whole hand. Shouldn't they all refer to the finger? Moreover, how will participants respond to the 6 questions? The section on planned analyses suggests a scale will be used.**

The questions have been updated so that they all refer to fingers rather than the hand, and a note on the scale to be used has also been added, as can be seen below:

Page 8, lines 349 – 358: "It felt like my finger was really stretching" / "It felt like the finger I saw was part of my body", two relate to disownership: "It felt like the finger I saw no longer belonged to me" / "It felt like the finger I saw was no longer part of my body", and two are control questions: "It felt as if my finger had disappeared" / "It felt as if I might have had an extra finger" (all questions will be directed towards the participants manipulated finger). Control questions are included to assess participant compliance effects, whilst disownership questions are included to assess if the potential analgesia from the illusions results from a disownership of the body part, or from subjective embodiment of said body part (McCabe, 2011). A visual analogue scale from 0 – 100 will be used for each statement, with 0 being strongly disagree, 50 being neutral and 100 being strongly agree."

- **There are 2 control questions to assess compliance effects. Will they be used to remove data sets?**

The control questions will be used to create an index by subtracting the median control scores from both the median illusion and median disownership scores. The 50 point threshold will be maintained for showing experience of the illusion and disownership after the control scores have been subtracted. This has been added to the preprocessing steps within the manuscript as can be seen below:

*Page 9, lines 371 – 376: "*Regarding questionnaire data, scores for both illusion experience questions will be combined to give median scores, along with both disownership questions and both control questions, resulting in 3 median scores per trial per participant. The median control scores will be used to create an index of the illusion and disownership scores by subtracting the median control score from the median illusion and median disownership scores, in line with previous research doing similarly (Matsumiya, 2021; Kilteni & Ehrsson, 2017; Kalckert & Ehrsson, 2012)."

Unaddressed details that seem necessary to ensure consistency and replicability

How long does a trial for a given condition last? How will the end of a trial be registered? What is the percentage of visual finger stretching?

Page 7, lines 299 – 302: "*Each trial will last 2.4 seconds for the manipulation phase,* where the finger will be stretched by 60 pixels (2.1 centimetres) in unimodal and multisensory conditions, followed by *a further 2.4 second habituation phase in which participants can view and move their augmented finger before the screen goes dark, indicating that the next trial can start*."

How is it ensured that the temporal structure of different trials is the same and that actual pulling and visual stretching are synchronized?

Page8, lines 326 – 327: "The researcher will use a button press to dictate the start of the manipulation, and will start pulling the finger, when needed, synchronously within the 2.4 second manipulation phase."

How long does it take until the video is augmented?

The following has been added to the Experimental Procedure section:

Page 7, lines 293 – 294: "There is a delay of 170ms in the video processing pipeline from the camera image to the augmented video image."

What happens if the experimenter makes an error (e.g. forgets to pull)?

The experimenter has undergone extensive training to prevent errors like this happening, and in previous similar experiments, this error has not occurred, however we have added a clause to the text regarding the possibility as can be seen below:

Page 8, lines 327 – 330: "If the experimenter forgets to pull the finger on a multisensory condition, or mistakenly pulls the finger in a control trial, then this will be noted during the experiment, and that trial will be removed from analysis."

Is the augmented reality system an item that has been purchased (if so from where?) or is it self-built? What are the basic specs of the augmented reality system (e.g., size of screen; resolution of screen, size of felt, overall height, type/brand of camera, etc.)?

Page 6, lines 279 – 284: "Within the self-built system there is a 1920 x 1080 Spedal Webcam Wide Angle Camera situated in the middle of the central area, away from the participant's view. 26cms above the felt base of this central area, there is a mirror, which is placed 26cms below a 1920 x 1200 resolution screen, with a width of 52cms. This screen is 54cms from the base of the system, and the base of the system is 82cms from the ground."

Will all fingers be outstretched or just the one that is being pulled?

As mentioned in the text, only the digit that is the most painful, or is a matched digit will be outstretched.

Page 7, lines 284 – 287: "Chronic pain participants will be asked which digit is in the most pain and will be asked to place this digit outstretched onto the felt. If multiple digits are equally painful, the digit that the participant chooses as their preference will be used. Healthy participants will be asked to outstretch a digit that has been matched to that of a chronic pain participant."

What are the other parameters of the sine wave (e.g., amplitude)?

Further parameters of the tactor have been added as can be seen below:

Page 8, lines 333 – 335: "The tactor is driven at 50% of the maximum (i.e. a peak input voltage of 3V) using a 26Hz sine-wave, and delivers a peak force of 0.18N."

Where will the experimenter be seated and how exactly will they pull a given digit?

Page 8, lines 317 – 319: "The experimenter will be seated opposite the participant, the other side of the augmented reality machine and will pull the digit by holding onto the distal interphalangeal joint and gently pulling the finger whilst the participant keeps their hand in place."

What software will be used to program the experiment?

Page 87, lines 322 - 323: **"**The experiment will be programmed in, and the conditions randomised using MATLAB R2017a…"

Who will operate the tablet?

Page 8, lines 344 – 346: "…the participant will be asked to complete the subjective illusory experience questionnaire for each trial using the Samsung Galaxy Tab A6 tablet via a questionnaire on Qualtrics (Qualtrics, Provo, UT)."

How will the pain rating be assessed (verbally/paper/tablet)?

Page 8, lines 337 – 340: "Clinical participants will be asked before each illusory manipulation and then again immediately after each manipulation to rate their pain on the 21-point NRS, which will be a verbal report that the experimenter will input onto a Samsung Galaxy A6 Tablet."

Will participants be seated in a shielded chamber?

Participants will not be seated in a shielded chamber. Our EEG system has electrodes with active shielding that reduce the impact of external noise, so a Faraday cage (or similar) is not required to obtain clean data.

Will there be ocular/reference electrodes?

Page 6, lines 274 – 275: "The whole head average will be used as a reference."

Will participants' finger be continuously stimulated (without any breaks) throughout a given block and also when they have to complete the final illusory experience questionnaire? Will the stimulator be attached to the finger that will be pulled? Etc.

There will be continuous stimulation during each trial, therefore not between trials or during the subjective illusory experience questionnaire, or pain ratings for the chronic pain participants. Text has been altered as can be seen below:

Page 8, lines 335 – 337: "The electromagnetic solenoid stimulator will be attached to the participant's finger that is outstretched and will receive the manipulations, between the knuckle and the first finger joint, using a black Velcro strip and will give continuous stimulation for the duration of each trial."

5. Appropriateness/clarity/detail of preprocessing steps

- **The outlined preprocessing of the EEG data seems very minimal. For instance, I would assume the EEG data need to also undergo segmentation, artifact correction, averaging, and a Fourier transform (some of this is mentioned in passing in 3.3. Pilot data). Moreover, what software will be used/what are the electrodes of interest/what about lateralization? It seems critical to specify these things, not least because they (might) determine the number of tests being performed as part of hypothesis 2.**

The preprocessing section regarding the EEG data has been updated to include mention of the software to be used, and the electrodes to be analysed as can be seen below:

Page 9, lines 365 – 370: "Using MATLAB r2019a and EEGlab, a 50Hz notch filter will first be applied to the raw EEG data for all electrodes, which will then be analysed to show standard errors for each electrode for each participant. The top 5% of standard errors will be calculated resulting in a standard error threshold. Any electrode with a standard error above this threshold, or with a value of 0, will be removed from analysis. Where a participant has over 50% of their electrodes over the standard error threshold, data will be removed."

There is no need for the EEG data to undergo artifact correction as you mention, as mentioned by Figueira et al's (2022) paper on the FreqTag toolbox, only a Fourier transform is typically needed for this type of EEG data. There will be averaging, as is mentioned in the analyses section of hypothesis 2 (2.4.2.2), which has had detail added about the use of Fourier transform as can be seen below:

Page 10, lines 408 – 420: "After pre-processing steps as mentioned in section 2.4.1 are taken, analysis of EEG data will first involve importing the waveforms from MATLAB into R, and then using R to take a Fourier transform for each waveform across all remaining electrodes, to obtain individual results per participant. These will then be averaged across all participants to give overall results, before running a dependent samples t test (two-sided) comparing MS to NI and one comparing UV to NI in the healthy group, along with a dependent samples t test (two-sided) comparing MS to NI and one comparing UV to NI in the chronic pain group. We will finally run a dependent samples t test (two-sided) comparing the healthy group baseline NI data to the chronic pain group's baseline NI data. The dependent variable will be SSSEP amplitude in µV, whilst the independent variable will be the different manipulations given in each comparison condition. No additional filtering or denoising steps will be applied to the EEG data, in line with Figueira et al.'s (2022) report that only a Fourier transform is typically needed for this type of EEG data. If differences are seen in these analyses, then there would be scope to run exploratory analyses between the healthy and chronic pain groups. Based on the pilot data in Figure 3, we would expect to see activation most pronounced over mid-frontal distributions, covering F1 and FC1 electrodes."

- **W.r.t. the selection of EEG data, it is not clear to me what is meant by calculating "the top 5% of standard errors" and using this as a threshold (maybe the 95th percentile?) and why it is sensible to use this threshold and keep a data set even if 50% of the electrodes have been removed.**

The criteria of having over 50% of the electrodes being removed resulting in removal of the entire dataset has been used previously (Hansford et al., 2022), therefore, to maintain consistency the same criteria has been used in the proposed study. The text regarding the EEG data preprocessing has been updated to add clarity about the top 5% of standard errors, as can be seen below:

Page 9, lines 365 – 370: "Using MATLAB r2019a and EEGlab, a 50Hz notch filter will first be applied to the raw EEG data for all electrodes, which will then be analysed to show standard errors for each electrode for each participant. The top 5% of standard errors will be calculated resulting in a standard error threshold. Any electrode with a standard error above this threshold, or with a value of 0, will be removed from analysis. Where a participant has over 50% of their electrodes over the standard error threshold, data will be removed."

- **The outlined preprocessing of the pain and illusory experience data involves averaging. In the section on planned analyses, it is mentioned that Likert scales will be used for the illusory experience questionnaire (rendering the data ordinal) and that the pain data are ordinal. By calculating averages, these data are treated as "interval". Besides, it is unclear to me why it is sensible to first average the pain data, then group according to chronic pain condition, and**

then average again. This procedure might give a lot of weight to a few individuals with a certain pain condition and does not seem to fit in with the proposed tests for hypothesis 3.

The questionnaire has now been changed to use a visual analogue scale from 0-100, which is described in the text at the end of the Experimental Procedure (2.3) section:

Page 8, lines 357 – 358: "A visual analogue scale from 0 – 100 will be used for each statement, with 0 being strongly disagree, 50 being neutral and 100 being strongly agree."

The Preprocessing section (2.4.1) has also been updated to show that median scores will now be used in place of mean scores, and the text regarding the averaging by each chronic pain condition has been removed:

Page 9, lines 371 – 373: "Regarding questionnaire data, scores for both illusion experience questions will be averaged to give median scores, along with both disownership questions and both control questions, resulting in 3 median scores per trial per participant. The median control scores will be used to create an index of the illusion and disownership scores by subtracting the median control score from the median illusion and median disownership scores, in line with previous research doing similarly (Matsumiya, 2021; Kilteni & Ehrsson, 2017; Kalckert & Ehrsson, 2012). Pain data will be averaged for pre and post all experimental conditions, resulting in 8 averages per participant."

6. Clarity of stated hypotheses

- **Given that there are 2 non-illusion conditions, it is unclear what is meant by "non-illusion condition" when hypotheses 1+2 and associated analyses are outlined.**

The hypotheses paragraph has been updated to reflect both non-illusion conditions as can be seen below:

Page 4, lines 190 – 200: "The first hypothesis, acting as a positive control (1), is that (1a) there will be a greater illusory experience, measured via a subjective illusory experience questionnaire, in the multisensory condition compared to the non-illusion conditions in both groups. Also, regarding subjective illusory experience, we hypothesise that (1b) there will be a greater illusory experience in the unimodal visual condition compared to the non-illusion conditions for those who experience the unimodal illusion. The main experimental hypothesis for this study is that (2) there will be a significant difference in SSEP response when comparing (2a) multisensory visuotactile illusory resizing to non-illusions, and when comparing (2b) unimodal visual illusory resizing to non-illusions in the Healthy Group. There will also be a significant difference in SSEP response when comparing (2c) multisensory visuotactile illusory resizing to non-illusions, and when comparing (2d) unimodal visual illusory resizing to non-illusions in the Chronic Pain Group."

- **Hypothesis 1 appears underspecified. Just like for hypothesis 2, shouldn't the subhypotheses for hypothesis 1 be specified separately for healthy and chronic pain participants, resulting in 4 instead of 2 subhypotheses?**

The additional subhypotheses have been added for hypothesis 1 as can be seen in the text below. This change has been updated throughout the manuscript:

Page 4, lines 190 – 195: "The first hypothesis, acting as a positive control (1), is that (1a) there will be a greater illusory experience, measured via a subjective illusory experience questionnaire, in the multisensory condition compared to the non-illusion conditions in the Healthy Group. There will also be (1b) a greater illusory experience, measured via a subjective illusory experience questionnaire, in the multisensory condition compared to the non-illusion conditions in the Chronic Pain Group."

- **Hypothesis 3 seems underspecified. As far as I understand, it effectively consists of 2 subhypotheses (3a: reduction in pain [pre vs post] for the multisensory illusory condition; 3b: reduction in pain [pre vs post] for the unisensory illusory condition) instead of a single hypothesis.**

The text regarding hypothesis 3 has also been updated to split it into 2 subhypotheses regarding each condition as can be seen below. This change has been updated for the entire manuscript:

Page 4, lines 202 – 205: "The final hypothesis is that (3) we expect to find a subjective reduction in pain, measured via a 21-point numeric rating scale, comparing before and after scores for (3a) MS and (3b) UV conditions, whilst we expect (3c) no reduction of pain following the NI condition, nor (3d) a reduction of pain following the NIT condition*."*

- **W.r.t. hypothesis 3, it is also unclear to me why the authors do not wish to include a control condition – I think that would facilitate the interpretation of results.**

Two additional subhypotheses have been added regarding the control non-illusion conditions which can be seen below, and this change has been reflected throughout the manuscript:

*Page 4, lines 203 – 205: "…*whilst we expect (3c) no reduction of pain following the non-illusion condition, nor (3d) a reduction of pain following the non-illusion tactile condition."

7. Appropriateness/clarity of planned analyses

- **Hypothesis 1: It is not clear to me why a Friedman test should be performed given that all subhypotheses seem to relate to contrasts between specific experimental conditions. Similarly, it is unclear to me why the text states that 3 comparisons will be made. Which groups/conditions do these 3 comparisons involve?**

The analysis plan for hypothesis 1 has been changed to an ANOVA, with more explicit mention of the now 4 comparisons, as can be seen below:

Page 9, lines 389 – 392: "A one-way ANOVA will be run to compare the dependent variable of median illusion score from each independent condition. Given significant findings, post-hoc tests will be run, with Bonferroni correction for 4 comparisons (MS / NI conditions, UV / NI conditions) at an initial alpha of 0.05."

- **Hypothesis 1b: I think the authors might run into a double-dipping/regression to the mean problem (see e.g., Kriegeskorte et al., 2009; Stoll et al., 2022). This is because the illusory experience data from the unisensory illusion condition will be used for selection (subset of individuals experiencing the illusion in this condition) and selective analysis (comparison of illusory experience data in this condition to illusory experience data from a control condition for this subset). This renders this procedure circular and likely results in regression towards the mean (or variants thereof), i.e., statistical artifacts. One way to break this circularity would be to assess the unisensory condition twice in each individual, so that one data set can be used for selection and the other for selective analysis. However, due to the lack of detail on the scales assessing illusory experience, it is not clear to me why individuals need to be selected in the first place or why individuals with an average illusion score above 1 in the unimodal illusion condition should be selected. In any case, the selection of individuals seems to come with an uncertainty about the sample size – it seems unclear how the proposed study accounts for that.**

This has been addressed through stating that the subjective illusory data will be used to give a subset of participants who experience the unimodal illusion and then with this set of participants only their EEG data will be analysed, not their questionnaire data. As such, hypotheses 1c and 1d regarding the prediction that there will be greater illusory experience in the unimodal condition compared to the non-illusion conditions for those who experience the unimodal illusion, have been removed. These changes have been reflected in the text below:

Introduction change:

Page 4, lines 188 – 190: "In line with previous findings regarding effective UV conditions (Hansford et al., 2022), subjective questionnaire data will be used to identify individuals who experience an effective UV condition, and these participant's SSEP data will then be analysed."

Planned analysis Hypothesis 1 change:

Page 9, lines 392 – 396: "Subjective data will also be used to identify participants who effectively experience the unimodal visual condition where participants will be included in further EEG analysis if their median illusion scores on the subjective illusory questionnaire scale for the unimodal-visual condition are greater than 50, in line with previous research using mean subjective embodiment scales (Carey et al., 2019), which will indicate experience of the illusion."

Planned analysis Hypothesis 2 change:

Page 10, lines 401 – 405: "*There will be a significant difference in SSEP response when comparing (2a) multisensory visuotactile illusory resizing to non-illusion, and when comparing (2b) effective unimodal visual illusory resizing to non-illusion in the healthy group. There will also be a significant difference in SSEP response when comparing (2c) multisensory visuotactile illusory resizing to non-illusion, and when comparing (2d) effective unimodal visual illusory resizing to non-illusion in the chronic pain group."*

- **It would be good to always specify what software will be used, whether a test is one-sided or two-sided and how multiple comparisons will be dealt with, which I think is also relevant for power analyses.**

Additions of which software to use and whether one-sided or two-sided tests will be used have been added where this was missing as can be seen in the text below (multiple comparisons are mentioned when needed):

Page 9, lines 389 – 390: "A one-way ANOVA will be run to compare the dependent variable of median illusion score from each independent condition."

Page 10, lines 408 – 414: "…analysis of EEG data will first involve importing the waveforms from MATLAB into R, and then using R to take a Fourier transform for each waveform across all remaining electrodes, to obtain individual results per participant. These will then be averaged across all participants to give overall results, before running a dependent samples t test (two-sided) comparing MS to NI and one comparing UV to NI in the healthy group, along with a dependent samples t test (two-sided) comparing MS to NI and one comparing UV to NI in the chronic pain group."

Page 10, lines 431 – 432: "…non-parametric Wilcoxon signed rank tests (one-sided) will be used to compare the dependent variable of mean pain scores before and after each independent condition."

*Page 11, lines 449 – 450: "A priori power analysis using G*Power for the smallest effect size of interest (f = .73) shows that for a repeated measures, within factors one way ANOVA …"*

*Page 11, lines 466 – 467: "A priori power analysis using G*Power shows that for a Wilcoxen signed-rank test (two-sided, matched pairs) …"*

7. Appropriateness/clarity of reported effect sizes and power analyses

- **Hypothesis 1: I think it needs to be ensured that the ultimate contrast of interest is sufficiently powered (see also my comments on planned analyses).**

The section on effect sizes for hypothesis 1 has been updated to show the specific study that the previous effect sizes have come from, and shows that needed sample size of 6 participants for each group, which will be achieved with the overall sample size needed being 34 participants for each group as seen in the power analysis for hypothesis 2. The text amended for hypothesis 1's power analysis can be seen below:

Page 10, lines 439 – 451: "Effect sizes are determined by research from Hansford et al (2022) using the subjective illusory experience questionnaire and comparing MS and UV hand-based resizing illusions which show an effect size of $n^2 = .33$ (converted to a Cohen's $f = .73$) when comparing all participants, and an effect size of $n^2 = .35$ (converted to a Cohen's $f = .74$) when looking at participants who experience an effective uni-modal visual illusion. Additional effect size information comes from a visual capture study using a subjective embodiment questionnaire and visual and tactile manipulations to a mannequin body (Carey et al., 2019), showing an effect size of $r = .64$ (converted to a Cohen's $f = .83$). An effect size of .73 was used for hypothesis 1a and .74 was used for hypothesis 1b to adhere to the lower end of previous effect sizes.

*A priori power analysis using G*Power for the smallest effect size of interest ($f = .73$) shows that for a repeated measures, within factors one way ANOVA, with an effect size (f) of 0.73, alpha of 0.05, power at 80% and 2 groups with three measurements, 6 participants are needed for each group."*

- **Hypothesis 2: A power analysis for a one-sided paired t-test has been performed. Hypothesis 2, however, is expressed in a way that seems to suggest a two-sided test.**

This is correct that it should have been a two-sided test, so the text for hypothesis 2 has now been updated to give a new sample size, which has been reflected throughout the manuscript:

Page 11, lines, 454 – 459: "This is the first study to investigate illusory finger stretching using SSEPs, so appropriate effect size estimates are not available. We therefore conducted power calculations based on a smallest effect size of interest (Lakens, 2014) of d = 0.5 (a medium effect, see Cohen, 1988).

*A priori power analysis using G*Power shows that for a matched pairs two-sided t test, with an effect size of d = .5, alpha of 0.05, power at 80%, a total sample size of 34 participants is needed for each participant group."*

- **The descriptions related to the effect sizes for hypothesis 1 are not easy to understand because it seems unclear what they refer to. For instance, what does it mean that hand-based resizing illusions show a certain effect size? What has been compared here?**

Clarity has been added to the effect sizes for hypothesis 1 as can be seen below:

Page 10, lines 439 – 443: "Effect sizes are determined by research from Hansford et al (2022) using the subjective illusory experience questionnaire and comparing multisensory and unimodal-visual finger-based resizing illusions using the same finger stretching illusions and the same equipment, which show an effect size of $n^2 = .33$ (converted to a Cohen's $f = .73$) when comparing all participants,

and an effect size of $n^2 = .35$ (converted to a Cohen's $f = .74$) when looking at participants who experience an effective uni-modal visual illusion."

Similarly, for hypothesis 3, it is not clear whether the same pain scale/multisensory resizing illusion have been used in prior work and what the effect size in Preston et al. (2020) amounts to.

Clarity has been added to the effect size rationale for hypothesis 3 as can be seen below:

Page 11, lines 462 – 465: "Effect size is determined using those listed in previous research using the 21-point numeric pain rating scale (Preston et al., 2020) and from previous pilot data using the same multisensory resizing illusions for analgesic effect, finding post illusion pain scores to be significantly lower than pre illusion scores (t(10)=3.32, p = .008, d = 1.0)."

8. Clarity of presented pilot data

- **The text states that a pilot study has been conducted to determine the ideal frequency. However, as far as I understand, it has only been tested how well a frequency of 26 Hz works (?).**

The text regarding the pilot data has been updated to reflect that 26Hz was tested to see if we could record a reliable SSEP at 26Hz using our equipment, as can be seen below:

Page 11, lines 473 – 480: "Previous literature states that the ideal vibration frequency to use to elicit somatosensory steady state evoked potentials (SSSEPs) ranges from 26-27Hz (Muller et al., 2001; Muller-Putz et al., 2001; Breitweiser et al., 2016; Pokorny et al., 2016; Snyder, 1992). Due to resizing illusions often manipulating the index finger, and previous studies using the index finger supporting around 26Hz as an optimal frequency (Muller-Putz et al., 2001; Breitweiser et al., 2016; Pokorny et al., 2016), it was hypothesised that 26Hz would elicit a dependable SSSEP. Therefore, we ran a pilot study to check that our setup and equipment can reliably elicit and record a SSSEP at 26Hz, using the resizing illusion and EEG."

- **It is not entirely clear to me what the exact data basis for the amplitude vs frequency graph is (currently Figure 3, although I think it should read Figure 2). For instance, have the data been averaged across all electrodes and all conditions and are they based on healthy participants? Have the data been cleaned?**

The pilot data text has been updated to show that the data are from healthy participants and that the data are across all conditions. No data cleaning took place, and this has now been mentioned in text. Due to the addition of a new figure 2, The figure has remained Figure 3:

Page 12, lines 481 – 487: "Pilot data was collected for 3 Healthy participants. Participants underwent the same experimental protocol as mentioned in the "Experimental Procedure" section, minus the subjective illusory experience and pain rating scales. No additional filtering or denoising steps were applied to the EEG data, in line with Figueira et al.'s (2022) report that only a Fourier transform is typically needed for this type of EEG data. A Fourier transform was calculated for each waveform at each electrode for all conditions, and then averaged across repetition to obtain individual results. These were then averaged across all 3 participants to give the result seen in Figure 3."

- **It is stated that for the pilot data, no illusion experience questions have been assessed. I thus wonder whether it is ensured that the illusory conditions still work properly when adding tactile input at a frequency of 26 Hz.**

This has been addressed in the additional pilot study run, which shows that the illusory conditions still show illusory experience with the addition of the vibrotactile input, as can be seen below and has been added to section 3 Pilot Data:

Page 12, lines 501 – 511: "*As can be seen, there is a greater subjective experience of the resizing illusion, indexed by participant's illusion score, in both experimental conditions (UV average = 64.25; MS average = 67.88) compared to both control conditions (NI average = 32.38; NIT average = 24.13). Scores below 50 are indicative of disagreement of experience of the illusion, whilst a score of 50 is a neutral option regarding the illusion experience, and scores above 50 are indicative of agreement of experiencing the illusion. This therefore shows that the addition of the vibrotactile stimulation does not remove the experience of the resizing illusion and can therefore be used in the proposed study to elicit SSEPs without affecting the subjective illusory experience of the resizing illusion.*

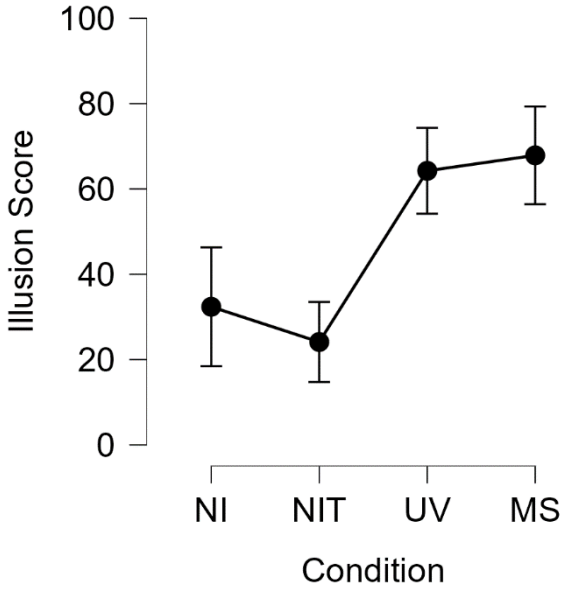

Figure 4. Averaged Illusion score for each condition. Error bars represent standard errors."

7. Consistent usage of abbreviations and capitalization

• Once introduced, it would be good to consistently use capitalizations/abbreviations, such as Healthy Group, OA, SSSEP, or MS (instead of e.g. going back to spelled-out versions or using variants of abbreviations, such as UVS or UV). For clarity, I think even common abbreviations (EEG) should be spelled out upon first usage.

All abbreviations and capitalisations have been made consistent throughout the manuscript, and EEG has been introduced at first use.

8. Clarity/comprehensiveness of appendices

• Appendix A: The single figure here is labeled "Figure 2". I think the figure numbering typically starts over in each appendix (Figure A1 in this case). The text here refers to 90 participants, but the main text to 94 participants.

The figure has been relabelled as A1, and the text has been updated to 68 (new sample size) participants as can be seen below:

Page 19: "Doubling this to c.100 participants tested, to account for participant drop-out for the current study needing 68 participants…"

- **Appendix B: I think it would be good to give the design table a title and table number (i.e., Table B1) and use the table number in the main text. I wonder why the design table does not outline the rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis and the theory that could be shown wrong by the outcomes. I also think the statements about achieved power can be removed.**

The table has been given a title and table number and this has been reflected when referenced in text throughout the manuscript. The manuscript has been updated throughout to show that the new sensitivity for confirming or disconfirming hypotheses is now at the typically used alpha of 0.05 and power at 80%, which has been updated in the design table also. Statements about achieved power have been removed, and an additional column for the theory that can be shown wrong by the outcomes has been added.