

I have now received three reviews of your manuscript, “The Phenomenological Control Scale: Measuring the capacity for creating illusory nonvolition, hallucination and delusion.” I also independently read the manuscript before consulting these reviews.

We thank the editor for the opportunity to respond to reviewers. We address the editor’s comments here and reviewer comments below.

The reviewers see value in the study, although they have many comments about the theoretical framing of the study, the reporting of the methods, the description of the results, and the interpretation of the results.

1. *One major set of objections from the reviewers stems from the embedded theoretical assumptions that frame the research. The reviewers both disagree with some of the assumptions and also feel that these assumptions were not made explicit. As Reviewer 1 put it, “I think the authors need to be more explicit about the theoretical assumptions and commitments associated with phenomenological control including the overarching assumption that responsiveness to suggestion indexes phenomenological control, which is never actually justified.” A few of the objections may stem from misunderstanding of what you are proposing about this hypothetical construct of phenomenological control.*

We had failed to clarify what we mean by phenomenological control. It is not intended to describe a new concept but rather as a new term to describe existing concepts. The phenomena referred to have been studied extensively under the label “hypnotizability”. Our introduction of this term is motivated by the scientifically misleading connotations of the term “hypnosis” and the lack of success that previous attempts to move beyond the hypnotic context have had. Barber introduced the term “creative imagination” to describe response to imaginative suggestion away from the hypnotic context. This may have failed because it fails to capture the range hypnotic experience (e.g. analgesia) – it clashes with our intuitions about imagination. We have added clarification of this point.

A related issue is that our motivation has not been stated clearly enough. We have added the following text to the first page, explicitly describing our motivation and goals:

“Although it is a new term, phenomenological control is not a new concept. The capacity we describe is most commonly referred to as “hypnotizability”, which in turn has its roots in the 18th and 19th century concept of mesmerism (Pintar & Lynn, 2008). The present work is motivated by our belief that, at least in recent years, the dominance of the hypnotic context has had two major consequences. Imaginative suggestion researchers immersed in the hypnotic tradition can overlook that the hypnotic context is not required and focus exclusively on this single context and its associated themes. Hypnosis is a mildly disreputable term, a situation which is likely to arise from the many scientifically inaccurate myths associated with the hypnotic context (e.g., trance or being controlled by an authority figure; see (Lynn et al., n.d.)). This may have led to the avoidance of the subject by many researchers, who may have little awareness of the copious evidence for stable trait differences in response to imaginative suggestion. We argue that focusing on ‘hypnosis’ has acted as a barrier walling off imaginative suggestion from the rest of psychology and resulted in a lack of attention to other contexts in which phenomenological control may be operative (which may be many and varied; for example, speaking in tongues, channelling spirits or responding in

scientific experiments; Dienes, Lush, et al., 2020) . In this paper, we present a scale for measuring the capacity of phenomenological control. Our aim is to accelerate research into imaginative suggestion away from just the hypnotic context (which comes with its cultural myths) and to raise awareness of imaginative suggestion research outside the field. Imaginative suggestion effects may confound experimental psychology measures (see Corneille & Lush, 2021; Dienes, Lush, et al., 2020; Lush, Dienes, Seth, et al., 2021; Lush et al., 2020); trait differences in response to imaginative suggestion (as measured by imaginative suggestion scales) are therefore potentially relevant to any researcher interested in reports of change in experience.”

2. *In some of their comments, the reviewers object to what they see as a rosy interpretation of the quantitative findings. They suggest that the viability of the phenomenological control construct is at the least highly uncertain and some of the reviewers’ view of the evidence in the broader literature is that it actually rules against the validity of phenomenological control and thus they object to the manner in which the interesting topics in here are being pursued. As Reviewer 1 put it, “I think the paper would be greatly strengthened if you provided some validation research demonstrating that responsiveness to suggestion measures what you call “phenomenological control””.*

This relates to concerns raised by R1 and R3 that responses may reflect compliance and bias. We emphasise that all we have done here is to remove the word “hypnosis” and references to relaxation (and trance etc) out of a hypnosis scale. All these concerns apply equally to hypnosis scales, but each of the reviewers seems to consider hypnotizability scales to reflect something other than compliance and bias. For example, each review contains reference to high hypnotizables. A high hypnotizable is someone who has a high score on a hypnosis scale.

We agree that the question of how to interpret reports of experience is critical to psychology. Our wider project of investigating phenomenological control in experimental measures (e.g., rubber hand illusion etc) is motivated by this question. How much can we trust our participants not to fake responses? How reliable is introspection? This is more ground than we can cover in this paper, but we have added references to a recent manuscript which addresses these issues with reference to demand characteristics and phenomenological control (Corneille & Lush, 2021).

There are several reasons we do not consider hypnosis scales to entirely be driven by conscious faking or imagination (we address these in detail in individual responses to reviewers). We consider the phenomenological control scale to be similar to the SWASH scale, because scores are comparable and highly correlated across scales, and because the scale items and measures are virtually identical across the scales. Relatedly, past research has also repeatedly found that responses to the same suggestions are highly correlated with or without a hypnotic induction; we now refer to relevant reviews. The method section has been expanded to contain more detail regarding the similarities and differences between the scales We have also elaborated on our position in the introduction and discussion. We have addressed the relationship between hypnotizability and phenomenological control in earlier theoretical papers (Dienes, Lush, et al., 2020; Dienes, Palfi, et al., 2020) and have now added further reference to these papers in the manuscript (see individual responses).

3. *I think if a revised manuscript were to express better the degree of uncertainty around the hypothesized construct and the different views that may be warranted about them, with substantial revision of the reporting of the study, and thoughtful attention to the other comments of the reviewers, your report could be acceptable for publication in Collabra:Psychology. By softening rhetoric where it may over-claim, such as the phrase “the role of any induction or altered state of consciousness can be approached in a scientifically neutral way”, and making in some cases substantive changes in response to the reviewers’ points, I think the manuscript will be much improved.*

We have added caveats to the manuscript throughout, and deleted strong claims (see response to reviewers). E.g., “the role of any induction or altered state of consciousness can be approached in a scientifically neutral way” has been deleted.

We have softened rhetoric and revised the manuscript extensively. In summary, this scale has been developed primarily for use in an ongoing project to investigate relationships between trait response to imaginative suggestion and reports of change in experience in other effects studied by psychological scientists. We are aware that responses on imaginative suggestion scales (including hypnotisability scales) will include a variety of different phenomena (e.g., faking, imagination and phenomenological control). Our proposal is that reports of changes in experience which are attributed to other mechanisms are confounded by the mechanisms which drive stable trait response to imaginative suggestion. So far our correlational results are consistent with this prediction: for hearing sounds when watching silent video (Lush, Dienes, Seth, et al., 2021), feeling touch when watching video of people being touched, feeling pain when watching video of people in pain and of experiencing that a fake hand is one’s own (Lush et al., 2020). Participants who are inclined to fake response to imaginative suggestion may also fake response to the procedures which produce these effects. Participants who intentionally imagine a required experience in response to imaginative suggestion may also intentionally imagine the required experience in these other procedures. Participants who respond with phenomenological control to imaginative suggestion may also respond with phenomenological control in these procedures. See (Corneille & Lush, 2021 for extended discussion). If a participant reports strong experience of an illusory mosquito when this is verbally suggested, the same mechanisms behind that report may be in play when that same participant reports strong experience of visually evoked auditory response, mirror touch synaesthesia, mirror pain, or the rubber hand illusion. We of course agree that it is necessary to disentangle the various possible mechanisms driving report of experiential change. This is, however, orthogonal to our current aims. We provide the norms of this scale in order to aid interpretation of our current and forthcoming results which employ this scale and as a simple and relatively efficient tool (the PC scale is based on the SWASH, which was a revision of the earlier WSGC intended to greatly reduce the time required to screen large numbers of participants) for researchers who wish to measure response to imaginative suggestion as it is generally measured in hypnosis research (minus the hypnosis). We have revised the manuscript to clarify that we do not claim that a particular response on this scale (nor on hypnosis scales) may not reflect genuine experience.

In the remainder of this letter, I provide a few comments from my own reading that I didn't notice were mentioned by the external reviewers:

4. *-There isn't a capsule summary of what the study is before the Methods section starts, and a resulting problem is that it takes a lot of reading through the Methods for a reader to discern what the main methodological features of the study are*

Added to p.10 "Here we present norms for a phenomenological control scale which is closely matched to an existing hypnosis scale (SWASH). Secondary analyses contrast scores for the phenomenological control and SWASH scales and investigate the stability of response to imaginative suggestion across the two contexts."

We have added a further summary at the start of the Discussion (p.29):

"While it is well established that response to imaginative suggestion does not require hypnosis, previous scales of imaginative suggestion outside the hypnotic context (e.g., Barber & Wilson, 1978) have failed to displace hypnosis as the primary context in which imaginative suggestion effects are studied. Because many of the cultural beliefs about hypnosis are scientifically inaccurate, there is a need to develop scales for measuring trait response to imaginative suggestion outside the hypnotic context."

5. *-I don't see what the interval was between test and retest, but I might have missed it.*

Added to p.9: "(M= 3.9 weeks, SD = 0.7)."

6. *-the manuscript argues that "imaginative suggestibility" and "creative imagination" is no good, but "phenomenological control" probably has its own problems in leading some readers to think that you mean something that maybe you don't, like that a lot of people can hallucinate at will (unless you do mean that?). There probably is no term that is perfect, I don't know what you can do except work harder to clarify what you mean and add qualifier or hedge words that help convey what you don't mean at various points in the manuscript.*

We have elaborated on the definition of the term in the manuscript, again with reference to our other work which has a more detailed treatment (see response to reviewers).

With regard to whether or not participants can 'hallucinate', we agree that all we show here is that people will, to varying degrees (and according to a stable trait), report as though they can. We believe this is relevant for the interpretation of other experiences studied in psychological science. That aside, we agree with consensus in the hypnosis and imaginative suggestion community that at least some participants who report anomalous experience in response to imaginative suggestion are genuinely having experience (we now refer to an earlier paper reviewing this evidence with regard to phenomenological control throughout the text; Dienes, Palfi, & Lush, 2020). We certainly are not claiming that a lot of people can vividly hallucinate visual or auditory experience (strong response to music and negative visual hallucination suggestions, for example, is rare). In addition we clarify that suggested 'hallucinations' may be very different to

other forms of hallucination. To illustrate this point, consider the theory of phenomenological control we work with - the cold control theory of hypnosis (though note that the scale is not restricted to this theory) . This posits that hypnotic hallucination is intentional imagination which is experienced as unintentional, to meet strategic goals. How compelling a particular experience is therefore depends on a number of factors, including the participant's ability to intentionally imagine that experience and their ability to not experience that intentional act as intentional. Take for example a suggestion for a sour taste (which around 50% of participants pass). Imagining sucking on a lemon can make people salivate, and experience a 'taste' of sourness. This experience may well be rather different to the experience of sucking on a lemon. According to cold control theory, all that is required for this to count as a phenomenological control or (hypnotic response) is that, given an appropriate context (and appropriate goals), participants misrepresent their intention to imagine a sour taste so that the experience is of something happening to them rather than something they have intentionally done.

Regarding whether or not people can do this sort of thing at will, hypnosis research has shown that response to imaginative suggestion is, while reported as feeling involuntary, voluntary and strategically goal directed. The term 'control' reflects this near consensus. Phenomenological just refers here to experience. So 'phenomenological control' just describes the ability to control experience to meet the demands of imaginative suggestions in accordance with situational goals.

Our proposed term is not perfect. One issue is that while we intend to refer to the "control" that a participant has, the context of hypnosis carries misleading ideas about the hypnotist being in control, which could lead to misinterpretations of what we mean by control.

We have added the following to p.6:

"In labelling the capacity as a trait for phenomenological control we imply that it is not a disposition for being controlled by others (such as may be implied by suggestibility)."

7. *-When you say "measures of changes in experience in psychological phenomena (e.g., the rubber hand illusion, mirror synaesthesia and visually evoked auditory illusions) are substantially predicted by response to direct imaginative suggestion in the hypnotic context", would it be better to say measures of changes in reported experience, not changes in experience? I think this point is related to some of the difficulties in knowing how strong your claim is about phenomenological control or what exactly is included in the construct.*

Absolutely- we cannot assume experience for any given response (anybody could be faking, but for a number of reasons it is unlikely that everybody is). This sentence should say "reported experience". We have clarified this throughout the manuscript.

8. *-Many would object to this sentence and call it a fragment as it contains no verb: "Thus, the potential usefulness of a phenomenological control scale."*

Sentence deleted.

We thank the reviewers for their attention to our manuscript.

Reviewer 1

This is an interesting line of research and a potentially valuable contribution to the literature but I have a number of concerns with this paper that I think need to be addressed before this paper is suitable for publication. Most notably, I think the authors need to be more explicit about the theoretical assumptions and commitments associated with phenomenological control including the overarching assumption that responsiveness to suggestion indexes phenomenological control, which is never actually justified. Perhaps more importantly, I also think the authors need to be more objective about the limitations of this scale as it has a number of significant psychometric shortcomings - the authors should be more transparent about these.

Phenomenological control & suggestion

In their recent work, the authors have advanced the notion of phenomenological control. Although this is an interesting phenomenon that is worthy of study there are some problematic elements pertaining to the use of this term.

- 1. The authors assume that suggestibility (responsiveness to suggestions) is a form of phenomenological control (the capacity to generate experiences to meet one's expectancies) and really just present this as a fact that doesn't require justification. However, at present, this is really a hypothesis.*

See point A above. Phenomenological control is merely a new term for existing phenomena. See (Dienes, Lush, et al., 2020) for detailed discussion.

- 2. The central claim is that you're study the capacity to generate experiences to meet your expectancies and thus expectancies figure prominently here. Indeed, expectancies do predict suggestibility on conventional hypnosis scales (e.g., Lynn et al., 2008, Oxford handbook of hypnosis). However, these data are correlational and sometimes the effects are actually not especially impressive (e.g., Benham et al., 2006, J Per Soc Psych). Critically, multiple studies have demonstrated that expectancy manipulations (which successfully increase expectancies) do not produce corresponding changes in suggestibility (Benham et al., 1998, J Pers Soc Psych; Lifshitz et al., 2012, Conscious Cognit). Although expectations play a role in response to suggestion, these and other data clearly indicate that expectation is unlikely to be a core mechanistic variable and this poses significant challenges for response to suggestion as a form of phenomenological control.*

We don't claim expectancies (particularly as measured in any particular way) are the only important feature of responding. It is indisputable that expectancies are involved in imaginative suggestion. Direct suggestions tell people what to expect. We argue that response to direct verbal suggestion and to indirect suggestion are both examples of phenomenological control. Note our preferred theory, cold control, explicitly states expectations are not the "core mechanistic variable"; so while we do not wish to beg cold control, which is just a particular theory, it should now be obvious we do not beg response expectancy theory either. We also do not claim for or against expectancy

manipulations producing changes in trait phenomenological control (or “suggestibility”).

Added to p.34: “Note that we do not argue that expectancies are the sole determinant of response (see e.g. Dienes & Perner, 2007, for counter-arguments to expectancies being the sole psychological determinant to response). Rather, we point out that suggestions inform people of what they should experience (or what they should expect to experience). This is true whether the suggestion is indirect and non-verbal (e.g., repetitive passes of a mesmerist’s hands or iron rod) or direct and verbal (e.g., the suggestion that one will have a particular experience as delivered by a ‘hypnotist’).

3. *I think the paper would be greatly strengthened if you provided some validation research demonstrating that responsiveness to suggestion measures what you call “phenomenological control”, rather than just assuming this. The authors maintain that phenomenological control indexes both “context-general direct and indirect suggestions” but this scale, like other hypnosis scales, is really measuring direct verbal suggestibility (Oakley et al., 2021, Conscious Cognit). For example, measures of indirect suggestibility have been shown to not reliably correlate with measures of direct suggestibility (e.g., hypnosis scales) (Polczyk, 2016, Curr Iss Pers Psych). This is relevant for other sections where you assume that direct and implicit suggestions tap the same domain.*

One key issue is that non-significant results (as reported in the cited Polczyk paper) do not in themselves show anything. However, our own theoretical preferences imply that not all responses to suggestion reflect the same processes; for example, direct effects of expectations are different from intentionally creating mental states without awareness of those intentions. But that does not rule against a distinction between direct and indirect suggestion, which is only a difference between whether the experimenter explicitly gave the suggestion or implied it; and so long as the subject responds psychologically in the same way, then those different external conditions are more or less irrelevant. Thus, our argument is that the division between direct suggestibility and indirect suggestibility is misguided. We have clarified our position in the manuscript.

Added to p5: “Similarly, a focus on ‘hypnosis’ has led to an emphasis on direct verbal suggestion in the study of imaginative suggestion effects, and even arguments that direct verbal suggestion and indirect suggestion are unrelated (e.g., see non-significant correlations between direct and indirect suggestion response reported by (Polczyk, 2016)), an argument which overlooks that Mesmerism (and therefore hypnosis) began as a non-verbal indirect suggestion effect (Hammond, 2013). This focus may also have led to a lack of attention of the role response to imaginative suggestion may play in other contexts. If we accept that people can respond to both direct and indirect suggestions via the same mechanisms, and if we accept that suggestion effects can be generated in contexts other than Mesmerism or hypnosis, it is possible that participant hypothesis awareness arising from demand characteristics (cues which inform participants of the aims of an experiment) may act as indirect suggestion effects and drive experience in experiments (Kirsch & Council, 1989; Lush, Dienes, Seth, et al., 2021; Lush et al., 2020; Michael et al., 2012).”

4. *“Social compliance (which is distinct from response to imaginative suggestion)”*

This is inaccurate. Suggestibility correlates with standardised measures of compliance (Polczyk & Pasek, 2006, Int J Clin Exp Hypn) - the magnitude of these effects is similar to what the authors found with the rubber hand illusion, what they called a “substantial” relationship. In addition, many participants report responding to suggestions via basic compliant responding (e.g., Bowers, 1981, Int J Clin Exp Hypn; Bowers et al., 1988, Int J Clin Exp Hypn). Compliance is almost certainly an element of suggestibility as indexed by direct verbal suggestibility scales, albeit not the primary mechanism.

We do not dispute that social compliance can be involved in response to imaginative suggestion, rather that response is not merely social compliance. We have added the following to clarify our position.

Edit on p.6: “Social compliance is a distinct concept to response to imaginative suggestion, e.g. Coe et al., 1973; Moore, 1964; Tasso et al., 2020; but see Polczyk & Pasek, 2006). It cannot, however, be ruled out that any given response on a hypnosis or phenomenological control scale (or indeed any subjective report scale) may reflect response due to social compliance such as faking or imagination rather than phenomenological control.”

5. *The authors should more explicitly define suggestion. Their example: “suggestion (e.g. “Did you think of trying this?”)” references how the term is used in the general public but this is different from how it’s used in the experimental/clinical literature where it’s typically understood as a communication for an involuntary response (“your pain will go down”). Individuals who score highly on a typical suggestion scale are unlikely to be more responsive to “did you think of trying this?” so this seems irrelevant. Please more clearly define suggestion.*

Our example was not to define a suggestion but to show the word does not have negative connotations, and the example succeeds for that purpose. The added definition should clarify.

Added to p.7: “We define the sort of suggestion that is relevant for phenomenological control as a communication to experience a counter-factual state of affairs as real, for example a voluntary movement as involuntary, or imagination as perceptual.”

Edit on p.6 “(see Dell, 2021 for further arguments against the suitability of the term “suggestible” to describe the ability to change experience in hypnosis). Note that our rejection of “suggestibility” does not extend to use of the term “suggestion”, which does not have the same connotations as “suggestible”; one can successfully respond to a suggestion (e.g. “Did you think of trying this?”) without being suggestible.”

6. *“which phenomenological control can surreptitiously bring about “
All of the studies you’ve done to date have been correlational - please avoid causal language.*

Changed to “may”

7. *“how people can strategically (although not knowingly) alter their conscious experience”*
Provide evidence that they don’t know, especially given the high compliance rates (Bowers et al., 1988).

We have revised the manuscript to clarify our position (see other points). For any given subjective response, we do not know if people are faking etc. However, we follow consensus in hypnosis research that at least some responses are not attributable to faking etc. This scale differs from a hypnosis scale only in the presentation of the procedure as hypnotic (the hypnotic context is avoided). We therefore interpret responses as we would interpret a hypnosis scale (and as hypnosis scales are routinely interpreted by hypnosis researchers – the reviewer refers to high hypnotisables and presumably does not believe they are all faking, but high hypnotisables are identified by the use of hypnosis scales, for which any given result may be faked). For reasons we provide at length elsewhere (Dienes, Lush, et al., 2020), we do not consider all response to imaginative suggestion faking. We presume that, although some responses will be driven by compliance, this is unlikely to fully account for imaginative suggestion phenomena.

Added to p.34 “Like most measures of hypnotisability, the experience of involuntariness is assumed for the PCS. Future scales would benefit from the addition of an involuntariness measure for each item (note that subjective report is susceptible to faking in response to imaginative suggestion, as in any situation in which demand characteristics are not controlled; see (Corneille & Lush, 2021).

8. *Context*
“in a scientifically neutral way “
But the scale assumes a theoretical framework (cold control) and thus is not scientifically neutral. To have this be completely neutral, you’d have to refer to imagination-mediated suggestion.

“scientifically neutral” removed. The scale does not assume cold control (see below).

9. *“By contrast the phrase “the capacity for phenomenological control” does not presume anything on this matter, one way or the other.”*
It does carry with it hidden assumptions regarding the underlying mechanisms (e.g., the whole model is premised on the authors’ cold control theory of suggestion/hypnosis) and thus, similar to “intentional binding” is not a theory-neutral descriptive term, especially given the way it’s defined (viz. expectation). It’s fine to have a hypothesis-driven term but I think this just needs to be more explicitly stated and not claim that the term itself is theory-neutral.

The term is not restricted to cold control theory, but applies to any theory of hypnosis in which participant response is not genuinely involuntary and involves genuine change in experience. An two earlier manuscripts addresses the relationship between cold control theory and phenomenological control in detail ((Dienes, Lush, et al., 2020; Dienes, Palfi, et al., 2020) The only theoretical claims for phenomenological control are that it involves phenomenology (rather than just faking) and control (rather than being genuinely involuntary).

We have clarified this on p.32: “The term phenomenological control is consistent with any theory of response to imaginative suggestion which posits that response is under control (as opposed to reflex), and any theory which accepts there is experiential change in response to imaginative suggestion. While we favour the theory that phenomenological control involves voluntary acts which are experienced as involuntary ((Dienes, 2012; Dienes, Lush, et al., 2020; Dienes, Palfi, et al., 2020)), the term is in no way limited to this ‘cold control’ theory alone.”

*10. The motivation to move beyond the hypnotic context is understandable but it's important to emphasise that you're introducing a new context (phenomenological control) to participants and this introduces other confounds and factors into the mix. To be sure, I think it's probably less problematic than a hypnotic context but it still may shape your results in various ways. I think this needs to be more readily acknowledged. "...a non-hypnotic scale which better matches the contextual expectancies of other experiments and avoids the hypnotic context is potentially better suited for such investigation."
This is a great point however this is still occurring in the context of an "imagination" task, which is very dissimilar from a lot of experimental work and carries with it various assumptions/biases/etc.*

We agree with the reviewer.

Added to P.30: “Of course, no experimental situation is ever free of context. The context of phenomenological control is however, relatively simple compared to hypnosis and arguably therefore introduces relatively minimal confounding contextual effects.”

*11. Imagination
I found references to imagination to be confusing or problematic in places. "While the term "imaginative suggestion" accurately conveys the nature of scale items..."
Many would contest this as evidence for a role of imagination in response to suggestion is weak and not all suggestions even reference imagery.*

The word ‘imagination’ in no way implies the necessary use of visual imagery. Aphantasics, for example, are quite capable of imagining different situations

We have clarified our working definition of imagination on p.7:

“Imagination involves the construction of non-present or counter-factual states of affairs (for a discussion of different ways of imagining, see Currie and Ravenscroft, 2002; the word ‘imagination,’ for example, does not necessarily imply the use of visual imagery).”

*12. "Previous terminology also employs the term "imagination", which is also problematic."
But you're referring to your suggestions as imaginative (e.g., "the suggestions which follow as exercises in using imagination to alter subjective experience") and using imagination as a preamble, how is this not a contradiction?*

We described the term imagination as problematic in reference to the term “creative imagination” in motivating the term “phenomenological control”. The point is that a

reader unfamiliar with imaginative suggestion effects might have a different idea in mind to that intended when they hear of a “creative imagination scale”. We don’t have a general aversion to the word “imagination”.

Edit p.6 “Previous terminology has employed the term “imagination”, which has also been problematic in the contexts used. For example, “Creative Imagination”

13. *Magnitude of effects*

The authors are variable in how they describe their effects and I think they need to be more consistent in their language and more objective. In the introduction, they refer to correlations between suggestibility and the rubber hand illusion as “substantial” but the correlations are actually in the weak-moderate range (~.30).

Putting aside questions of the interpretation of standardized effect sizes (see Lush, Dienes, & Seth, 2021 for an argument justifying the interpretation of standardized correlation coefficients between rubber hand report and hypnotisability scores), we usually interpret raw effects in linear models throughout our work, because they are much more informative.

We judge the relation of the rubber hand to hypnotisability as “substantial” because of the RAW effect size in each case (see (Lush et al., 2020; Lush, Dienes, & Seth, 2021). For example, the degree of proprioceptive drift increase by 0.6 cm for each unit increase in hypnotic response (and the mean amount of drift was 1.1cm).

Added to p.8: “These relationships are substantial. For example, a 1 point increase in hypnotisability score (6 point scale) predicts reports of experience of “ownership” of a fake hand by 0.8 points (7 point scale) (Lush, Dienes, & Seth, 2021).”

14. *In the discussion of this paper, the authors also refer to correlations of around .5-.6 as “substantial”. ...” Later on, item-total correlations (which were around .48 for the SWASH) were described as “moderate” in the Discussion.”*

This is a matter of context. A .5 correlation between two measures requires different interpretation than a .5 item-total correlation.

15. *They also describe Cronbach’s alphas that are below commonly-used thresholds (.70) as substantial, when these are better described as “acceptable”.*

‘Substantial’ replaced with acceptable for references to Alpha.

16. *Current study*

The motivation for this study could be made clearer. It seems that it’s about presenting norms but it’s also about comparing the PCS vs. a hypnosis scale so it really has bearing on the impact of an induction (and a different preamble) on response to suggestion. Indeed, this secondary aim about comparing the scales is the first sentence of the discussion.

Capsule summary added (see response to editor)

17. *Methods*

Please justify the sample size for the two screenings.

Added to P.11 : “An opportunity sample of...”. And: “Retest screenings were advertised until there were no more responses (after 1 month).”

18. *Please clarify why both re-tests used the SWASH (hypnosis) scale. This means, of course, that you don't have test-retest for the PCs. Or is this a typo?*

This is an error – which has now been corrected:

“62 participants previously screened on the SWASH (49 female, 13 male; mean age = 19.2, SD = 1.7) returned for PC screening”

19. *Was the subjective scale exactly the same for each item or does the wording vary? Please clarify.*

Added to p14: “The suggestions are identical for the two scales except for the removal of references to relaxation or hypnosis in the PC script. E.g., “just relax..” removed from Hand Lowering; “and relax.... “ removed from Moving Hands Together; “Now relax, relax completely” removed from Mosquito; “and you just continue to relax... more and more relaxed.” removed from Taste; “and relax” removed from “Arm Rigidity”; “because of the relaxed state you are in”, “relax” and “Just relax, relax all over” removed from Arm Immobilization. “Now ... just sit back and enjoy being hypnotized.” removed from the music hallucination suggestion; “Just relax and become even more deeply hypnotized as you continue to breathe comfortably and effortlessly.” And “relax completely” removed from Negative Visual Hallucination”.

The SWASH script ends with a de-induction procedure during which the hypnotist counts down from twenty and participants are told that they will be fully awake when the count reaches one. During the de-induction, the amnesia and post-hypnotic suggestions are given (e.g., “you will have difficulty in remembering all the things I have told you and all the things you did or felt, since you closed your eyes”). There is no counting down de-induction for the PC scale and the amnesia suggestion and post-session suggestion are presented as further exercise “For the next exercise, feel you will have difficulty in remembering all the things I have told you and all the things you did or felt, since you closed your eyes.

The final section of audio provides information about the reporting procedure (see materials at OSF). Following this, participants were verbally instructed to rate either the degree to which they entered a hypnotic state or how absorbed they felt in their imagination for the hypnosis and PC conditions respectively, with response on a 0 – 5 “depth scale”. However, due to a programming error, these data were not recorded (see explanatory note at OSF). The Penn State Worry Questionnaire (Meyer et al., 1990) was presented following the screening. These data were for classroom use and were not analysed as part of this study.”

20. *The Materials section could be clearer, such as by explicitly describing that there were two distinct scales with a short paragraph for each. As is standard for Materials sections, more information is needed regarding the PCS and SWASH including the number of items,*

Added to p.11: “Response to two scales (SWASH and PC) were measured. Both consist of the same 10 items. There are four ‘ideomotor’ suggestions (suggestions for apparently involuntary movement: hand lowering (a suggested experience of a heavy object in the participants outstretched hand) and moving hands together (a suggested experience of a ‘magnetic’ force pulling the participants outstretched hands together). There are two ideomotor challenges (suggestions that the participant cannot move): arm rigidity (the arm is so rigid it cannot be bent) and arm immobilisation (the arm cannot be lifted from the participants lap). There is one post-session suggestion, in which participants are told they will press the space bar six times in a row but will not remember being told to do so. The remainder are cognitive-perceptual suggestions: experience of a mosquito (tactile or auditory experience), music (hearing happy birthday played), negative visual hallucination (shown a slide of three coloured balls and being told they can only see two balls), amnesia (that they can remember nothing of the session until that point), taste (suggestions for experiences of sweet and sour taste).”

21. *the nature of the audio files for delivery,*

Added: “44.1 kHz .wav files.”

22. *the anchors for the subjective scoring, details about the preamble, instructions, and so on. There is a reference to an ‘induction’ in the procedure - is this the hypnotic induction of the SWASH or is there an induction in the PCS?*

Added to p.13:

“The SWASH induction (adapted from the WSGC; see <https://osf.io/g72ae/> for the full script) is 862 words in length and establishes a hypnotic context through multiple references to entering a state of hypnosis and the presentation of a counting down relaxation procedure as hypnotic. For example, “I am about to help you to relax, and meanwhile I will give you some instructions that will help you to gradually enter a state of hypnosis. You can become hypnotized if you are willing to do what I tell you to, and if you concentrate on what I say.” And “Soon you will be deeply hypnotized, but you will have no trouble hearing me. You will remain deeply hypnotized until I tell you to awaken later on. Soon I shall begin to count from one to twenty. As I count, you will feel yourself going down further and further into a deeply relaxed, a deeply hypnotized state. At the end of the induction participants are told they are “now hypnotized”.

The PCS preamble (see <https://osf.io/pzmbw/> for the full script) removes all reference to relaxation, sleep, hypnosis or an altered state. The study is presented as an exercise in imagination rather than hypnosis.

E.g., “You will shortly be given some exercises in the use of your imagination to create certain experiences. The aim is to see how much you can control the way you experience some simple events, such as moving your hand. For example, first you will be asked to lower your hand, and imagine it is being pulled down by itself. Engage yourself in that imagination, until it really feels like it is being pulled down by itself. Focus on the sensations and on the imaginary situation so you can immerse yourself in that reality, as if, for example, you were getting carried away by the narrative of a film, as if it were real.” and “We will warm up with a simple exercise in imagination. Soon I shall begin to count from one to twenty. As I count, imagine yourself going down some steps...feel yourself stepping and see the steps .. experience them in every way you can”.

Reference to hypnosis was removed from the introduction of the report section: “the specific happenings which were suggested to you during the hypnotic procedure” in the SWASH is changed to “the experiences which were proposed to you during the procedure” in the PCS. The wording in the rest of this section was revised to avoid reference to suggestion: “the specific happenings which were suggested to you during the hypnotic procedure.” In the SWASH was replaced in the PCS with “each of the experiences which were proposed to you during the procedure” and “how strongly you experienced the effects of the suggestion.” was replaced with “how strongly you experienced each phenomenon.”. At the end of the preamble participants are told “Now we will start with our exercises.”

23. *What were the subjective scale anchors?*

Added to p.13:

“Subjective scale anchors are taken from the SWASH and differ for each item. For example, The anchors for taste are “No taste” and “Strong taste” and for “moving hands together”, in which participants are told they will experience a force pulling their hands together, the labels are “No force” and “Strong force”. See the response booklet at <https://osf.io/hqdnf/>.”

24. *Why were two subjective responses requested for two items and which ones? Why was this done? It seems peculiar to have extra scoring for only two items - this could be much clearer.*

Edited on p.16:

“However, two of the items (taste and post-session experience) involve two experiences, and therefore include two responses (see Lush et al., 2018). The taste suggestion consists of responses to two suggestions for sweet and for sour tastes. The final subjective response score for taste is the mean of the sweet and sour responses. The post-session (traditionally called post-hypnotic) suggestion suggests both that the participant will perform an action and also forget that this was suggested. The score for this item is the geometric mean of response to urge and amnesia statements, so that a subjective response for this item would be zero if either of the components of the suggestion did not generate a subjective response.”

25. *Please explicitly state in the Methods the number of days (M, SD) between the test and retest sessions - sorry if I missed this.*

Added to p.9: “(M= 3.9 weeks, SD = 0.7).”

26. *Analyses*

“to estimate the stability of phenomenological control capacities across the two screening procedures.”

Yes, but it should be clear that you didn’t administer the PCS both times (see above) - unless again if this was a typo.

Edited on p.18:

“across the two screening procedures which are presented in different contexts.”

27. *Results*

The scale differences (Tables 1 and 2) are mostly driven by ideomotor and motor items (except taste; the mosquito hallucination often functions as a motor item because it involves a motor response; Woody et al. 2005, Psych Assess). All of the cognitive items except taste, which is the easiest one, don't really differ between the two scales and posthypnotic suggestion is actually higher. This aligns with various spectral analyses showing that the different suggestions differentially target different underlying abilities. Of note here, compliance is more strongly related to the easy items (Polczyk & Basek, 2006, Int J Clin Exp Hypn) - increases in the PCS might be compliance effects.

First as a preliminary, as we have noted to the editor, we use the same scale items as the SWASH, which in turn are most of the items from the Waterloo scale of hypnotisability, excepting two because of possible negative effects (see (Cardeña & Terhune, 2009, who warn against negative effects of the age regression item)). Now, as for which items improve for the PC scale, note that according to sample means 8 out of 10 items scored higher for PC than SWASH (on both objective and subjective scales), including for hallucinations and challenge suggestions. Only 1 item scored lower on the subjective scale. We believe that no case can be made out of these data for specifically compliance playing a role in PC rather than SWASH.

28. *It looks like you've done Welch t-tests (based on the df) - this should be stated.*

“Welch” added to P.16

29. *Please comment on how the correlation between behavioural and subjective scales is greater for the SWASH than the PCS. The latter is clearly significantly lower. This suggests that the PCS is not tapping the classic suggestion effect to the same extent as the SWASH - this has important implications for the interpretation of this measure, particularly since you're saying it indexes changes in volition.*

The correlation between behavioural and subjective scales was numerically greater for SWASH than for PCS. For individual items, point biserial correlations were comparable with a few exceptions. E.g., the lower limit of PC scale music point-biserial correlation was greater than the upper limit for SWASH. To know if the “classic suggestion effect” (i.e. the experience of involuntariness) is stronger or weaker in PC than SWASH we would need to take involuntariness ratings. Nonetheless we do have some support for a stronger feeling of involuntariness in PC than SWASH, because for motor suggestions that is the key part of the subjective experience that we measure with subjective ratings; and, as the reviewer has just mentioned, on these items people score higher on PC than SWASH.

30. *The fact that the amnesia and posthypnotic suggestions do not correlate (behaviour-subjective association) indicates that these are poor items. This aligns with previous research, where they also have poor psychometric properties (e.g., Sadler & Woody, 2004, Int J Clin Exp Hypn; Acunzo & Terhune, 2021, Int J Clin Exp Hypn). I recommend removing these items and replacing them with better items.*

See response to point 32.

31. *“The upper limit of the 95% CI for the PC subjective scale alpha was lower than the lower limit for the SWASH, .68, 95% CI [.62, .74], but also showed substantial consistency.”*

“Subjective scale alpha was substantial for both PC and SWASH.”I’m sorry but this is inaccurate, given that the alpha was below a common threshold for what is considered “acceptable” for a psychometric instrument (alpha > .70), it’s clearly not “substantial”, especially given the other problems with the scale. At the very least, change this to “acceptable”. Also, specify that this is Cronbach’s alpha in the Discussion as some readers might find this unclear.

Cronbach’s added. Changed to acceptable.

32. *“Alpha did not show good reliability for the objective SWASH scale .49, 95% CI [.39, .57] ...”*

This was also the case with the original SWASH paper, which begs the question why this scale was not revised. The item/suggestion content of these scales should really be reconsidered.

We have abandoned ‘objective’ scale measurement. Regarding the conservative approach to scale development, there are two motivations. First the scale was not revised because we anticipated that changing anything other than the basic hypnotic context would be seen as problematic for comparison with hypnotisability scores generated using existing scales. Second, we have been investigating relationships between the SWASH and other measures (Stroop word blindness, RHI etc). We are moving carefully, one step at a time. This scale is not intended as the final word on imaginative suggestion measurement.

Added to p.33: “Our approach to scale development has been conservative thus far. The SWASH is an adaptation of the long established WSGC, itself a group adaptation of the Stanford. The PCS is a close adaptation of the SWASH. While this strategy is advantageous with regard to relating results to earlier studies, a disadvantage is that weaknesses of these historical scales have not been removed. There is, however, a need for development of imaginative suggestion scales, which have undergone relatively little development since their initial development many decades ago (see Acunzo & Terhune, 2020). Future development of phenomenological control scales can afford to be less conservative, for example, by revising the post-session/post-hypnotic suggestion and by introducing new items.”

33. *A number of the items do not correlate well with the total scale scores on the PCS (negative visual hallucination, arm immobilisation, posthypnotic suggestion) - again, this points to the poor structure of these scales. This needs to be more readily acknowledged and is at odds with the statements in the Discussion arguing that this is a good scale.*

We leave this long standing point of debate in the hypnosis literature to be resolved in the future (see previous response regarding future scale development). Note that while having a single factor structure implies simplicity in interpreting the scale, if in fact phenomenological control is multidimensional, then it is also good that the scale reflects

this. We have added a principal components analysis to the supplementary material however (which yields one major dimension only, though with low loadings by a couple of items; <https://osf.io/7x3fy/>).

34. *The retest data and design are unclear given the description of the retest design in the Methods section (I assume the description in the Methods is incorrect?)*

The description in the methods has been corrected. See responses to points 18 and 26.

35. *In the context of the retest again it would be nice to know the amount of time between sessions.*

See response to point 25.

36. *Why are BFs reported for some analyses but not for others? This could be clearer.*

Added: “Bayes factors are provided for inferential analyses (rather than when estimation alone suffices).”

37. *It would be valuable if the Results section more clearly distinguished a priori from exploratory analyses as it's easy for the reader to mix up the two since the pre-registration is only mentioned in passing in the Methods section.*

Headings added.

38. *Please consider reporting the subjective-behavioural correlations across measures.*

Table 2 shows subjective/objective correlations for both scales.

Discussion

39. *It would be valuable in the Discussion to compare the test-retest reliability of the scales to other measures (e.g., the Elkins scale or the Harvard/Stanford scales).*

The closest comparison is with test-retest of the WSGC experiential scale across hypnotic and non-hypnotic contexts.

Added to p.30: “Test retest correlations across the two scales were comparable to the test-retest correlation (.50) for experiential scales across hypnotic and non-hypnotic presentations of the WSGC (in which the hypnotic context was presented second; Meyer & Lynn, 2011)”.

40. *The increase in response to suggestions in the PCS relative to the SWASH was really specific to low and medium responders - this was also found by Scacchia and De Pascalis (2020) and should be noted as it has implications for suggestion research, which is often on very high responders.*

Added to p.29: “are driven by differences at the lower end of the scale”

41. *“The hypnotic context and induction may reduce correlations between hypnotisability and the use of phenomenological control in other contexts.” Quite the opposite actually. The PCS has lower internal consistency, which places an upper boundary on the correlations you can expect to observe. All things being equal, these data suggest you’re going to get poorer correlations with the PCS than the SWASH. For example, this is probably why you observed better correlations with the subjective scale than the behavioural scale in your rubber hand illusion research (this has also been observed with other trait studies on hypnotic suggestibility).*

There are two distinct issues here. Yes, the statistical ceiling is lower. It is an empirical question as to whether this is offset by the more closely matched contexts. Evidence from our first study employing this effect indicates that it may be (the correlation between PCS score and visually evoked auditory response is .37(Lush, Dienes, Seth, et al., 2021)).

42. *“We suggest that the presentation of imaginative suggestions in a scientific context (the context of a scientific experiment rather than the context of hypnosis) ...”
...
This implies that hypnosis can’t be used in scientific experiments - please rephrase.*

We agree with the reviewer.

Added to p.30: “though note that hypnosis can, of course, be performed within the context of a scientific experiment.”

*“While we do not recommend the use of scores generated by the objective scales ...”
Why have you used them in your previous studies then?*

Because we have become more aware of the problems with them during the course of scale development. It is partly the repeated use of them that has informed our beliefs.

Edited on p.31: “While we no longer use and do not recommend the use of scores generated by the objective scales (which are really no more objective than the subjective scales; see Lush, Seth, & Dienes, 2021 for a related discussion on the interpretation of measures which are labelled as ‘objective’

43. *“The scale thus has a mix of the suggestions requiring possibly different subskills ...”
Please reference seminal studies by Woody and McConkey (2003, Int J Clin Exp Hypn; Woody et al. 2005, Psych Assess), who were the main proponents of this idea.*

Added ref to (Woody & McConkey, 2003)

44. *“The terms “induction” and “hypnosis” presume there is a special procedure for inducing an altered state (Coe, 1992).”*

Only to an extent. Many researchers have conceptualised the induction as a preamble/introduction similar to your imagination preamble or as a simple first suggestion (e.g., Nash).

Added "For some researchers"

Minor

48. *"However, hypnosis is not required for successful responding."*
To what?

Added "to imaginative suggestion"

45. *"Thus, the potential usefulness of a phenomenological control scale."*
Missing word?

Sentence deleted.

46. *"to avoid potentially unpleasant effects of two WSGC suggestions."*
Unclear - more detail is needed.

This is discussed in the paper in which norms for the SWASH were presented. We have clarified that this information is available in Lush et al (2018)

47. *Define PC at first use.*

We do – the first sentence in the manuscript is a definition. Repeated here:

"People to varying degrees are capable of altering their subjective experience such that it misrepresents reality in ways consistent with goals, and such that the misrepresentation can be sustained over at least minutes despite clear contrary evidence, a capacity we call phenomenological control"

48. *Some references are no in the reference list (e.g., Parra...)*

(Parra & Rey, 2019) added to reference list.

49. *There is some repetition (e.g., mentioning groups of 50 multiple times seems unnecessary).*

"in groups of up to 50" removed from p.10

50. *The reference to taste in the analysis section is unclear as there is no mention of taste or what this means in the Materials (again, it would be nice to know which suggestions are on this scale). Same goes for urge and amnesia responses - this is unclear because it was never introduced in the materials*

Suggestions are now described in the methods section (see earlier response)

51. There is a reference to a posthypnotic suggestion but there is no such item on the PCS - this could be clearer.

Changed to "post-session suggestion".

52. What was the difference score reported in Lush et al. (2018)? Between which two conditions? This could be clearer.

Clarified:

"These were based on the difference between the minimum possible score on the scale and the scale scores reported in Lush et al, (2018) ~~and minimum scale scores~~ (subjective scale, mean of 1.7 and minimum score 0; objective scale, mean 3.7 and minimum score 0)."

53. "The ratio of these ORs was 4.9 ...", sorry but I found this unclear - please provide more information.

By ratio we mean one OR divided by the other – the literal meaning of ratio.

54. There a handful of random typos in, Table 5, the reporting of BFs (e.g., p. 17), correlation on p. 18, etc. so please correct these.

Typos corrected where found.

55. There are no a) and b) panels in Figure 1.

Corrected

56. Table 6: does the term hypnotisable have any meaning here if no hypnosis was used with the PCS?

"Hypnotisable" removed.

Reviewer 2

1. *The manuscript addresses a very interesting topic, with substantive theoretical and empirical implications for the advancement of the knowledge of consciousness and its alterations. The main problem is that authors do not mention previous scientific works that already carried out almost identical investigations, but with other scales. For example, T.X. Barber created two scales (Barber Suggestibility Scale, and Creative Imagination Scale, both without any prior hypnosis inductions. The Babers's aim was to demonstrate that all hypnotic "phenomena", included the experience of non volition when succeeding suggestions, could be elicited without hypnotic inductions... debunking, in this way, the concept of hypnotic trance. Even more, he created Task Motivational Instructions as an alternative to hypnotic suggestions. TMI elicited objective reactions and subjective experience similar to obtained using hypnosis induction and suggestions. Therefore, again, Barber and his colleagues' series of studies showed evidence about the weakness of theoretical perspectives about de "essence" of hypnosis based on the intrinsic properties of hypnotic induction and hypnotic suggestions. I miss references to this line of work already old (from 1960 on...), which matches with the aims and goals of the manuscript.*

Our intention is not to ignore Barber's work, and we previously stated this (on p.3):

"Here we follow Barber & Wilson (1978) in developing a scale to measure response to imaginative suggestion outside the hypnotic context (see also Oakley et al, 2021).

We also discussed why we are moving on from Barber's terminology.

We have added the following text further to emphasise that earlier non-hypnotic imaginative suggestion scales exist and to clarify our motivation for the present scale:

P.4: "There are existing scales which measure response to non-hypnotic imaginative suggestion, most notably the scales arising from Barber and colleagues work on non-hypnotic imaginative suggestion (Barber & Glass, 1962; Barber & Wilson, 1978). Indeed, any hypnosis scale can be turned into a non-hypnotic scale simply by removing reference to hypnosis and related concepts such as trance, sleep or relaxation (e.g., Meyer & Lynn, 2011), and a non-hypnotic scale transformed into a hypnosis scale by adding reference to hypnosis and related concepts (e.g., Braffman & Kirsch, 1999). However, this has so far failed to displace hypnosis as the dominant context in which imaginative suggestion research is conducted. Because the hypnotic context refers to scientifically outdated conceptions of the phenomena in question, we consider this to be a barrier to progress."

P.7: "Terminology may have played an important role in the failure of Barber's scales to displace hypnosis scales. The term "creative imagination" perhaps fails to convey the range of response to imaginative suggestion, for example, surgical anaesthesia (Esdaile, 1852; Wobst, 2007). A reader unfamiliar with the history of imaginative suggestion research may have a misunderstanding of the phenomena which a researcher studying "creative imaginative" is investigating."

P.34: "Although we believe that imaginative suggestion research has been hampered by the outdated label of 'hypnosis', we are aware that this claim may not be welcomed by

researchers who have much invested in the hypnotic context. It is for this reason that we have changed only as much as was necessary to remove the hypnotic context in adapting the SWASH scale (e.g., references to relaxation, sleep, trance states; (Lynn et al., n.d.). We hope that fears that the scale measures something different to hypnosis scales may be allayed somewhat by this conservative approach. We do not discount hypnosis research, the context has proven fruitful in the study of imaginative suggestion effects. We also do not discount the use of the hypnotic context in clinical treatment, whenever it proves useful (see (Lynn et al., 2019). However, continued adherence to a term rooted in 19th century understanding and which is so laden with misleading mythology may be of more harm than good to progress in scientific research on the phenomena in question.”

2. On the other hand, The Phenomenological Control Scale statistical analysis, at least, lacks from an exploratory factorial analysis. Much better would be to use and confirmatory factorial analysis, so that internal consistency of the scale could be addressed from another perspective, and for scales’s low level on internal consistency could be better understood.

We have added a principal components analysis (the simplest cousin of factor analysis), to the supplemental material (<https://osf.io/7x3fy/>).

3. Therefore, the manuscript aims and goals are strongly interesting, but it have to mention previous research, justify better why to use The Phenomenological Control Scale (only a change of name of the Sussex Waterloo Scale of Hypnotisability), when there are other scales (as already I mentioned) created for the same purposes... and reanalyze data with more updated statistical methodology.

We have used the SWASH to test hypotheses regarding phenomenological control in other effects studied by psychologists (Lush et al., 2020). We developed this revised scale to remove the hypnotic context for future related work (e.g., Lush, Dienes, Seth, et al., 2021). Rather than keep it to ourselves, this norm paper is being made available to aid interpretation of our current and forthcoming work using this scale and so that others can use it if they want.

Reviewer 3

The current study describes a new “phenomenological control scale” that measures trait response to imaginative suggestions. The scale is a modification of a previous hypnotic suggestibility scale. The results are expected, and the statistical analysis seems to be correct. However, the interpretation of the results, the literature review, the conclusions, and the concept of “phenomenological control” are all deeply problematic. Many of the conclusions are not supported by the results, and there are many overstatements and incorrect statements in the text. The study’s limitations are not critically considered. Relevant literature is not cited, or the results of cited studies are misrepresented to fit the authors’ narrative. My overall assessment is that the current manuscript is unacceptable. Below, I list my main specific concerns.

1. *I find the concept of “phenomenological control” very problematic. Healthy psychology undergraduates, like the current sample, do not hallucinate, and they are not delusional. Normal humans cannot “control their conscious perception” and see things in front of their eyes that are not there. The whole idea of phenomenological control seems rather absurd to the reader, at least from how it is presented in the current manuscript. Phenomenological control is confounded by mental imagery. The participants probably follow the instructions and suggestions and engage in various forms of mental imagery. This alternative view is not considered, and the relationship between mental imagery and “phenomenological control” is not discussed. This is a major weakness of the study.*

We had failed to clarify what we mean by phenomenological control. It is not intended to describe a new concept but rather as a new term to describe existing concepts. Many hypnosis experiments on undergraduates over the past 100 years have routinely shown they will report hallucinations and delusions in a controlled way, in responding to suggestions. It is accepted in hypnosis research that they do, as the reviewer states themselves in the next comment. Whether imagery is needed for this is a matter of debate; the previous reviewer for example believes there is no relationship between imagery and successful response to at least most suggestions – though presumably imagery is needed for hallucinations. A recurring claim in theories of hypnosis is the argument that mental imagery plus feeling of involuntariness is enough to count as a ‘hallucination’ in terms of report. We argue elsewhere that this is the same thing going on, for example, in rubber hand illusions etc (for example, via the cold control theory of phenomenological control; Dienes, Lush, et al., 2020; Dienes, Palfi, et al., 2020). It is unclear to us why merely removing reference to hypnosis (a context which is laden with scientifically inaccurate beliefs) from an imaginative suggestion screening procedure transforms a phenomenon from something the reviewer apparently accepts to something absurd.

P.3: “Notably, merely the addition of the word ‘hypnosis’ to a screening procedure is sufficient to produce such a boost (Gandhi & Oakley, 2005). Rather than this word having any unique status, a simple explanation is that it drives particular demand characteristics in participants who arrive with a culturally acquired knowledge of the hypnotic context. Again, we emphasise that response to imaginative suggestion predates the introduction of the term ‘hypnosis’ (most famously in mesmerism but also other contexts, Hammond, 2013).

2. *True automatic hallucinations are possible to induce with a hypnotic suggestion. However, this is a rare ability, and even very highly hypnotizable individuals are an extremely heterogeneous group.*

We agree that reports of some hallucinations are rare; others are common. For example, PCS pass rates on objective criteria for auditory hallucination are 8% and for negative visual hallucination 4%. Taste hallucination is 45%. Mosquito hallucination (auditory or tactile experience) is 45%. This varying rates of responding are similar for hypnosis screenings (as shown in our manuscript).

High hypnotisables are identified by the use of imaginative suggestion scales. The SWASH derives from the Waterloo-Stanford scale, which is a direct descendent of the Stanford scales. The only difference between PCS and SWASH is the reference to hypnosis (and phenomena associated with hypnosis, such as trance and relaxation). The reviewer seems to accept reports of hallucination as genuine when interpreting hypnosis scales, but not when the hypnotic context is removed. This seems to us to prejudge an important issue.

3. *There exists strong converging support for the idea that hypnosis may involve a special state of consciousness in some rare individuals. Hypnotic hallucinations are thus a rare phenomenon (like synaesthesia), and the current way of defining it is heavily confounded by mental imagery. These critical views must be discussed.*

Many current researchers investigate imaginative suggestibility without postulating an altered state. The use of a scale without an induction leaves open what is added to experience by adding an induction. It is simply an empirical issue. But whatever subjects experience without an induction (e.g. on the PC scale), is surely very interesting because they report hallucinations and other transformations of subjective experience as real or more real as when a standard hypnotic context is used. Researchers who want to operate solely within the hypnotic context, or explore effects of inductions on reports, are welcome to do so. Meanwhile others can explore the reported transformations of experience that occur outside of that context. And note in the latter case it is still very possible to select those most highly skilled in PC as special case studies if a researcher wishes.

Added to p.31:

“As mentioned, it has long been known that response to imaginative suggestion It has long been known that response to imaginative suggestion does not require a special state ((Barber & Glass, 1962; Hull, 1933) and the postulation of a state has not so far been shown necessary to explain response to imaginative suggestions (see Jensen et al., 2017; Lynn et al., n.d.)”

4. *The study presents no evidence that the “generated experiences” were non-voluntary. Most of the participants were likely aware that they were generating mental imagery voluntarily, at least to some degree, or their experience was a mixture of volitional and automatic. Therefore, the statements on lines 367-368 are speculative and most likely incorrect. This is a critical point because “hallucinations” should be non-voluntary and experienced as real. The latter*

crucial metacognitive judgment was also not investigated in the current study. To what extent did the participants experience the “hypnotic experiences” are real instead of simply imagining things or acting?

We are, for this issue (strength of experience), treating the subjective reports with as much seriousness for PC as for SWASH - i.e. we are not using a double standard for the two scales. Subjects rate their experiences similarly but as a bit more compelling for PC than for SWASH. We agree that this issue should be addressed repeatedly in future research, as it is a central one.

The extent of involuntariness experienced in the hypnotic setting has been well researched; McConkey (2008) has investigated how real subjects find the experiences, and individual differences in involuntariness and other experiences; SJ Lynn has explored the malleability of involuntariness ratings; others such as Kihlstrom, Wagstaff or Coe have explored trying to breach the experience. Our goal is not to review the whole hypnosis field, but to rely on its broad conclusions for taking things further.

On a related note, the PC scale is a relatively blunt tool, just like other existing imaginative suggestion and hypnosis scales. It is sufficient for our current research questions. There is plenty of room in future research to refine this tool or to develop others (as we have noted in the manuscript).

Added to p.34 “On a practical note, in terms of the usefulness of imaginative suggestions tested out of versus within the hypnotic context, we have investigated relationships between response to imaginative suggestion and other experimental reports using both the SWASH (e.g., rubber hand illusion and mirror synaesthesia; (Lush et al., 2020) and phenomenological control scales (e.g., visually evoked auditory response; Lush, Dienes, Seth, et al., 2021). As for the SWASH, relationships between the PC scale and other measures are substantial; reports of visually evoked auditory response are predicted to increase half a point for each 1 point increase in PC scale, (both on 6 point scales), $r_s = .37$. At least for the purposes of predicting reports of experiential change from trait response to imaginative suggestion (i.e., phenomenological control as a potential confound in psychological experiments), the PC scale appears to be no less effective than the SWASH.”

Regarding involuntariness, see our earlier edit in response to R1 (Point 7).

We consider the evidence taken overall indicates that not everybody is faking in both phenomenological control research (within and outside the hypnotic context) and in these other phenomena. At least some people really seem to believe they have had a change in experience in response to imaginative suggestion, and at least some people really seem to believe they have had anomalous experience of ownership of a fake hand in the rubber hand illusion. In both cases, brain imaging and other physiological measures (e.g., skin conductance response) are consistent with there being some genuine change in experience. Consider the frequency of report of experience in each case. Around 70-80% of participants report experience for ideomotor suggestions, for example that their hands are drawn together as though by a magnetic force. A similar percentage report anomalous experience in the rubber hand illusion. We consider it more plausible that most people providing such reports genuinely believe they have had a relevant experience than that they are all faking. In other words, we do not employ a

double standard to interpret reports of direct verbal imaginative suggestion effects and reports of anomalous experience which follow other procedures (such as the stimuli used in the rubber hand illusion).

5. *No evidence for genuine alternations in immediate perceptual awareness is reported in the article. All results could stem from visual, auditory, tactile, and taste imagery (so lines 369-371 are overstatements). I also do not think that any of the current students were genuinely delusional as claimed on lines 370-373; at least, no evidence for this is presented. The participants probably just responded to social compliance and engaged in mental imagery and role-playing. I do not think you are justified in referring to the current experiences as “hallucinations” or “delusions”.*

McGeown and colleagues showed V4 activation for suggestions for colour hallucinations; Derbyshire and colleagues changes in the pain matrix for pain hallucinations; Coltheart and colleagues showed breaching hypnotic delusions was difficult. These are all now classics in the field. We now refer explicitly to a review of this literature as it pertains to phenomenological control (Dienes, Palfi, & Lush, in press). Against a virtually consensus conclusion in the literature about the majority experience of subjects, claims that it is all compliance will need solid evidence. Even Wagstaff (who tried to push it as much as he could) eventually gave up on the claim.

Added to p.9: “There is much evidence that response to imaginative suggestion involves genuine change in experience. For example, McGeown et al. (2021) showed V4 activation for suggestions for colour hallucinations; Derbyshire et al. (2009) changes in the pain matrix for pain hallucinations in people with fibromyalgia. We have previously reviewed this evidence as it pertains to phenomenological control (see Dienes, Palfi, & Lush, in press; for a review of other evidence, see McConkey, 2008). That imaginative suggestion can apparently lead to change in experience does not, of course, imply that any given response on a hypnosis scale can be considered to reflect genuine experience. Demand characteristic effects are omnipresent wherever expectancies are uncontrolled, and demand characteristics are transparent in imaginative suggestion (Orne, 2009; see Corneille & Lush, 2021 for a simplified model of demand characteristic effects including phenomenological control).”

6. *The authors should discuss the distinction between “suggestions” and “instructions”. To what extent were the subjects simply instructed to perform mental imagery. Instructions almost always lead to a response since only ordinal social co-operation is required (and in the current study, the capacity to perform mental imagery). Automatic responses and purposeful imagining are two fundamentally different processes in highly hypnotizable individuals. I am concerned that the authors mainly gave instructions to imagining things and not “suggestions”.*

We have now made explicit that the suggestions were worded almost identically (with differences now noted, i.e. removal of words such as “relax”) in the PC and SWASH, and the responses were rated on exactly the same scales. response to R1, point 19

We have added in the Discussion (p.32):

“It is possible that the preamble to the PCS, inviting subjects to actively engage with the exercises, promotes a different way of responding than a hypnotic context. However, the results of this study do not easily support this interpretation. The suggestions in both the PCS and SWASH scales are worded as suggestions for changes in experiences rather than instructions or requests. The response scales ask about the experiences in terms of distortion in volition and perception; thus, prima facie, subjects have similar experiences on both scales. Responding in the hypnotic context has also long been recognized as an active striving (White, 1942), a process with which subjects actively engage (Sheehan & McConkey, 1982), with the subjective distortions building up over time within each suggestion (McConkey et al., 1999).”

The suggestions on the PC and SWASH scales are standard imaginative suggestions, used in countless studies and dating back to the Stanford scales of the 1950s. If the reviewer is a hypnosis researcher, their work almost certainly makes use of these suggestions or close relatives of them. Criticism of these standard suggestions applies therefore to a much wider body of literature than this study. Note that anticipation that some hypnosis researchers would be uncomfortable with the idea of abandoning the hypnotic context is the reason for our conservative approach to scale development, sticking closely to established hypnosis scales and changing things one step at a time (see R1 and R2 comments asking why we did not just develop an all new scale).

As noted by R1 and R2, that imaginative suggestion does not require the hypnotic context is not a novel claim (e.g., Hull, 1933). Note also that, while the preamble urges subjects to engage with the exercises, the ratings ask about how the response was *experienced* – and subjects report the suggestions as being experienced like hypnotic suggestions. This fits the already substantial literature on suggestions given without inductions being effective to a similar degree.

7. The issue of response bias is critical in the current study. The results might not reflect “phenomenological control”, or even the capacity to engage in vivid mental imagery, but just various forms of response bias. Biases that are particularly relevant to consider are demand characteristics, social desirability, and acquiescence bias. This issue should be explicitly discussed.

We agree, and are very interested in the question of the degree to which any given report is faking, imagination, or something else. See Corneille & Lush (2021) for a simplified model of demand characteristics which proposes that, when demand characteristics are not controlled, any given response may be attributable to faking, imagination or phenomenological control. This, of course, applies to any given response in a hypnosis screening and in a non-hypnotic imaginative suggestion screening equally (demand characteristics are transparent in such screening procedures). As we have clarified elsewhere, there is good evidence that imaginative suggestion is at least not entirely faking. There has been a huge body of research into this issue conducted over the last century, which we have reviewed elsewhere and now refer to in the text. That hypnotic responding at least sometimes involves the experiences reported is not a contentious issue (we are aware of no researchers currently publishing on imaginative suggestion effects who hold this position). Again, the only difference between this scale and the hypnosis scale on which it is based is that the hypnotic context is not established (largely because hypnosis is not mentioned). We choose not to apply a

double standard to the interpretations of response to imaginative suggestion in 'hypnosis' and in other contexts.

We now have extended the discussion of demand characteristics for PC and more generally (see point 4)

p.33: "Regardless of the degree to which one believes imaginative suggestion scales (within or outside the hypnotic context) reflect genuine experience (for a review of evidence for this see Dienes, Palfi, & Lush, in press; for a review of other evidence see McConkey, 2008), other demand characteristics effects such as faking, wilful imagination (see Corneille & Lush, 2021) or (as is most likely) some mixture of these possibilities, it is important to take stable trait differences in a tendency to report anomalous experiences in experimental situations seriously."

8. In the abstract and introduction, the rubber hand illusion is mentioned. However, the authors should remove this text because Lush et al Nat Communications 2020 present clear evidence against a relationship between hypnotic suggestibility and the rubber hand illusion. When a control condition is used in the analysis, which is required to obtain meaningful results, there is no relationship between the rubber hand illusion and the hypnotizability score. Thus, this finding actually seems to falsify the phenomenological control theory.

The reviewer's stated argument disregards the logic of controlled experiments. See (Lush, 2020; Lush, Dienes, & Seth, 2021; Lush, Seth, & Dienes, 2021; Reader, 2021; Roseboom & Lush, 2020).

For a control measure to be valid, everything except the independent variable must be held constant. In the asynchronous control condition this is not the case. Expectancies for asynchronous stroking differ from expectancies for synchronous stroking. The change of timing of brush strokes from synchronous to asynchronous does not just change the timing of multisensory stimuli, but also the demand characteristics of the task. Therefore, even if the results of RHI experiments are entirely attributable to demand characteristics, we should still see a difference between synchronous and asynchronous conditions. The confounding effects of demand characteristics are not limited to phenomenological control (Corneille & Lush, 2021; Orne, 1962; Sharpe & Whelton, 2016). Because expectancies do not differ across levels of hypnotisability, the null correlation we reported between the difference between synchronous and asynchronous conditions and hypnotisability (Lush et al., 2020) can easily be explained (see Lush, Dienes, & Seth, 2021).

In short, an asynchronous condition is not a valid control measure for a synchronous condition in embodiment illusions. In common with much of the RHI literature (e.g., see Botvinick & Cohen, 1998, on which the field is largely based), we interpret synchronous condition agreement scores in the rubber hand illusion. These are substantially predicted by hypnotisability.

Finally, note further that even if the reviewer were correct in their interpretation of Lush et al 2020 and the RHI was not related to phenomenological control, this would

not be evidence against phenomenological control but merely evidence against it driving this particular effect.

Added to p.33: “With reference to effects for which relationships between phenomenological control have been shown (thus far the rubber hand illusion, mirror touch synaesthesia, vicarious pain and visually evoked auditory response), there is a measurable trait which predicts reports of experience. When these relationships are taken into account, these effects require re-interpretation. For example, without higher PC participants there is not agreement on average for an illusion of ownership of a fake hand at typical sample sizes (see Roseboom & Lush, 2020); this puts pressure on theories of the rubber hand illusion not based on phenomenological control.”

On page 5, the authors misrepresent the “constructive nature of consciousness”. Perceptual conscious experiences are not just constructed in the brain out of nothing but results from an interplay between top-down predictions and bottom-up sensory signals. The brain tries to predict the sensory inputs in order to minimize prediction errors (e.g., predictive coding); not construct “fake” sensory data. Hallucinations generated according to the phenomenological control theory would presumably maximize prediction errors. So, I do not follow the authors’ argument here.

Sensory imagination involves the construction of sensory representations, for example in the visual case perhaps all the way down to V1. And somehow we manage this, despite what must be some prediction error.

Although psychosis and some other clinical conditions are interesting to consider in this context, nothing in the cited theories supports that healthy individuals can construct genuine perceptual-sensory experiences in the way that the authors claim in the manuscript.

The reviewer stated in point 2 that high hypnotisables can do just this.

9. The authors claim that “phenomenological control” and hypnotisability are single stable traits without thoroughly discussing the likely possibility that these scales capture different kinds of subskills. There might not exist a single stable trait for “phenomenological control” or hypnotic suggestibility.

We agree. We have added a principal components analysis to the supplemental material (<https://osf.io/7x3fy/>) to show what light our data might shed on this interesting and long standing issue. This was chosen over factor analysis for simplicity, as it has minimal researcher degrees of freedom.

Added comparison of the underlying structure of the two scales to the manuscript (p.28).

“Exploratory PCA results were similar for the two scales (see supplementary results at <https://osf.io/7x3fy/>), showing a primary component of overall response to suggestion and a secondary small component that contrasts perceptual-cognitive suggestions with motor suggestions, consistent with the motor vs perceptual-cognitive by challenge vs direct classification of Woody and Barnier (2008)

10. *The scale was developed outside the context of hypnotic suggestions. One drawback could be that the scale no longer captures what is really interesting: those rare, highly suggestible individuals who experience genuine hallucinations. This limitation should be discussed.*

The existence of hypnotic virtuosos is not a question we address here. However, such virtuosos are typically identified by the use of hypnosis scales. As explained in the manuscript, the scale was developed directly from a hypnosis scale (in turn closely based on an earlier hypnosis scale which is a direct descendent of scales which have been in use for over 60 years). The difference between SWASH and PCS is the hypnotic context. The hypnotic context can be reduced to including or excluding the word “hypnosis” from a procedure (Gandhi and Oakley). This is an awful lot to hang on a word.

P.3-4: “Notably, merely the addition of the word ‘hypnosis’ to a screening procedure is sufficient to produce such a boost (Gandhi & Oakley, 2005). Rather than this word having any unique status, a simple explanation is that it drives particular demand characteristics in participants who arrive with a culturally acquired knowledge of the hypnotic context. Again, we emphasise that response to imaginative suggestion predates the introduction of the term ‘hypnosis’ (most famously in mesmerism but also other contexts, Hammond, 2013).”

The last sentence (line 394) should probably be deleted.

We actually quite like it as a finishing note. It highlights a key message. We have added further clarification: “Phenomenological control may be widespread in psychological experiments without people noticing. And people may not have noticed because PC has been called hypnosis, a word and context that suggests something both magical, yet also mildly disreputable. Our aim is to encourage people to look and see.”

References

- Barber, T. X., & Glass, L. B. (1962). Significant factors in hypnotic behavior. *The Journal of Abnormal and Social Psychology*, 64(3), 222–228.
<https://doi.org/10.1037/h0041347>
- Barber, T. X., & Wilson, S. C. (1978). The Barber Suggestibility Scale and the Creative Imagination Scale: Experimental and Clinical Applications. *American Journal of Clinical Hypnosis*, 21(2–3), 84–108.
<https://doi.org/10.1080/00029157.1978.10403966>
- Botvinick, M., & Cohen, J. (1998). Rubber hands ‘feel’ touch that eyes see. *Nature*, 391(6669), 756–756. <https://doi.org/10.1038/35784>
- Cardeña, E., & Terhune, D. B. (2009). A Note of Caution on the Waterloo-Stanford Group Scale of Hypnotic Susceptibility: A Brief Communication. *International Journal of Clinical and Experimental Hypnosis*, 57(2), 222–226. <https://doi.org/10.1080/00207140802665484>
- Corneille, O., & Lush, P. (2021). *Sixty years after Orne’s American Psychologist article: A conceptual analysis of “Demand Characteristics”*. PsyArXiv.
<https://doi.org/10.31234/osf.io/jqyvz>
- Dell, P. F. (2021). Hypnotizability and the Natural Human Ability to Alter Experience. *International Journal of Clinical and Experimental Hypnosis*, 69(1), 7–26. <https://doi.org/10.1080/00207144.2021.1834859>
- Dienes, Z. (2012). Is hypnotic responding the strategic relinquishment of metacognition? In *Foundations of metacognition* (pp. 267–278). Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780199646739.003.0017>
- Dienes, Z., Palfi, B., & Lush, P. (in press). Controlling phenomenology by being unaware of intentions. In Weisberg, J. (Ed.), *Qualitative Consciousness: Themes from the Philosophy of David Rosenthal*. Cambridge University Press
- Dienes, Z., Palfi, B., & Lush, P. (in press). Controlling phenomenology by being unaware of intentions. In Weisberg, J. (Ed.), *Qualitative Consciousness: Themes from the Philosophy of David Rosenthal*. Cambridge University Press. <https://doi.org/10.31234/osf.io/4zw6g>
- Esdaile, J. (1852). *The introduction of mesmerism, as an anæsthetic and curative agent, into the hospitals of India*. Dewar.
- Gandhi, B., & Oakley, D. A. (2005). Does ‘hypnosis’ by any other name smell as sweet? The efficacy of ‘hypnotic’ inductions depends on the label ‘hypnosis’. *Consciousness and Cognition*, 14(2), 304–315.
<https://doi.org/10.1016/j.concog.2004.12.004>
- Hammond, D. C. (2013). A Review of the History of Hypnosis Through the Late 19th Century. *American Journal of Clinical Hypnosis*, 56(2), 174–191.
<https://doi.org/10.1080/00029157.2013.826172>
- Hull, C. L. (1933). *Hypnosis and suggestibility* (pp. xii, 416). Appleton-Century.
- Kirsch, I., & Council, J. R. (1989). Response expectancy as a determinant of hypnotic behavior. In *Hypnosis: The cognitive-behavioral perspective* (pp. 360–379). Prometheus Books.

- Lush, P. (2020). Demand Characteristics Confound the Rubber Hand Illusion. *Collabra: Psychology*, 6(1), 22. <https://doi.org/10.1525/collabra.325>
- Lush, P., Botan, V., Scott, R. B., Seth, A. K., Ward, J., & Dienes, Z. (2020). Trait phenomenological control predicts experience of mirror synaesthesia and the rubber hand illusion. *Nature Communications*, 11(1), 4853. <https://doi.org/10.1038/s41467-020-18591-6>
- Lush, P., Dienes, Z., & Seth, A. (2021). *Rubber hand illusion reports remain confounded by demand characteristics and are substantially related to trait phenomenological control*. PsyArXiv. <https://doi.org/10.31234/osf.io/qh4ag>
- Lush, P., Dienes, Z., Seth, A., & Scott, R. B. (2021). *Trait phenomenological control predicts visually evoked auditory response*. PsyArXiv. <https://doi.org/10.31234/osf.io/x9eud>
- Lush, P., Moga, G., McLatchie, N., & Dienes, Z. (2018). The Sussex-Waterloo Scale of Hypnotizability (SWASH): Measuring capacity for altering conscious experience. *Neuroscience of Consciousness*, 2018(1). <https://doi.org/10.1093/nc/niy006>
- Lush, P., Seth, A., & Dienes, Z. (2021). *Demand characteristics confound asynchronous control conditions in indirect measures of the rubber hand illusion*. PsyArXiv. <https://doi.org/10.31234/osf.io/w67xc>
- Lynn, S. J., Green, J. P., Polizzi, C. P., Ellenberg, S., Gautam, A., & Aksen, D. (2019). Hypnosis, Hypnotic Phenomena, and Hypnotic Responsiveness: Clinical and Research Foundations—A 40-Year Perspective. *International Journal of Clinical and Experimental Hypnosis*, 67(4), 475–511. <https://doi.org/10.1080/00207144.2019.1649541>
- McConkey, K. M., & Sheehan, P. W. (1982). Effort and experience on the creative imagination scale. *International Journal of Clinical and Experimental Hypnosis*, 30(3), 280–288. <https://doi.org/10.1080/00207148208407265>
- McConkey, K. M., Wende, V., & Barnier, A. J. (1999). Measuring change in the subjective experience of hypnosis. *International Journal of Clinical and Experimental Hypnosis*, 47(1), 23–39. <https://doi.org/10.1080/00207149908410020>
- McConkey, K. M. (2008). Generations and landscapes of hypnosis: Questions we've asked, questions we should ask. *The Oxford handbook of hypnosis: Theory, research and practice*, 53-77.
- Meyer, E. C., & Lynn, S. J. (2011). Responding to Hypnotic and Nonhypnotic Suggestions: Performance Standards, Imaginative Suggestibility, and Response Expectancies. *International Journal of Clinical and Experimental Hypnosis*, 59(3), 327–349. <https://doi.org/10.1080/00207144.2011.570660>
- Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the penn state worry questionnaire. *Behaviour Research and Therapy*, 28(6), 487–495. [https://doi.org/10.1016/0005-7967\(90\)90135-6](https://doi.org/10.1016/0005-7967(90)90135-6)
- Michael, R. B., Garry, M., & Kirsch, I. (2012). Suggestion, Cognition, and Behavior. *Current Directions in Psychological Science*, 21(3), 151–156. <https://doi.org/10.1177/0963721412446369>
- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their

- implications. *American Psychologist*, 17(11), 776–783.
<https://dx.doi.org/10.1037/h0043424>
- Parra, A., & Rey, A. (2019). The interoception and imagination loop in hypnotic phenomena. *Consciousness and Cognition*, 73, 102765.
<https://doi.org/10.1016/j.concog.2019.102765>
- Pintar, J., & Lynn, S. J. (2008). *Hypnosis*. Wiley-Blackwell.
<https://doi.org/10.1002/9781444305296>
- Polczyk, R., & Pasek, T. (2006). Types of Suggestibility: Relationships Among Compliance, Indirect, and Direct Suggestibility. *International Journal of Clinical and Experimental Hypnosis*, 54(4), 392–415.
<https://doi.org/10.1080/00207140600856764>
- Polczyk, R. (2016). Factor structure of suggestibility revisited: New evidence for direct and indirect suggestibility. *Current Issues in Personality Psychology*, 4(2), 87–96. <https://doi.org/10.5114/cipp.2016.60249>
- Reader, A. T. (2021). *What do participants expect to experience in the rubber hand illusion?* PsyArXiv. <https://doi.org/10.31234/osf.io/d8x9y>
- Roseboom, W., & Lush, P. (2020). *Serious problems with interpreting rubber hand illusion experiments* [Preprint]. PsyArXiv.
<https://doi.org/10.31234/osf.io/uhdzs>
- Sharpe, D., & Whelton, W. J. (2016). Frightened by an Old Scarecrow: The Remarkable Resilience of Demand Characteristics. *Review of General Psychology*, 20(4), 349–368. <https://doi.org/10.1037/gpr0000087>
- White, R. W., & Shevach, B. J. (1942). Hypnosis and the concept of dissociation. *The Journal of Abnormal and Social Psychology*, 37(3), 309–328.
<https://doi.org/10.1037/h0057048>
- Wobst, A. H. K. (2007). Hypnosis and Surgery: Past, Present, and Future. *Anesthesia & Analgesia*, 104(5), 1199.
<https://doi.org/10.1213/01.ane.0000260616.49050.6d>
- Woody, E. Z., & McConkey, K. M. (2003). What we don't know about the Brain and Hypnosis, but need to: A View from the Buckhorn Inn. *International Journal of Clinical and Experimental Hypnosis*, 51(3), 309–338.
<https://doi.org/10.1076/iceh.51.3.309.15523>