1 Response to Editor

1.1

Reviewers have now commented on your paper. As you can see from their comments, which are appended below, two of the three original reviewers are now pretty satisfied with the revision. The other reviewer, however, is not equally satisfied, and still raises issues concerning what is seen as a lack of conceptual clarity, both at the abstract and at some other points over the introduction and discussion sections.

We are very happy to hear that two of the reviewers view our manuscript as ready for publication. As you will read below, we think the remaining reviewer raised excellent points, and we think this version of our manuscript is even better than the last. Our sincere thanks to everyone involved.

1.2

I believe that the main message of the experiment is strong and compelling, and that a paper reporting on these results could eventually deserve publication in JEP:LMC. After all, it is one of the very few experimental examples that shows some conditions in which learning can be facilitated by distraction, at least when this learning involves unlearning about previously acquired procedural tendencies. However, I tend to agree with Reviewer 1 in the impression that the message could have been transmitted more clearly over the abstract, and the introduction sections, and I also endorse Reviewer 1 diagnostic with respect to the need to mention and discuss the limitations of the study, specially with respect to the lack of significant differences between several conditions, which, according to your hypotheses, should have produced a difference.

These are valid points. We made several significant revisions in response to these concerns. We changed the title to "Increased Cognitive Load Enables Unlearning in Procedural Category Learning." We also completely revised the abstract. The new abstract is now focused on the core finding of our paper: increased cognitive load enables unlearning. Finally, the lack of dose dependency and overlap effects are now included in our discussion section.

1.3

If you undertake this subsequent revision seriously, I hope to be able to take a final decision myself, without asking for more external reviews. I would like to encourage to do so, and to submit your revised manuscript together with list of changes or a rebuttal against each of these points.

Thank you again for the opportunity to revise. We think this final round of revisions is a significant improvement over the previous versions.

2 Response to R1

2.1

This revision is clearer than the originalthe introduction is greatly improved–and the results more convincing (e.g., Figure 6 is very nice). Very interesting findings! However, I do have remaining concerns as follows.

Thank you! Your comments have helped us pin down the significance of our results and communicate them clearly. As you will read below, we have taken your remaining comments seriously, and think that our paper is again improved over previous versions. Thank you for you careful thinking.

2.2

It is still very hard for people like me who have not been immersed in this line of research to keep the argument straight. Could the underlying argument not be presented more simply and consistently in the abstract and introduction? As I think I understand it, for the present experiment (e.g., page 7) the argument is that detecting that feedback is random prevents unlearning, whereas not detecting the randomness enables it. If this is the case, could this not be stated more clearly in the abstract and earlier on? Particularly in the abstract, upon first reading the argument wasn't making sense to me. First it is stated that "our results suggested that modification of procedural knowledge (note: which I take to be unlearning of the original habit) is possible only if feedback contingency is high." This sentence led me to anticipate that unlearning should be facilitated by being able to detect a contingency during intervention. So this seemed on the surface to contradict a subsequent statement that "increasing cognitive load during intervention via the concurrent performance of an additional task should disrupt the accurate estimation of contingency, thereby keeping the gate on procedural learning open." Although the reasoning becomes somewhat clearer in the introduction, it all still seems more convoluted than would be ideal, and I'm still not sure I've got it right. (Or perhaps I've just got my own mental block on thisI'm feeling dense!) Would it help to try to cut down on the number of terms I wonder–for example, using "unlearning" more consistently instead of alternatives?

First, thank you very sincerely for helping us communicate this information clearly to a broad readership. The original passage on P. 7 of our last revision that you highlight in your comment read as follows:

> If feedback contingency is estimated by executive mechanisms, then increasing cognitive load during the intervention phase (by requiring participants to simultaneously perform a dual task) should impair the ability of participants to detect random feedback. This should in turn cause the TANs gate to remain open, thereby allowing RF to modify the procedural knowledge that was acquired during initial learning.

To our reading, one element of confusion in this writing is the use of the phrase "detect random feedback," which isn't well defined. We have revised this section to read:

> If feedback contingency is estimated by executive mechanisms, then increasing cognitive load during the intervention phase (by requiring participants to simultaneously perform a dual task) should disrupt its estimation. This disruption should deprive the TANs gate of the clear signal they require to close the gate during RF, thereby allowing RF to modify the procedural knowledge that was acquired during initial learning.

2.3

The new discussion is very unsatisfying and oddly removed from the results. There is no explicit mention of limitations of the present study, nor any consideration of the implications of some of the findings. For example, what is the implication of the findings (see bottom of page 13) that there was no evidence of dose dependency or of any influence of the overlap of the dual task from acquisition to intervention? Although I am the reviewer who raised the question about the procedural nature of the original learning, given that this is a brief report, I don't think it is appropriate to take up half of the discussion reviewing the earlier evidence that it is. Instead I'd suggest the lack of direct evidence here might be acknowledged as a limitation, but a sentence citing earlier work that it is procedural be included.

We changed the lengthy discussion about category learning as a procedural skill to the following:

A natural question for readers unfamiliar with the category-learning literature is whether our behavioral paradigm is a good choice for studying procedural behaviors. In other words, how can a task with such simple motor demands (e.g., push a button) possibly recruit procedural networks that are strongly tied to motor processes? In fact, the empirical evidence is strong that performance improvements in the classification task used here are mediated via procedural learning and memory [**?**, **?**, **?**]. Nevertheless, a limitation of the present study is that we did not directly probe the learning to ensure that was procedural in nature.

We also now include a couple paragraphs that to address the absence of dose and overlap effects in the discussion section titled "Dose Dependency and Intervention Onset"

Our design allowed us to ask not just whether contingency estimation relies on executive function, but also whether the effects of disrupted contingency estimation are dose dependent (i.e., whether effects increase with dual-task exposure). We did not find dose effects. The absence of a dose effect indicates that 150 trials of dual-task was equally effective as 350 trials of dual-task. Could it be that 50, 10, or even a single trial of dual-task could again be equally effective? A true absence of dose effect would imply that this is indeed the case, though this prospect seems intuitively unlikely to be true. More likely, all the doses explored in this paper were past a saturation point, at which point dose effects are washed out. In the absence of further data, we are left only to our speculations.

Finally, our design also allowed us to ask whether it is important for the dual task to overlap with the transition from acquisition to intervention. One possibility is that in order for procedural learning to remain vulnerable to modification, the increase in cognitive load would need to precede the onset of random feedback intervention. The thinking here is that the gate that protects procedural learning during random feedback may be sensitive to *changes* in feedback contingency. Another possibility is that any disruption in the computation of feedback contingency at any time can cause the gate on learning to open. This possibility predicts that any increased cognitive load during intervention, regardless where it is placed should enable unlearning via random feedback. We found no evidence that the overlap was important.

2.4

More minor considerations: (a) The term "II" is not defined before its first use (page 3) nor is RB (page 6). Both are defined only on page 14. (b) Figure 2 is referred to in the text after Figure 3 is.

Thanks for the catch. We fixed these.

3 Response to R2

3.1

This manuscript presents an interesting experiment on how feedback contingency is estimated and used to allow or prevent the learning of stimulus-response associations. I was a reviewer on the original version of this manuscript. The authors have been successful at addressing my earlier concerns, and I believe that the manuscript is now adequate for publication in JEP:LMC.

Thank you very much for your help bringing the manuscript to it's final state.

3.2

Figure 3 is called before Figure 2. Please re-number figures.

Thanks for the catch. We made the change.

3.3

It might be useful to give more meaningful names to the conditions instead of Condition 1, Condition 2, etc. I found myself going back and forth in the manuscript to remind myself of what the conditions are. Maybe something like Prior/100, Prior/200, ... Post/100 (with the word indicating if the dual task was introduced before or after the treatment and the number indicating the number of dual task trials).

Thanks for the helpful suggestion. We now label our conditions Overlap-150, Overlap-250, Overlap-350, and No-Overlap-300. The no dual-task control condition is simply referred to as that.

3.4

p. 11: For the intervention ANOVA, why did you pick 4 blocks? Why not 3 or 5?

We chose 4 blocks because visual inspection of Figure 5 (the learning curves) showed that the first 4 blocks is where the differences, if any, would be. Choosing more blocks would washout the signal we were attempting to report, and choosing fewer blocks would needlessly exclude data for the analysis. We added a parenthetical to this section that calls out the importance of visual inspection for this choice.

This is clearly seen in the first four blocks of the intervention phase (visual inspection of Figure **??**), and is supported by the results of a 5 condition $\times$ 4 block repeated-measures ANOVA.

3.5

p. 12, ln 3: You begin your sentence with "First..." but there is no "Second...". Consider revising.

Thanks. We edited accordingly.

4 Response to R3

4.1

I am satisfied with the revisions and think the manuscript is ready for publication.

Thank you very much for your help bringing the manuscript to it's final state.