Dear Tim:

We have extensively revised our MS JEAB-2019-0183 entitled "Number and Time in Acquisition, Extinction and Recovery" in the light of the very helpful comments and suggestions from you and the referees.

What follows are those comments and suggestions, followed by our response (in italics)

Editor's Suggestions  
Like Reviewer 1, I thought some critical sections of the Discussion could use a bit more development. For example, the reviewer notes “In particular, how do you connect the predictive power of a CS and the span of the experiment? It seems that this a critical point for your paper but it is difficult to follow your reasoning here.” In addition, the reviewer notes “I understand that short CS with long inter-stimulus intervals are more informative and that larger experimental span increases the stability of the perceived predictive value of the CS. However, the second part of the prediction, regarding extinction, is unclear as stated.” I too think these sections are critical and require some expansion in order to clearly get your point across.

*Done. We have revised, expanded and, we hope, clarify the reasoning*  
Like Reviewer 2, I found the immediate transition to reporting data from Exp 4 a bit disconcerting. I do think rearranging the experiment numbering would be worthwhile. *Done. We completely agree that this is desirable. It was, however, surprisingly tricky to carry out the renumbering throughout the text AND in the computer code (so as to preserve the data trail).*

In addition, I thought that the expanded (and yes possibly redundant) data presentation suggested by this reviewer might be helpful to many readers.

*We have added Figure 6 and accompanying text in response to this suggestion*

Finally, the reviewer’s issue with how the manipulation of intersession interval and span of training are presented in the abstract and analyses deserves some attention. *Done*  
  
For my part, I thought the section on challenges to associative theories in the Discussion could use some additional development. You say “Here is not the place to elaborate on those challenges nor on how they might be met.” I realize that you don’t really want to get into it, and I sympathize. But, in the Introduction you lay out some basic properties of how different associative theories approach recovery. It seems fair, and symmetrical in terms of the structure of the paper, to circle back to these issues in the Discussion, at least briefly. Your description of the Kraemer & Spear, Bouton, and Estes type approaches gives the impression that they might have something to say about span of training.

*In response to this, we have added a section to the Discussion titled The Challenge for Computational Neuroscience and we have removed this material from the introduction*

I also kept thinking about the Devenport temporal-weighting-rule account of spontaneous recovery---it is not really an associative account, but it did seem worthy of some sort of mention. *We very much agree. We've added several paragraphs devoted to the TWR in the Discussion*

Incidentally, I learned about the weighting rule from previous discussions of related issues in your work on choice. Anyway, if you have anything at all to say about all of this, I think it could make for a more complete and possibly controversial manuscript. If you ignore the associative stuff here, I worry associative theorists might just ignore or totally dismiss this work. If you say a bit more about how their approach is challenged by the current data, maybe it will capture their attention a bit more. *I say more than a bit more now in the section devoted to the challenge in the Discussion*  
I also wondered if it would be useful to mark with ticks or symbols on the x-axis when extinction and the post-delay test begin. *Done (added dashed vertical lines)*  
Finally, for JEAB, you will need to move the Methods into the body of the text, move the figures to the end, and remove the color from Figure 2 (unless you want to pay for it). *Done*  
  
Reviewer(s)' Comments to Author:  
  
Reviewer: 1  
  
Comments to the Author  
In this article, the authors conducted a series of Pavlovian experiments designed to probe the effects of the overall number of training trials, the number of trials per session and the temporal span of training sessions on spontaneous recovery following extinction. They found that the rate of acquisition depends on the number of training trials per session but not the rate of extinction. Importantly, they report that the span of training sessions affects recovery. The authors propose that content-based theories of associative learning are better suited to account for these data than conventional associative theories of conditioning.   
  
The paper is well written, the experiments are well conducted, the results are sound, and the topic is very interesting. However, there are some minor points that could be clarified to improve the paper. I will list them following the order of the manuscript sections. In my opinion, there are several points in the text that are either too assertive or too compact to be convincing. Overall, I think that the impact of the paper would be higher if some theoretical claims were more carefully presented. *We hope you will agree that they are now more carefully presented.*  
Introduction section:   
Page 2 line 36. Two conclusions are expected but there is only one in this paragraph. *Now they are both there*

P. 3, L. 14. There are several claims in the manuscript that subject do computations over something. In my view this is a very strong claim that would deserve a lengthy discussion. I would recommend restraining these claims as one might argue that behaving as if a computation is done doesn’t imply that the computation actually occurred. Because this is not the main topic of the present research, the impact of the paper would not be reduced by stating that responses are well correlated with previous wait times rather than saying that subjects “do computations with the wait times they have learned”. The main point would be preserved and it would be much more convincing to some readers.  
L. 22: same as previous point.  
*The reviewer is certainly right that we are making strong claims and that some readers will be put off by them and we now explicitly acknowledge this at this point in the MS. Also, I've moved almost all of this to the above mentioned final section of the Discussion. I am grinding an axe. But, 1) I've earned the right to grind this axe and 2) In my opinion, it badly needs grinding. I think no small part of the reason why animal learning has gone from the central place in experimental psychology and neuroscience that it occupied when I was a graduate student to the peripheral place it now occupies (in both fields) is that researchers have lacked either the conviction or the courage to stress the implications for neuroscience of what they have discovered. The whole reason I do behavioral experiments is to learn what we should be looking for in the brain. In the light of what we have learned since the 1960s, the current neuroscience story about memory is intellectually bankrupt. We know now that our animal subjects often base their behavior on the ratios of abstract quantities (duration and number, rate and probability). This is no longer controversial. Yet, no neuroscientist is willing even to suggest how a plastic synapse (the associative bond) might store the symbols for durations, numbers, rates and probabilities. The plastic synapse story is not even wrong, because it does not even attempt to explain what obviously needs to be explained. As far as I am concerned, the time is long past when we should have stopped tip toeing around the neuroscientific implications of what we have discovered. I try not to lose any opportunity to call to the attention to all those who worry about neurobiological plausibility the fact that what neuroscientists believe about memory is utterly incapable of explaining what we have learned from the study of animal learning during the last half century. If this makes them uncomfortable, so much the better. The glove is at their feet. Why don't they have the courage to pick it up. As scientists, we are supposed to think about the implications of the facts we uncover. Thinking, "Well, someday, someone will think of a way to apply the concept of a plastic synapse to all those troublesome behavioral findings," is not thinking. It's a refusal to think. Anyway, in response to this comment and your suggestions, I have put this into the new and final section of the Discussion. Theorists have been averting their gaze from the PREE for more than half a century and from the Gibbon and Balsam C/T finding for almost 40 years, even though Rescorla called attention to its importance 30 years ago. And the story they try to tell about Rescorla's famous finding STILL rests on fantasized "trials" or "microstimuli", constructs that ought to make them blush, given their claimed allegiance to neural plausibility.*

Results section:  
Figure 1: The caption doesn’t state the difference between the top 6 panels and the bottom six ones. *Now it does.* I find the changes in scale for the y-axis confusing. *Changes eliminated; all have same scale now.* Also, this masks the “bumps” at the end of the record (as stated in the text), maybe an inset of this would help. *Done* It would also be helpful to mark the onset of each session type (training, extinction…). *We have revised the figure and its caption in accord with these excellent suggestions.*

P. 7 L. 10: There is a strong emphasis on the trials to acquisition measurement (also in the methods section), which could be simply stated as the first time the cumulative account reaches a value higher than the minimum plus 20. *The reviewer is right; we were unnecessarily verbose about this. We have followed his phrasing--in the Statistical Analysis part of the Methods section.*

P. 7 L. 25: I did not understand the sentence “From a content-theoretic…”.   
Figure 2: It might be better to use different symbols rather than colors. *Done* Also it would be helpful to connect the classification used here (e.g. FewSessWithFewTrls) to the one used before (e.g. 3.1). This is done to some extent in the text but it is incomplete and it might be useful to have it in the figure caption. *Now, there are more systematically made connections*

P. 9 L.7. It would be helpful to report the extent of recovery at 7 days.

*We added Figure 1 in response to this suggestion*

Figure 3. As this is the most important figure, it would be useful to label each panel to guide the reader (e.g. rate of acquisition wrt total number of training trials). Also, given the large inter-subject variability I would recommend using boxplots to present these data so the reader would get a better sense of the actual distributions.

*This is a puzzling comment, because these plots all have 54 points, one for every subject, so readers have the actual distributions right in from of them.*     
Regarding the data analysis, it would be reassuring to report also the effect sizes (the reader can not infer it, having no access to the distributions used to compute the t tests). *We have replaced the pooled standard deviation with Cohen's d (a common measure of effect size) in the two panels where there is a significant effect. We note that the BF, that is, the odds ratio is also sometimes used as a measure of effect size.*

I would also think that one needs to correct the significant threshold (bonferroin correction for instance) because the comparisons here are not orthogonal. Finally, I am not sure that the Bayesian statistics used here are necessary given the clear-cut probabilities obtained with the t tests. *Given the p values, a reader sophisticated enough to worry about the Bonferroni correction would/should understand that it is not worth bothering with in this case. Also, one of the many reasons for preferring BFs to p's is that one does not have to worry about that correction.*

Discussion section:  
P. 11 L. 15. There is an extra “here”. *Gone now*  
L. 18. I do not see the point of this paragraph in the context of this manuscript.

*Paragraph removed.*

L. 28: this paragraph is quite confusing and I am not sure whether the authors just state a general opinion (in which case they might consider deleting these sentences) or whether they draw a logic conclusion based on the outcome of their experiments (in which case this should be better explained).

*The Discussion has been entirely rewritten with this and the following comments in mind*P. 12 L. 18 this paragraph should be expanded to better explain your point. In particular, how do you connect the predictive power of a CS and the span of the experiment? It seems that this a critical point for your paper but it is difficult to follow your reasoning here.  
P. 13 L 6: A space is missing (oneuses). *Corrected*  
P. 13 L 7: I understand that short CS with long inter-stimulus intervals are more informative and that larger experimental span increases the stability of the perceived predictive value of the CS. However, the second part of the prediction, regarding extinction, is unclear as stated. Also, it seems that you predict the opposite of what the behavioral-momentum theory proposes. It might be worth stating that and explaining the theoretical differences between these two views on extinction.   
P. 13 L. 20: I am not sure I know what is meant by “persuade the brain”. The authors might want to rephrase this paragraph. *No longer use this phrase*  
  
Methods section:  
P. 14 L. 22. As mentioned before, I do not think that the matlab code is necessary here. *Gone*Table 2. I do not think that the Matlab code is useful here. *Gone*  
  
  
  
  
Reviewer: 2  
  
Comments to the Author  
This is a very interesting set of experiments that make the point that different parameters of training differentially impact acquisition, extinction and spontaneous recovery.  Furthermore, the experiments make the point that span of time over which training occurs can differentially impact these different aspects of learning. There are some points that I think should be considered in a revision.  
  
In the abstract, it says that intersession interval and span of training were manipulated. This is true but the effect of intersession interval is never analyzed. Since span and intersession interval are confounded in some experiments it would be useful to analyze whether span,per se, does have an effect on spontaneous recover that it is not because of intersession interval (if this is true). *The problem is that intersession interval and span covary. Teasing them apart will have to wait for future work*

I am not sure why it is asserted that the computation of stochastic models largely occurs “off line”. Surely, models are being developed and tested during learning. My read of Wilkes and Gallistel is that this is exactly what is happening over the course of training. The models may be refined and perhaps tested off line but I don’t think the emphasis on off line processing is warranted. If it is warranted, it should be more clearly justified. *This now reads*

*"It seems likely that some of the second component—the computation of stochastic models—occurs off line." It's there to lead into the point that this assumption explains consolidation and reconsolidation phenomena. To the best of my knowledge, no other computational theory gives a computationally motivated explanation for these phenomena.*   
In the second paragraph, I think it might be clearer to talk about the extinguished elements rather than the unconditioned elements. *Done*  
You might want to point out that not only do animals use time and number they may even use both to solve the same task (Light et al., 2019). *Done--see Discussion*  
The reference to the “so-far unsolved problem for associative theories” will be obscure to all but the experts. You should explain the unsolved problem.

*That's what the (new) final section of the Discussion does*  
  
I like the consolidated presentation of data across experiments but it was jarring to see the results start with experiment 4.  Perhaps that experiment could just get relabeled as experiment 1 since it does not matter in the current manuscript.  *Done*  
With respect to the number of trials per session, Jenkins et al.’ chapter in the Autoshaping and Conditioning Theory book presents data showing that with a single trial a day birds acquire autoshaped keypecking after 2 pairings! *Grateful to referee for calling attention to this! I did not know the Jenkins et al paper. It's a gem. We now cite it repeatedly.*  
  
It may seem a bit redundant but I think the data exposition would be more easily understood if for each dependent variable there was a plot against trials per session, total training trials and training span. Even though it would be more figures, the reader could easily whether or not there is a relationship.

*We added Figure 6 in response to this suggestion*  
Within session extinction of contextual conditioning is demonstrated in Mustaca, Gabelli, Papini and Balsam, 1991 (see figure 4). *Grateful again. We now have a paragraph devoted to this paper*  
  
Additional evidence that CS extinction depends on number of trials and not duration of CS exposure is found in two papers by Drew (Drew et al., 2004; Drew et al., 2017). *We now cite them*  
Further support for the idea that span of training influences the belief that a particular set of training conditions might prevail comes from a few studies that examined the effects of session spacing in extinction. Tapias-Espinosa, Kádár , & Segura-Torres 2018 and Tsao & Craske (2000) found that giving extinction sessions over a longer span reduced spontaneous recover. This makes the effects of span of extinction very important to the sustained success of exposure therapies. *Tapias-Espinosa, Kádár , & Segura-Torres 2018 are now cited. Tsao & Craske not cited because human subjects, clinical focus, self-report, messy results.*