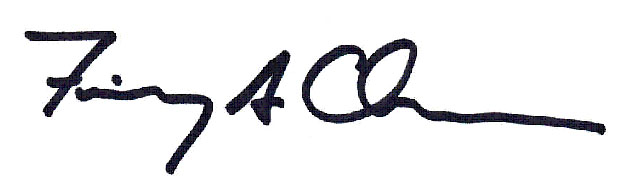
July 3, 2015

Dear Editor,

We appreciate the opportunity to submit a revision of our manuscript, “Habitual control of goal selection in humans”. We are grateful for the additional feedback of the reviewrs, as well as their enthusiasm for our initial revision. This feedback has lead to further improvements to the manuscript, which we summarize below.

We thank you again for your careful attention to this manuscript, and we look forward to your final decision on its suitability for publication in the *Proceedings of the National Academy of Science.*

Sincerely,



Fiery Cushman

Assistant Professor

Department of Psychology

Harvard University



Adam Morris

Department of Psychology

Harvard University

**Reply to reviewers:**

Reviewer 1

*This is a really excellent and thorough revision to an already exciting paper and I congratulate the authors. The new experiments significantly extend the interpretation of the result, ruling out alternatives I actually didn't think were possible to fully address. Just to reiterate, this is an incredibly careful and targeted behavioral attack on the mechanisms of approximate planning in the brain, which is an issue of strong current interest, and goes well beyond previous attempts in this area in its specificity.*

We thank the Reviewer for an enthusiastic endorsement of our revised manuscript.

*(1) It seems like some methodological information might have gotten lost in the reorganization (or I have lost track of it anyway?). I don't see a discrete methods section anywhere and I can't find things like the numbers of subjects and trials, the fact that the subjects were recruited from Turk, an assertion that informed consent was obtained, and the process by which random rewards were generated. I think I have reviewed all that information previously and much of it also appears in the response letter, so it's not an issue for me but of course it should appear in any published version.*

We now include more a complete presentation of the methods for each experiment. These details are distributed across the main text and supporting information consistent with space restrictions. Adam – can you do this?

*(1) I think it's worth briefly comparing to the results of Dezfouli & Balleine (PLoS CB 2013). Does the current model explain their action sequencing effect? I think one key difference is that they are lacking an experiment like Expt 2 here and so can't really substantiate their claim that option selection is model-based.*

Both reviewers urge more discussion of a recent model and associated experimental results by Dezfouli and colleagues. We have revised the manuscript accordingly.

Below we make a number of detailed points for the sake of completeness, but these can be summarized simply: Our theoretical proposal and Dezfouli’s are compatible, and indeed they share the theme that habitual and goal-directed control may be mixed at independent levels of hierarchical control. Our data do not contradict their model, nor do their data contradict our model.

In brief, Dezfouli and colleagues make two arguments. First, they propose model-based control over habit-based options. Second, they propose that habitual control occurs by action sequencing rather than model-free value representation.

Regarding the first claim, of course we provide evidence for the opposite arrangement: habitual control of the selection of model-based options. Consistent with prior versions of our manuscript, we do not regard these proposals as mutually exclusive. To the contrary, we regard both as reflections of a single guiding principle: It is desirable to “[tailor] the means of control (habit vs. planning) to the affordances of a particular level of behavioral abstraction” (i.e., a particular level of hierarchical control). In our prior manuscript we approvingly cited both theoretical and empirical studies indicating habit-based options. In our current revision we add substantial new emphasis to Dezfouli and colleagues’ important work showing that these options may be under model-based control.

As Reviewer 1 anticipates, a novel method we employ in Experiments 2a-b allows us to provide evidence for model-free control over options. Specifically, we construct a task in which a terminal reward distribution is equally available from all Stage 2 states (and thus all Stage 1 actions). In this context, model-based control over options does not predict that rewards from this distribution will influence subsequent choice.

Dezfouli and colleagues’ second argument is that past evidence (e.g. by Daw and colleagues) for a model-free RL account of habits is better accommodated by an action sequencing account. We have revised the present manuscript to note the evidence of Dezfouli and colleagues in favor of the existence of options defined as action sequences.

Reviewer 1 asks whether “the current model explain[s] [Dezfouli and colleagues’] action sequencing effect”. However, we do not believe that *our data or model* provide an alternative explanation of *Dezfouli and colleagues’ data* on action sequencing. Future research might explore an extension of our novel methods to this question, but we regard this possibility as beyond the scope of the present manuscript.

Conversely, does Dezfouli and colleagues’ model of action sequencing provide an alternative interpretation of our data, which indicates genuine model-free control at the highest level of our task? It does not: The concept of “action *sequences*” does not apply to the critical trials of our Experiments 2a-b, which involve novel sets of numbers to be summed.

Our data would contradict the strong claim that *no processes of model-free value assignment exist*, and that *any* such influences can be captured by model-based valuation of action sequences. However, we do not find that Dezfouli and all make such a strong claim. E.g., in their 2013 publication in PLOS CB, they write “… here we assumed.. action sequences are also under goal-directed control. Alternatively, it is possible that the value of any action sequence is learned in a model-free manner… Our results are silent with respect to this latter assumption”.

Our data provides strong evidence for model-free control over options (in this case, options defined by goals), and it cannot be explained by model-based control over options. As we have noted, however, we believe that model-based control over options may be exercised in other contexts, consistent with Dezfouli and colleagues’ model.

*(2) The control for win-stay-lose-shift seems a bit fishy to me since it doesn't appear to control for the reward obtained on the first trial after the setup (which is presumably correlated with the regressor of interest). Personally I don't find this control at all necessary in any case since win-stay-lose-shift is to my mind a bona fide instance of model-free learning (it just arises in the limit as learning rates approach 1).*

Did he just miss the point that we analyzed rewards and losses separately? Or is the point that we need to control for reward *magnitude* within losses and gains? If so, fine, let’s do that. Even if he’s not worried about WSLS I think other readers might be, and since it’s all in the SI there’s not much reason *not* to include it.

Reviewer 2

*The authors have taken care to address many of the issues raised in the last review. However, there are several sticking points for this reviewer.*

*1. On the assertion that model fitting as requested in too difficult. There are several papers from Bernard Balleine that also argue for hierarchy in these two-steps tasks, but they assert the opposite: the highest level of option selection is controlled by a model-based system, and the internal option policies are model-free.*

*Cushman cites one of these papers in passing (ref #32), but doesn't discuss this connection. This is important.*

As discussed above in reply to Reviewer 1, we agree that this past work has an important place in our manuscript and have revised it accordingly.

*Also, Balleine's group does formal model fitting with option models, so the assertion in the first reply that model fitting here is too hard or the data too noisy is hard to understand.*

Fit a model.

*2. What I am mostly puzzled about though is that it seems like they set up a straw man type argument, essentially asking: "is there evidence of model-free behavior at the top level?" (and secondary to that, whether the problem representation is hierarchical), whereas what they want to ask is "is there ONLY model-free behavior at the top level, and no model-based control at all?" This seems pretty silly, so maybe I missed something in my re-reading. In exp 1a/b, the fact that rewarded rare transitions are reinforced is taken as evidence for Model-Free. Yes. But it's not evidence against any possible Model-based influence.*

We agree with Reviewer 2 that Experiments 1a-b do not provide strong evidence of this kind, and we have revised the manuscript to emphasize this point. Importantly, Expeirments 2a-b provide the requisite evidence, as we have noted above in our reply to Reviewer 1.

(Of course, as a general matter, we believe that in different contexts both model-based and model-free control will be exercised at superordinate levels of hierarchical organization. We frame our work in a manner that emphasizes the advantages of “tailoring the means of control (habit vs. planning) to the affordances of a particular level of behavioral abstraction” (i.e., a particular level of hierarchical control). In some contexts this may favor superordinate model-based control (as proposed by Dezfouli and colleagues, among others), while in other contexts this may favor superordinate model-free control (which is the focus of our manuscript).)

*3. Then there is the secondary question of whether the problem representation is hierarchical and exactly where the option boundaries are, which exp 2 is meant to address. The focus is on what constitutes a "goal state". However, the other extremely important piece of what constitutes an option -- the the option policy -- is completely trivial. It's not even a two-step problem, it's one step: 1,3 maps to blue state, 2,4 to red state. The experiment, as designed, can't speak strongly as to whether or not the same option policies are being invoked in each instance. I point these out because this was not highlighted by the authors in either version of the ms.*

Indeed, it is important to understand the nature of the intra-option policies. In this comment, Reviewer 2 brings up Experiment 2 initially and then details a more specific critique of Experiment 1. We agree with the critique of Experiment 1 and, crucially, Experiment 2 is designed to address it. Specifically, regarding the question of whether “the same option policies are being invoked in each instance”, Experiment 2 provides strong evidence against this possibility. This is because the actions necessary to execute an option are entirely novel in Experiment 2 (i.e., a novel set of three numbers than can be summed to 16 or 21). The use of fully novel actions requires that the intra-option policy be derived by model-based methods, and it precludes the possibility that the *same* option policies are being invoked across instances.

*4. In the SI, why do t-tests on model output which can be made arbitrarily precise by running a larger number of simulations? This SI is more akin to a power analysis than actually testing the predictions of the models. Unless I'm missing something here, the results speak against them because even without model-free goal learning the difference between the quantities of interest are close to "significance" at 0.1.*

I hadn’t really focused on it, but it is a little weird that we get a marginal effect without MFonMB. What is that all about? If we run the simulation to death, does it go away?

*5. SI equation 4 is a bit confusing. I understand they want to unroll the tree, but what is "a"? In the equation it says it is the action set which leads to a goal state (so actions at the first stage, since goal states are at the second stage?), but above it says "from each stage 2 action a".*

Adam – can you find a way to make this notation more clear?