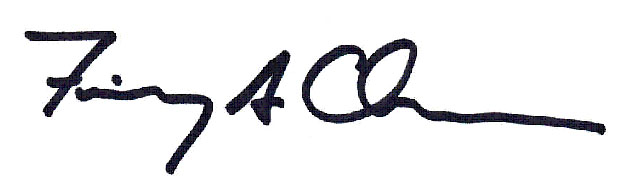
July 3, 2015

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 a dedicated section with a more complete presentation of the methods for each experiment.

*(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.

One key element of Dezfouli’s proposal is that there may be goal-directed control of habitual “options” (i.e., temporally extended sequences of actions). We, of course, provide evidence for the opposite arrangement: Habitual control of goal-directed planning. 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 suggestion that these options may be under model-based control.

The second key element of Dezfouli’s and colleagues’ proposal is that humans make use of “action sequences”, in which sets of actions are chunked together into a single fixed routine available for selection by a higher-level controller. The revised manuscript explicitly notes this aspect of their work as well. Reviewer 1 asks whether we provide an alternative account for Dezfouli and colleagues’ action sequencing model. Although our methods might be adapted for this purpose, we believe that the current evidence provided is not dispositive. In any event, we regard this possibility as beyond the scope of the present manuscript, and better left to future research. Our purpose is not to address the sufficiency of prior evidence for action sequencing, but rather to provide new evidence in favor of a model of habitual goal selection.

*(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).*

We thank the Reviewer for raising this concern. We have followed the Reviewer’s advice and removed this section from our manuscript.

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.*

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

*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.*

We agree with the Reviewer that our argument would be strengthened by fitting our formal model to behavioral data. We have conducted this analysis, and the results strongly support our conclusions. The analysis is presented in our revised manuscript.

*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.*

As a general matter, we believe that in different contexts both model-based and model-free control are exercised at multiple levels of hierarchical organization. Thus, 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).

We concur with the reviewer that we provide strong evidence for an influence of model-free value representation at the highest level of control in our task. We also concur that we do not provide evidence against the possibility of some *additional and independent* model-based influence. Our experiments were not designed to test this possibility. The interesting, novel and important conclusion of research is that habitual control over goal selection in humans exists, not that it is the exclusive mechanism of goal selection.

*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.*

We agree with the reviewer that it is important to understand the nature of the intra-option policies. In this comment Reviewer 2 briefly mentions Experiment 2, but then presents a critique specific to Experiment 1. We agree with this critique of Experiment 1. Crucially, Experiment 2 answers 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.

Motivated by the reviewer’s suggestion, we have edited our discussion of the limitations of Experiment 1 for clarity.

*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.*

We agree with the Reviewer’s first concern. We have removed the t-tests on model output, and instead report means and standard errors. In order to address the Reviewer’s concern about a marginally significant result in the absence of model-free goal learning, we reran the simulations with 2.5x more agents per simulation. The results, reported in our revised manuscript, more clearly validate our methodological approach. Our analysis detects a large effect in the presence of model-free goal learning, but no effect whatsoever in its absence.

*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".*

Our revised manuscript presents clearer notation and exposition in this section of Supporting Information.