Dear Editor,

Thank you very much for reviewing our manuscript, “Latent structure in random sequences drives neural learning toward a rational bias” (PNAS MS# 2014-22036), and giving us the opportunity to revise it. We sincerely appreciate the constructive comments you and all the reviewers have provided.

The manuscript has been revised carefully. In summary, two major areas are revised to address the concerns by Reviewer 2. First, we emphasize that our model is not designed to merely mimic human behavior by arbitrary parameter fitting. Instead, our results are from strong predictions of the model, based on a detailed analysis of the learning environment and a biological understanding of the neural mechanism of temporal integration. Second, we expand the points regarding the connection between the neural model and Bayesian models, by clarifying the analytical procedures leading to the new parameter interpretations then addressing the theoretical implications. And, the revision puts more emphasis on a unified perspective regarding the implicit statistical learning and rule-based Bayesian inference. (Please see detailed point-by-point responses attached in the following).

We hope that you find the revised manuscript much improved and publishable.

Sincerely,

Yanlong Sun, Randall C. O'Reilly, Rajan Bhattacharyya, Jack W. Smith, Xun Liu, Hongbin Wang

=============================

Point-to-point responses (Reviewer 2’s comment points are quoted in italic. Locations of corresponding revisions are enclosed in brackets):

“*In the first part of the manuscript, the authors offer up a simple neural network model, trained on sequences with varying probabilities of successes, and demonstrate that, given their architecture, the model mimics human behavior via sensitivity to alternating and repeating sequences. Given the inherent flexibility of this type of neural network model, I am not surprised that one can be constructed to behave in this way. It's not clear to me, as a decision theorist, exactly what has been learned from this exercise. Other than a demonstration of the rich types of behavior that a neural network model can mimic, what exactly is to be learned and interpreted via human behavior?*”

In this revision [page 2, paragraphs 2 and 3], we emphasize that this neural model is not designed to merely mimic certain types of human behavior by arbitrary parameter fitting. Instead, based on a set of biologically-plausible parameter values that have been tested with a wide variety of tasks, we make a strong prediction that the model should be able to naturally capture the waiting time statistics of temporal patterns. We also note that in Supporting Information, we have shown an extensive exploration of the model parameters, and the alternation bias was a robust result in a wide range of parameter values. Overall, these results demonstrate that the alternation bias and the gambler’s fallacy emerge naturally as the consequence of the neural mechanism of temporal integration reacting to the environmental statistics, rather than arbitrary parameter fitting.

“*This exercise, in and of itself, does not constitute a rationalization. For that to occur, the model would have to have been derived according to rational principles, axioms, etc. I may be missing something here, but I don't see what the impact is for decision theory other than a demonstration of model fit.*”

In this revision, we emphasize that the model was indeed reacting to a set of “rational principles”. Specifically, in the introduction section [page 1, paragraph 2], we highlight that the waiting time is in effect a measure of the variance of pattern inter-arrival times. The rational principle is further demonstrated regarding the connection between the neural model and Bayesian models (see responses below). In the Conclusion section [first paragraph], we emphasize that both mean time and waiting time are normative measures of the learning environment, and “a rationalization” of human behavior depends on the criterion of the particular measuring device.

“*In the second part of the manuscript, the authors connect their approach to existing Bayesian models of random sequence production, offering new parameter interpretations. I found this to be more interesting, but to rigorously evaluate whether this is a viable interpretation of the modeling parameters would require additional data and carefully designed tests to answer this specific question.*”

We have significantly expanded this section of the manuscript:

First, we address to the general audience that the theoretical motivation behind this set of analyses is to provide “a unifying perspective on human statistical learning” and to bridge the gap between the implicit learning without instruction, and the generalization of the learned patterns under structured and expressive rules [page 2, right column, first paragraph after section title].

Second, we clarify the analytical procedures leading to the new parameter interpretations. In greater detail, we show the derivation of the Bayesian model (starting from a “rational” belief function) [around Equation 3], and offer an interpretation of the parameter lambda in the original Bayesian model (i.e., as a free scaling parameter that adjusts the balance between the numbers of heads and tails) [page 3, first paragraph; and the two paragraphs beneath Equation 5]. Then, we show the mathematical equivalence from the perspective of the neural model and offer a new interpretation of this parameter (i.e., as a bias-gain parameter that modulates the alternation bias) [around Equations 6 and 7].

Third, we address the theoretical implications of the parameter lambda. We show that the new derivation is consistent with the normative measures of pattern time statistics, and argue that it provides a quantitative connection between different learning mechanisms and different modeling approaches [page 3, after Equation 7].

Finally, we agree with Reviewer 2 that additional data and tests would present a stronger case to support our claims. However, we also note that the empirical data examined in this study represent a strong consensus in the literature on human randomness perception. The result of the subjective probability of alternation came from a comprehensive review of existing studies (Falk & Konold, 1997), and the random sequence production data (Goodfellow, 1938) came from a massive database, which to our knowledge is the largest one in the field. Moreover, the neural model is based on a biologically-motivated computation framework, and the model parameters we used here have been tested with various tasks in a wide range of different domains (O’Reilly, et al., 2012). In this revision, we highlight these points where the corresponding studies are cited.