

RNNs develop history biases in an expectation-guided two-alternative forced choice task

Manuel Molano-Mazon (molano@clinic.cat)
IDIBAPS, Barcelona, Spain

Guangyu Robert Yang (robert.yang@columbia.edu)
Center for Theoretical Neuroscience, New York, USA

Ainhoa Hermoso-Mendizabal (ainhoahem@gmail.com)
IDIBAPS, Barcelona, Spain

Jaime de la Rocha (jrochav@clinic.cat)
IDIBAPS, Barcelona, Spain

Abstract

Understanding how expectations bias perceptual decisions constitutes an unavoidable step towards deciphering how we make decisions. Here, we trained Recurrent Neural Networks (RNNs) in a novel two-alternative forced-choice (2AFC) task where both the current sensory evidence and the recent trial history provide information about the identity of the correct choice. We found that RNNs learned both to integrate the stimuli and to capitalize on the serial correlations of the trial sequence by developing history biases. Interestingly, during early stages of training, all networks reset their biases after an error response, which is consistent with data from rats performing the same task. At later stages of the training, approximately half of the networks moved from this initial, sub-optimal, strategy and developed after-error biases. A more detailed characterization of these different behaviors revealed that the percentage of networks showing after-error reset could be increased by limiting the resources of the networks, such as reducing their size, the information they receive or the training time. Together, these results suggest that rats develop a sub-optimal but easier to reach strategy to solve the task due to some limiting factor such as lack of computational capacity or time constraints.

Keywords: Expectations; Recurrent Neural Networks; Perceptual decision making

Introduction

The role of expectation biasing perceptual decisions has been extensively studied in human psychophysics in the context of two-alternative forced-choice (2AFC) tasks (Summerfield & De Lange, 2014). However, despite multiple psychophysical studies and considerable theoretical work (Ratcliff & McKoon, 2008), the way the brain flexibly combines past trial history with incoming stimuli to make statistically informed decisions is still not fully understood.

Here we trained Recurrent Neural Networks (RNNs) on a 2AFC task that requires the categorization of stimuli that are

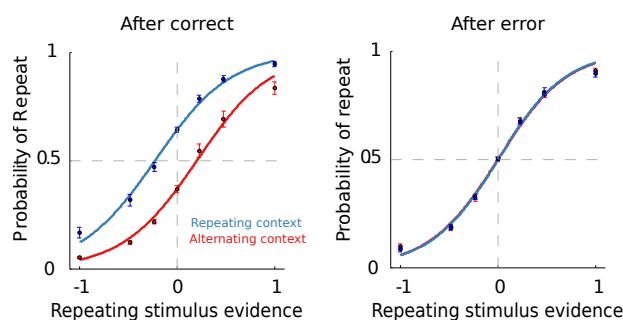




Figure 1: Rats develop after-correct transition biases. Psychometric curves show the probability that rats repeat the previous choice as a function of the evidence of that choice, after-correct (left) and after-error (right) trials (Hermoso-Mendizabal et al., 2019).

presented in a sequence exhibiting serial correlations. In particular, the probability, P_{rep} , that the correct answer at trial t is the same as at trial $t - 1$ alternates between $P_{rep} = 0.8$ (repeating block) and $P_{rep} = 0.2$ (alternating block), every 200 trials (Hermoso-Mendizabal et al., 2019):

Repeating block ($P_{rep}=0.8$) 
Alternating block ($P_{rep}=0.2$) 
right trials left trials

This setup allowed us to investigate how RNNs can integrate decision-relevant information present at different temporal scales: a fast source of information, the current stimulus, and a much slower one, the trial-to-trial correlations, that can be interpreted as the context in which the current trial is perceived. We found that RNNs developed a trial-history bias (transition bias, b , see Methods): a tendency to repeat ($b > 0$) or to alternate ($b < 0$) the previous choice depending on the number of previous repetitions vs. alternations. We further characterized this transition bias by separating trials depending on the outcome of the last trial (b^+ and b^- for after-correct/error biases, respectively) and found that trained net-



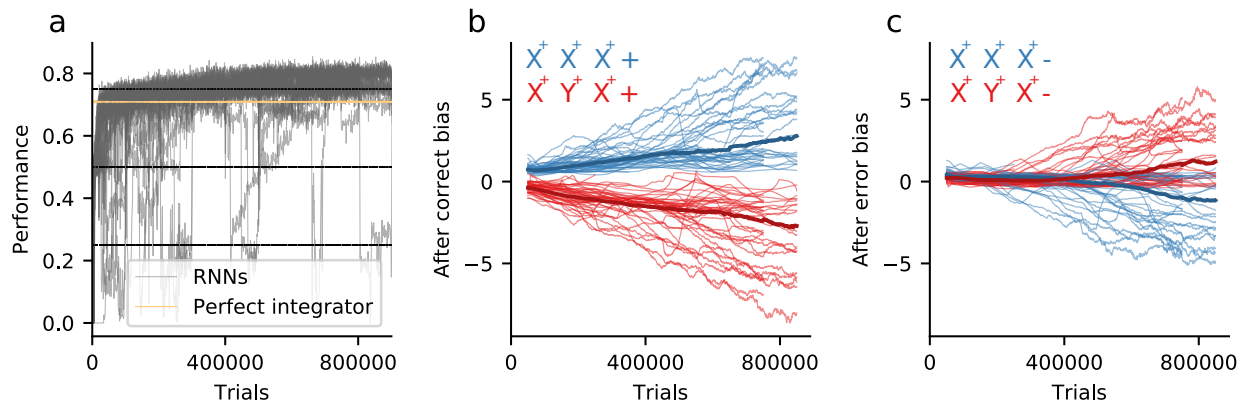


Figure 2: Performance and transition bias of RNNs trained on the 2AFC task. **(a)** Performance of 100 randomly selected networks across training (gray lines), compared to that of a perfect integrator that only uses the current stimulus to choose between the two sides (yellow line). **(b) and (c)** After-correct/error transition biases across training, for Repeating (blue) and Alternating (red) contexts (see main text). Thick lines correspond to the median values.

works follow two different strategies to solve the task: one in which the transition bias is present after correct responses but vanishes after error trials ($b^+ \gg b^- \approx 0$), as has been found in rats (Fig. 1) (Hermoso-Mendizabal et al., 2019); and another strategy in which networks show a transition bias after correct and after error of comparable magnitude but with an opposite sign ($b^+ \approx -b^-$). Interestingly, the former strategy characterizes the behavior of all networks during the first stages of training and only at a later stage, a sub-population of networks are able to move to the second strategy, which has a positive impact in their performance. We further characterized these two types of solutions found by the RNNs by varying the size of the networks and the amount of information they receive.

RNNs develop history transition biases

We trained 217 16-unit networks in the 2AFC task described above (see also Methods). Fig. 2a shows the performance of 100 randomly selected networks across training. For comparison, the performance of an ideal observer that perfectly integrates the current stimulus information but it is blind to trial history is also shown. The performance of most networks goes beyond that of the ideal observer, which indicates that RNNs are able to leverage the context information.

To investigate the extent to which RNNs are using the information provided by the recent trial history, we computed their transition biases separately for Repeating and Alternating contexts. We conditioned on the trials preceded by two correct transitions, i.e. two correct repetitions or two correct alternations, to make sure the network had experience the corresponding statistics of each context (e.g. that should display a tendency to repeat in the Repeating context). Then, within each context, we separated trials following a correct response (+) from error trials (-) (see legends in Fig. 2b and c). The reason to separate after-correct/-error trials was to test the extend

to which the RNN had developed a flexible adaptive strategy in each context. An ideal agent behaving on e.g. a repeating context, should be biased to repeat its previous choice after correct responses and to alternate it after errors. Thus, to avoid that after-correct/-error transition biases cancelled each other we separated trials depending on the previous outcome.

We found that on average, all history biases increased with training (Fig. 2b, c, thick lines). However, this increase was not symmetric between after-correct and after-error conditions: b^- grew at a slower pace (Fig. 2c). Interestingly, this slow learning was mainly due to some of the networks never learning to reverse their biases after an error trial, thus presenting a form of after-error resetting reminiscent of what has been found in rats performing the same 2AFC task (Fig. 1, right panel) (Hermoso-Mendizabal et al., 2019). This after-error bias resetting was present for all networks during early stages of training, and it was only after this initial period that some of the networks developed the capacity to reverse their bias after making a mistake (Fig. 3c). This could indicate that after-error resetting constitutes a solution that, although sub-optimal, is easier to reach and corresponds to a local minimum in which some networks remain throughout the entire training.

To investigate how the two strategies explained above emerge, we separated the networks in three different sub-groups, depending on their transition bias at the end of the training (see Fig. 3c, inset): **Reset networks** were those that presented only after-correct biases ($|b^+| > 0.5$ and $|b^-| < 0.5$) (orange points) (see Fig. 3a); **Reverse networks** showed both after-correct and after-error biases ($|b^+| > 0.5$ and $|b^-| > 0.5$) (green points) (see Fig. 3b); **Null Networks**, showed almost no bias ($|b^+| < 0.5$ and $|b^-| < 0.5$) (gray points). We found that the percentage of Reset and Reverse networks was very similar and much higher than that of the Null networks (Fig. 3c, inset).

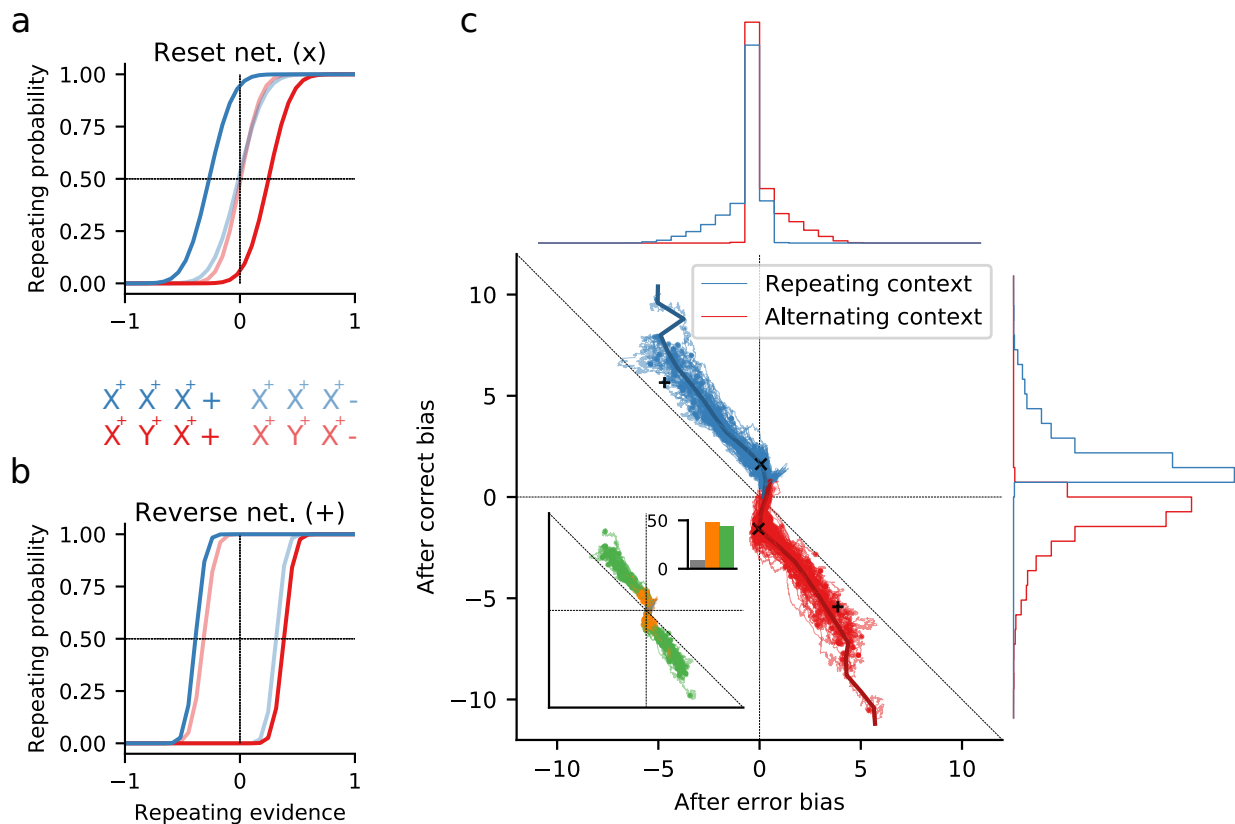


Figure 3: RNNs developed different types of strategies. **(a)** and **(b)**: Psychophysical curves examples for a Reset (top) and a Reverse (bottom) network. Legend is shown between the two panels. **(c)** Evolution of the after-correct VS after-error transition biases during training (thick lines correspond to the median values). Histograms show the distribution of after-error (top) -correct (right) history biases. Inset shows the same data colored by network type (gray: Null; orange: Reset, green: Reverse) and the percentage of each type (bars).

History bias are smaller for limited networks

Does the magnitude of the different history biases depend on the capacity of the network? To answer this question we trained RNNs of different sizes on the 2AFC task. We found that transition biases and the percentage of Reverse networks grew with the size of the network (Fig. 4a), indicating that capacity was an important factor curtailing the ability of the networks to develop after-error biases.

We then investigated how the extra information passed to the networks (previous action and reward, see Methods) influenced our results. Receiving information about the previous reward was essential for the networks to develop history biases (Fig. 4b), and the percentage of Reverse networks greatly increased when this information was provided. Taken together, these results seem to indicate that the Reset strategy is the preferred solution when a network has limited resources, be it short training time, lack of capacity of the network or limited information about the environment.

Methods

All networks were gated Recurrent Neural Networks (RNNs) (Song, Yang, & Wang, 2017) and were trained using standard supervised learning techniques. Trials were composed of 4 steps (1 x fixation + 2 x stimulus + 1 x decision). At each step, networks received as input the fixation cue, a stimulus, which was formed by two fluctuating streams drawn from two Gaussian distributions with different means, and the reward and action from the previous step (Wang et al., 2018). The network had to choose between 3 actions: fixate, respond left or respond right. At the decision step, the network should choose the action (left/right) associated with the stimulus with larger mean. Transition biases were obtained from the fitting of a 2-parameter probit function to the proportion of repeating choices as a function of the repeating stimulus evidence, x , defined as the stimulus evidence multiplied by the sign of the previous response (Hermoso-Mendizabal et al., 2019):

$$f(x) = 1/2(1 + \operatorname{erf}(\frac{\beta * x + b}{\sqrt{2}}))$$

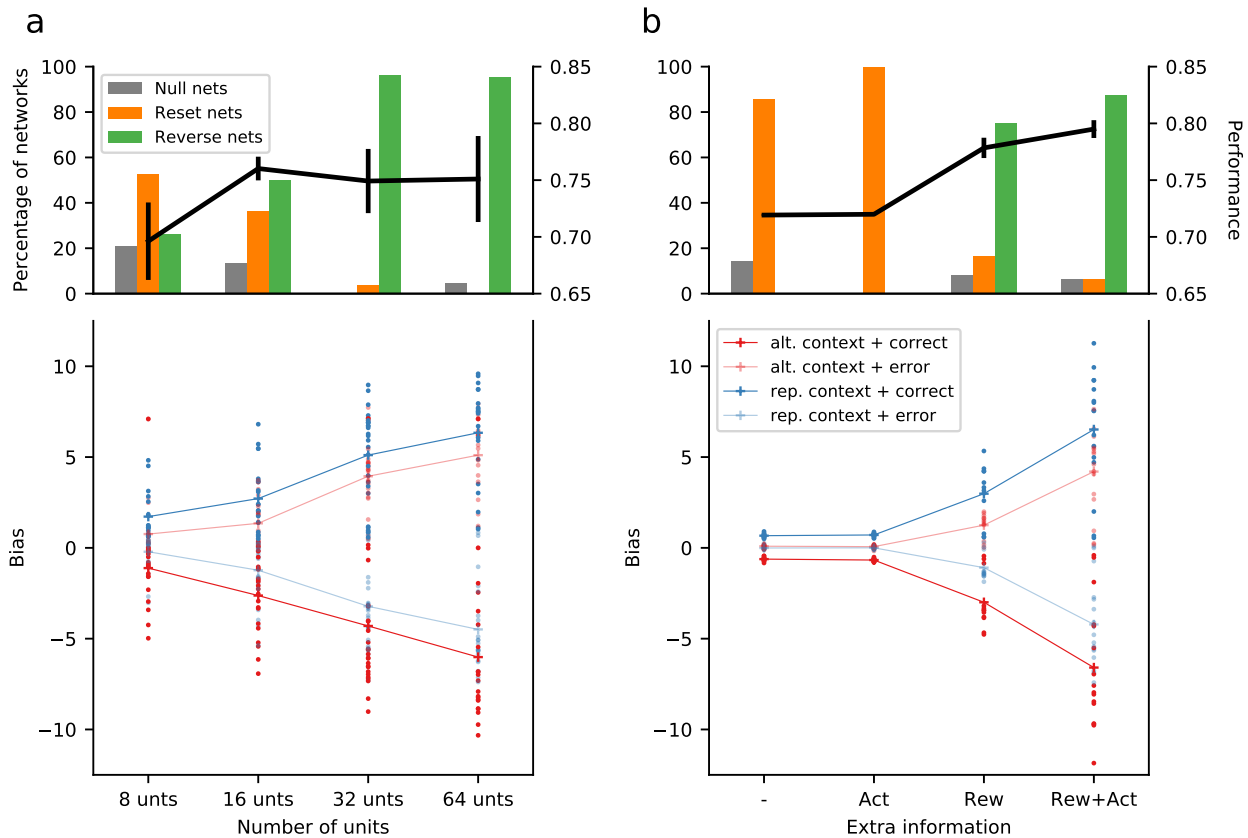


Figure 4: Performance, percentage of network types and transition biases as a function of: **(a)** The network size (number of units); **(b)** The amount of extra information provided to the networks. **Top:** bars show the percentage of Null, Reset and Reverse networks (gray, orange and green, respectively). Black line shows the average performance of all networks for the different configurations (errorbars indicate standard error). **Bottom:** Transition biases for the different configurations. Each trained network contributes 4 points: repeating/alternating context + correct/error (see methods).

Where β is the stimulus sensitivity of the network and b represents the transition bias of the network, i.e. the prior expectation towards repeat ($b > 0$) or alternate ($b < 0$).

Acknowledgments

Generalitat de Catalunya (grant 2017-BP-00305) and the European Research Council (ERC-2015-CoG 683209-PRIORS).

References

Hermoso-Mendizabal, A., Hyafil, A., Rueda-Orozco, P. E., Jaramillo, S., Robbe, D., & de la Rocha, J. (2019). Response outcomes gate the impact of expectations on perceptual decisions. *bioRxiv*, 433409.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation*, 20(4), 873–922.

Song, H. F., Yang, G. R., & Wang, X.-J. (2017). Reward-based training of recurrent neural networks for cognitive and value-based tasks. *Elife*, 6, e21492.

Summerfield, C., & De Lange, F. P. (2014). Expectation in perceptual decision making: neural and computational mechanisms. *Nature Reviews Neuroscience*, 15(11), 745.

Wang, J. X., Kurth-Nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Leibo, J. Z., ... Botvinick, M. (2018). Prefrontal cortex as a meta-reinforcement learning system. *Nature neuroscience*, 21(6), 860.