Episodic memory contributions to working memory-supported reinforcement learning

Soobin H. Hong*[3], Amy R. Zou*[1], Aspen H. Yoo[1,2], and Anne G. E. Collins[1,2]

[1]Department of Psychology, UC Berkeley

[2]Helen Wills Neuroscience Institute, UC Berkeley

[3]Cognitive Science Program, UC Berkeley

Author Note

*These authors contributed equally to this work.

This study was not preregistered. Data will be made available on OSF at publication; study materials and analysis code will be available upon request.

Correspondence concerning this article should be addressed to Anne G. E. Collins, Department of Psychology, University of California, Berkeley, 2121 Berkeley Way, 94720, CA, United States. Email: annecollins@berkeley.edu.

Here is the link to the OSF: osf.io/preprints/psyarxiv/64hxe. We also have our data publicly available on https://osf.io/cew9n/.

Appendix

Supplemental Figures

Figure A1

Experiment duration before participant exclusion. The figure highlights a cluster of participants with experiment duration time <1h, and a long tail of outliers who took much longer breaks between blocks, making test phases not comparable, and were thus excluded from further analyses.

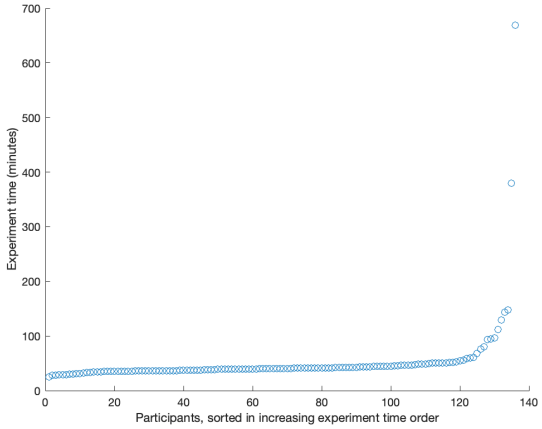

Figure A2

Confusion matrices for model recoverability and identifiability. For a given simulated model data, we find the proportion of agents where each model best fits the data, denoted by rows of the confusion matrices. For example, simulating model RL (y-axis)–recovering model RL2a value (x-axis) corresponds to the proportion of RL model generated data that is best explained by the RL2a model. Each row sums up to 1. A) Model recovery with randomly generated parameters. RL2a model was best recovered by the RL model when random parameter values were used. This was likely due to the models sharing most of the parameters. RL2a model converges to the RL model when α_{pos} and α_{neg} values are similar. B) Model recovery with fitted parameters for each model simulation. RL2a model was best recovered by the RLEM-SAR model when fitted parameter values were used. This was also likely due to the models sharing most of the parameters. RLEM-SAR model converges to the RL2a model when w value is close to 1.

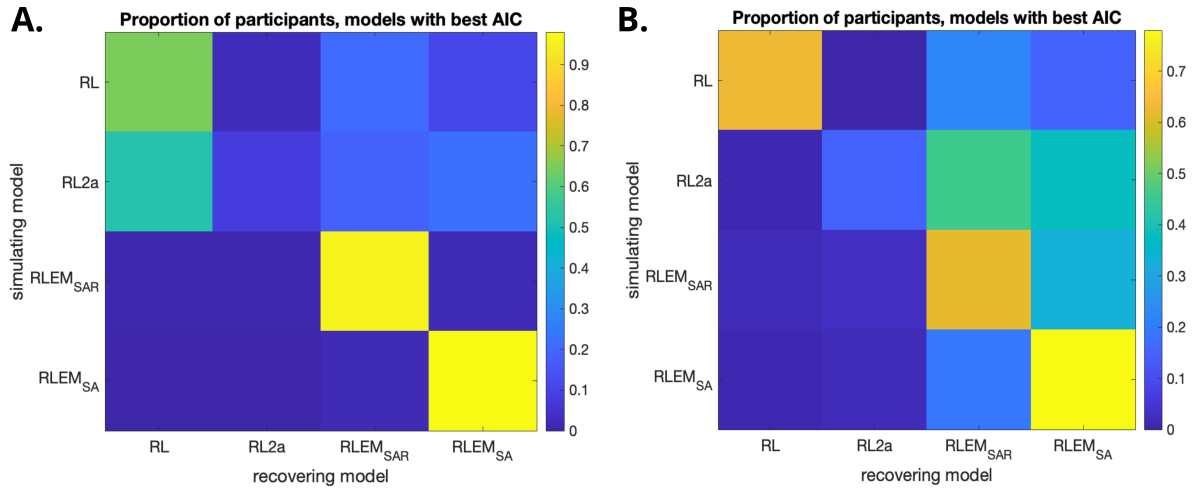

Figure A3

Fitted parameter values for the winning model in experiment 1: RLEM-SA. The level of parameter fit was individual participants; in other words, each point on the figure represents a fitted parameter value of an RLEM-SA model to an individual participant data. The inset, a zoomed in version of the main plot, was added to visualize the fitted parameter values for α_{pos}, α_{neg}, w, f, and st better as they have relatively low values compared to β and β_{test}. The y-axis of the inset is identical with the main figure; it represents parameter values.

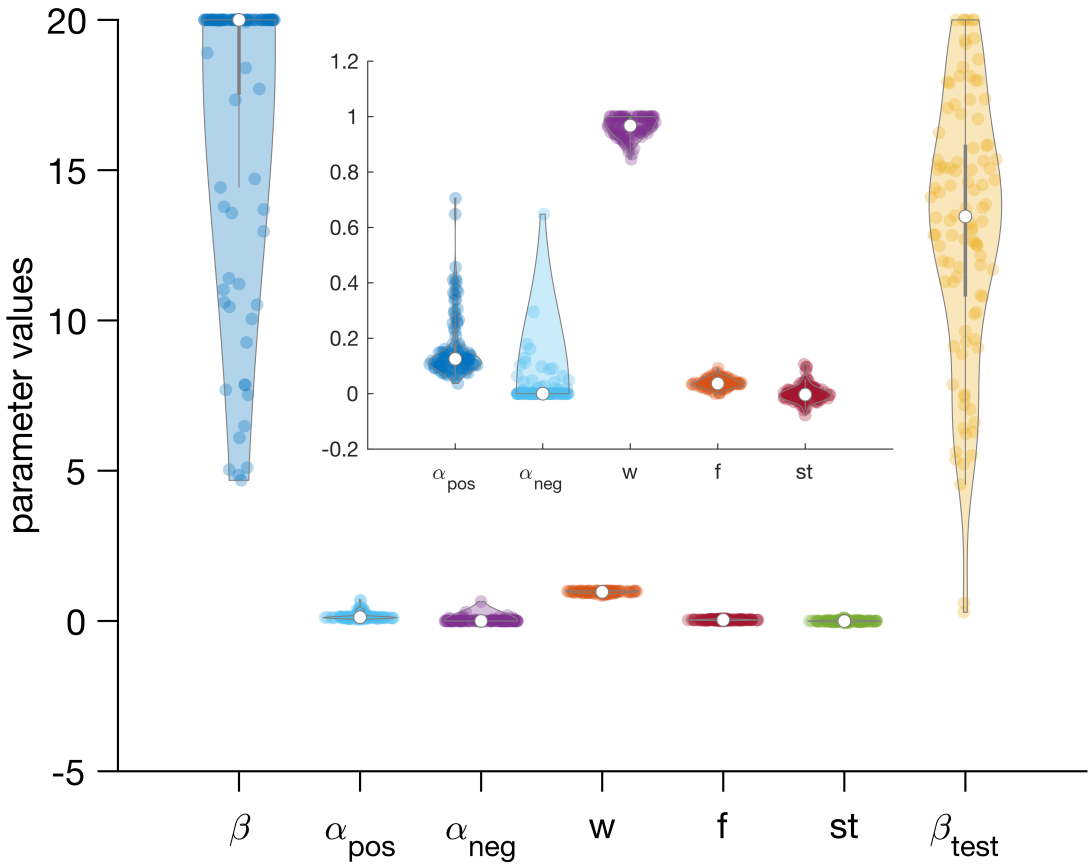

Figure A4

Generate and recover plot for the RLEM-SA model. We randomly chose parameters of the RLEM-SA model within their possible ranges, generated data with the chosen set of parameters, and then recovered parameters that best describe the generated data. We performed such generate and recover steps for 100 sets of parameters, identical to the number of participant data analyzed.

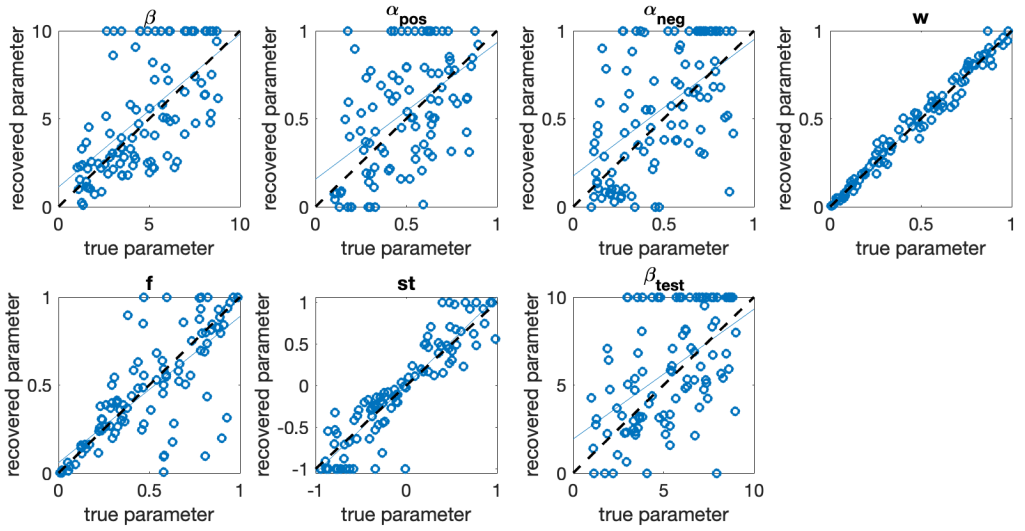

Figure A5

A) Block effects on learning phase accuracy (orange; $\beta = 0.0605, p < .001$) and testing phase accuracy (blue; $\beta = 0.0275, p < .001$). Performance on testing phase trials correctly answered in learning indicated by a solid line, incorrectly answered by a dashed line. B) Effects of within-block temporal order (learning trial) and initial feedback received (FB) on learning phase choice accuracy. FB had a significant main effect ($\beta = 0.592, p < .001$) while learning trial's main effect approached significance ($\beta = 0.063, p = .0814$). C) Testing phase behavior looking at the effect of including vs excluding block 16 on accuracy. D) Effects of within-block temporal order (learning trial) and initial feedback received on testing phase reaction times. Learning trial itself was not a significant predictor ($\beta = -0.068, p = .981$), but FB and its interaction with learning trial were ($\beta = -37.0, p < .001$ and $\beta = 8.13, p = .039$, respectively). Block was not a significant predictor ($\beta = -0.504, p = .303$).

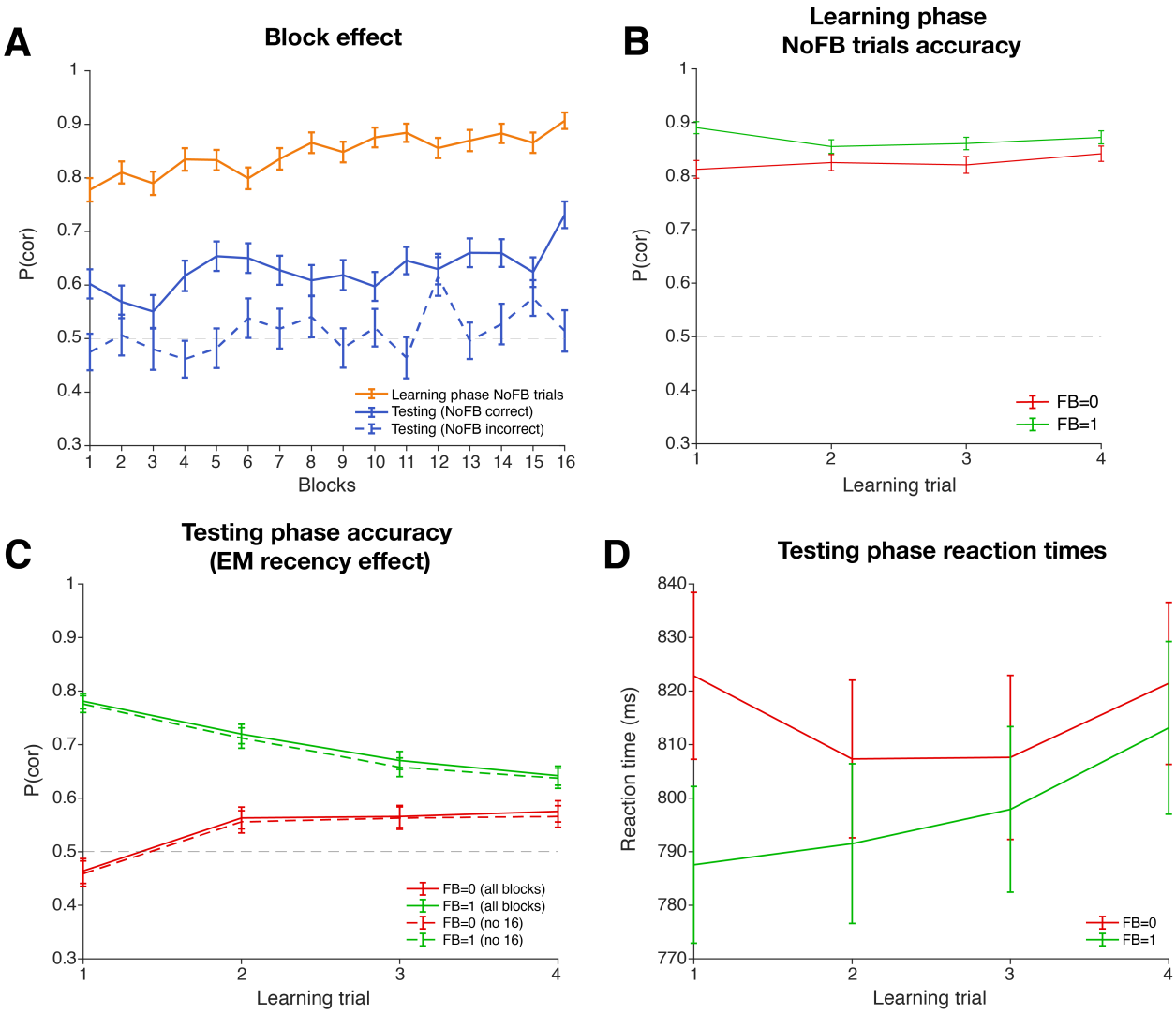

Figure A6

Behavioral analyses of final included participants (left) vs. participants excluded only by RT criteria (middle) vs. participants excluded by more stringent criteria (RT or spammed blocks). A) Testing phase accuracy as a function of temporal order (learning trial) and initial feedback received (FB). When we include participants excluded by RT criteria, predictors values were consistent with the main analysis (block: $\beta = 0.0247, p < .001$*; learning trial:* $\beta = 0.0891, p < .001$*; FB:* $\beta = 1.29, p < .001$*; learning trial* $\times$ *FB :* $\beta = -0.261, p < .001$*). This was also the case when we further included those excluded by the spammed blocks criterion (block:* $\beta = 0.0254, p < .001$*; learning trial:* $\beta = 0.0879, p < .001$*; FB:* $\beta = 1.29, p < .001$*; learning trial* $\times$ *FB :* $\beta = -0.251, p < .001$*). B) Primacy effect for initially unrewarded trials (red) and initially rewarded trials (green). When we included participants excluded by RT criteria, the slopes for initially rewarded vs. initially unrewarded trials were different from each other (*$t(197) = -5.45, p < .001$*, 95% CI* $[-0.3587, -0.1680]$*) and from zero (initially rewarded:* $t(200) = 5.33$*, 95% CI* $[0.0934, 0.2032]$*; initially unrewarded:* $t(199) = -3.34$*, 95% CI* $[-0.1856, -0.0478]$*). This was also the case when we further included those excluded by the spammed blocks criterion: the slopes for initially rewarded vs initially unrewarded trials were different from each other (*$t(213) = -5.56, p < .001$*, 95% CI* $[-0.3558, -0.1694]$*) and from zero (initially rewarded:* $t(223) = 5.25, p < .001$*, 95% CI* $[0.0899, 0.1979]$*; initially unrewarded:* $t(215) = -3.44, p < .001$*, 95% CI* $[-0.1819, -0.0495]$*).*

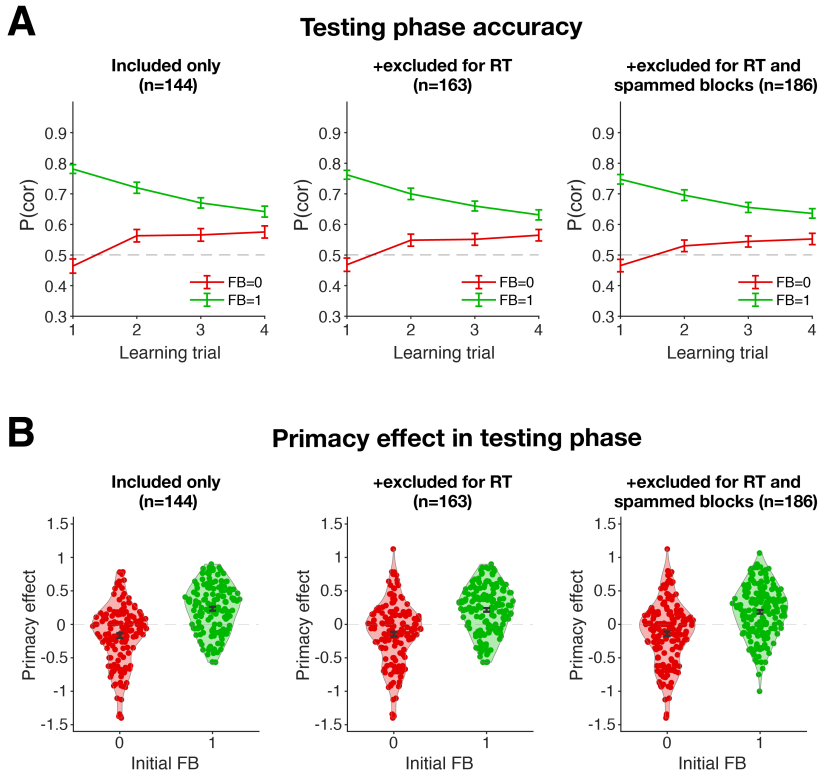

Figure A7

Parameter recovery for the full model in Experiment 2 shows that true parameters used to simulate synthetic datasets are well recovered when fit by the same model.

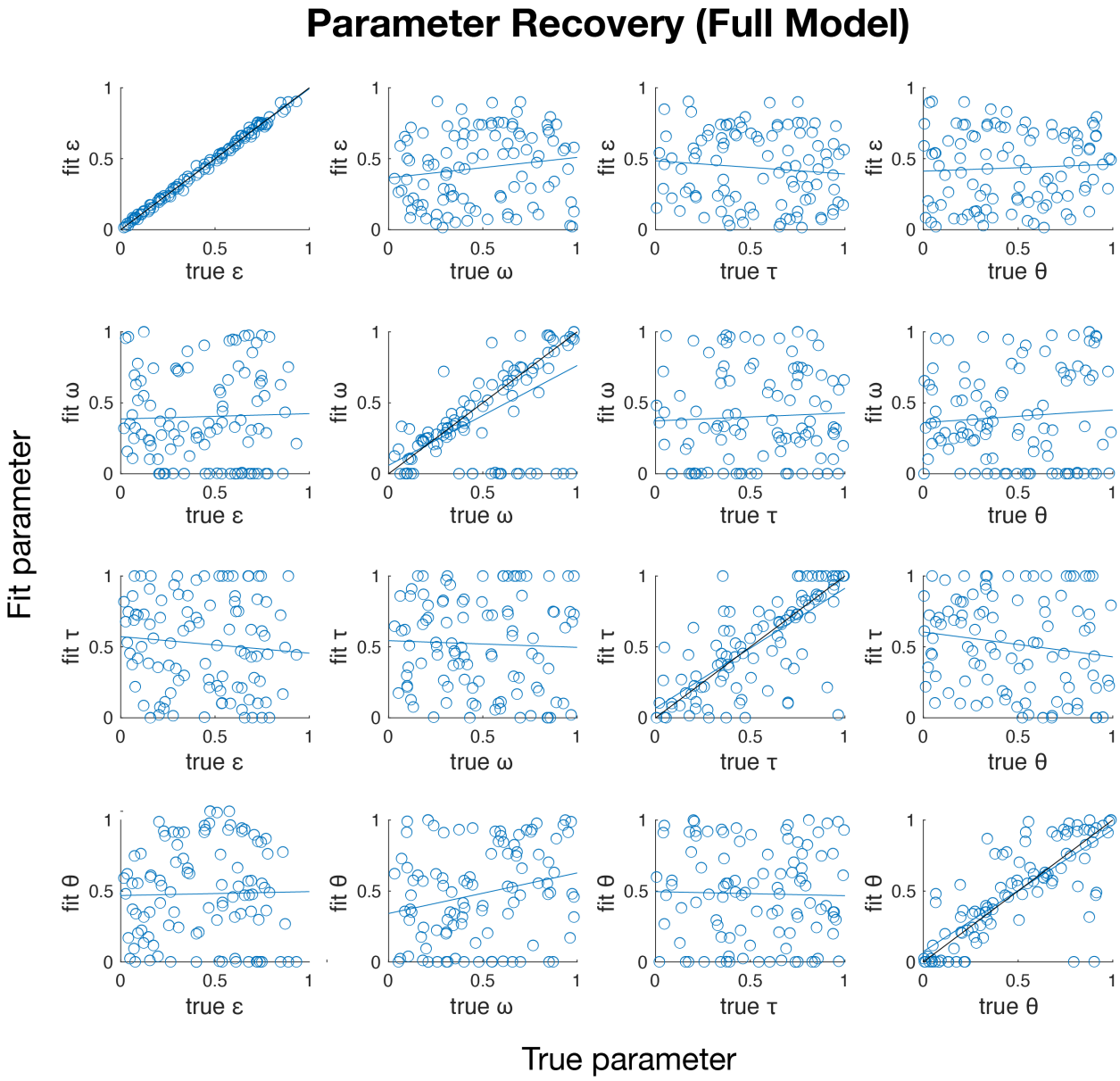

Figure A8

Validation of the model comparison procedure for Experiment 2. A) Difference in out-of-sample -LLH between all models and target model across 50 iterations (dots). While the generating model systematically has the best LLH, other models produce comparable LLH—this is expected because of the nested nature of the models and the parameter fitting on a large dataset (12 folds with 12 participants of 64 trials each = 8448 trials), leading to no overfitting in even higher-parameterized models. B) To ensure that models are nevertheless identifiable, we show that the full model, when fit to data generated to a nested model, identifies the corresponding fixed parameter value to a very close value (e.g., high fit θ when fit to data simulated with fixed $\theta = 1$); and crucially, that the full model parameters fit to the participants' data are outside of the range expected for the reduced models.