

# Contents

<b>1</b>	<b>Results</b>	<b>6</b>
1.1	Learning Performance . . . . .	7
1.2	Computational Models . . . . .	7
1.2.1	Model Comparison . . . . .	9
1.2.2	Model Validation . . . . .	11
1.3	Recognition Memory Results . . . . .	12
1.4	Memory as a function of model-derived PE . . . . .	12
<b>2</b>	<b>Discussion</b>	<b>17</b>
<b>3</b>	<b>Methods</b>	<b>21</b>
3.1	Participants . . . . .	21
3.2	Materials . . . . .	22
3.3	Design and procedure . . . . .	22
3.4	Computational Models . . . . .	24
3.4.1	Action Selection . . . . .	27
3.4.2	Parameter Recovery . . . . .	27
3.4.3	Model Recovery . . . . .	27
3.5	Parameter Estimation and Model Comparison . . . . .	28
3.6	Statistical Analysis . . . . .	29
<b>4</b>	<b>Supplemental Material</b>	<b>1</b>
4.1	Delta-rule formulation . . . . .	1

# Prediction Error and Episodic Memory - Insights from Computational Models

Francesco Pupillo<sup>1</sup>, Javier Ortiz-Tudela<sup>1</sup>, Rasmus Bruckner<sup>2</sup>, and Yee Lee Shing<sup>1</sup>

<sup>1</sup>Goethe-Universität Frankfurt

<sup>2</sup>Freie Universität Berlin

November 2021

Raw data and scripts used for this project is available on GitHub. <https://github.com/FPupillo/PreM-Computational>

## Abstract

Predictive processing accounts propose that our brain constantly tries to match top-down internal representations with bottom-up incoming information from the environment. Predictions of varying degrees can lead to corresponding prediction errors depending on whether or not the information encountered in the environment conforms with prior expectations. Theoretical and computational models assume that prediction error has beneficial effects on learning and memory. However, while there is strong evidence on the effects of prediction error on learning, relatively less evidence is available regarding its effects on memory. Previous studies on the effects of prediction error on memory using computational models have manipulated the amount of reward participants received on probabilistic learning task and showed contrasting evidence. We used a task in which the reward is not explicitly manipulated but learned from the correct or incorrect outcome of the predictions. Participants first formed expectations of different strengths for context/object-category associations in a contingency-learning paradigm. In the encoding phase that followed, they were asked to predict the expected category of the objects after being cued by specific contexts. The objects that were then shown could either match or violate their previously learned expectations. Finally, participants were asked to complete a surprise recognition memory test. We used a reinforcement learning model to derive subject-specific trial-to-trial estimates of prediction error. In two different experiments, results showed that prediction error at encoding influenced subsequent memory as a function of the outcome of participants' predictions (correct vs incorrect). Precisely, when participants correctly predicted the object category, stronger prediction error (as an outcome of weak prior expectations) led to enhanced memory. In contrast, when participants incorrectly predicted the object category, stronger prediction error (as an outcome of strong prior expectations)

led to impaired memory. These results reveal a computationally specific influence of prediction error on memory formation, highlighting the important moderating role of prediction accuracy.

In our daily interaction with the environment we are confronted with a massive amount of information which cannot all be processed in detail, given the limited resources of our cognitive systems. In order to simplify the complexity of incoming information, our brain tries to extract its regularities in order to be able to react to environmental demands in efficient ways. One way of extracting regularities from experiences is to rely on repetitions and associations of events, which lead to the creation of increasingly complex knowledge (Ghosh and Gilboa, 2014; Tse et al., 2007). The accumulation of knowledge across similar experiences that can subsequently form expectations and orient our actions is what characterizes reinforcement learning (Sutton and Barto, 1998). In reinforcement learning, individuals incrementally learn, over experiences, to form expectations in order to successfully predict future events. As an example from real life, we may learn after several experiences that book stores tend to specialize on different genres, such as crime or romance. Learning this association will allow us to anticipate the type of books in a given book store and thus guide us to enter it or not if we are looking for a particular book.

Events can, however, sometimes deviate from our expectations. We may go to a book store we have learnt being specialized in crime genre and discover that the crime novel we are looking for is not there. We might then decide to go to another book store, which we know has equal coverage of different genres. Despite our low expectations, we may find the crime novel we are looking for in that generalist book store. In such situations, the difference between expectations and the experienced event generates a prediction error (PE) signal which may change future expectations and decisions. PE is crucial in promoting learning by driving the updating of prior knowledge (Ergo et al., 2020; Friston, 2018). An aspect that may modulate the effect of PE is the outcome of the predictions. In fact, correctly predicting an event generally results in an increase in the strength of the expectations that led to the prediction, whereas incorrectly predicting leads to a weakening of the related expectations (Daw and Tobler, 2014). For example, correctly predicting that the book we are looking for is in a specific book store will increase our belief that we need to go to that store to find similar books. Conversely, when we do not find the book in the store our belief will not be as strong as before. Going beyond incremental learning, our current study examined the effects of PE on the formation of episodic memory, with learning rate and prediction outcome as potential modulators.

In reinforcement learning models, the extent to which PE is used to update the expectations is regulated by the learning rate (Sutton and Barto, 2018). A higher learning

rate results in an increased update of the expectations, shifting the expectations towards the most recent outcomes. On the contrary, lower learning rate shifts the expectations towards older estimates. The amount of update that follows a PE can also be modulated by the outcome of the predictions, as individuals tend to update their belief more as a consequence of positive outcomes compared to negative ones (Lefebvre et al., 2017; Sharot et al., 2007, 2011; Sharot and Garrett, 2016).

In addition to incrementally learning from multiple episodes, individuals are also capable of encoding distinct, temporally specific episodic memories of events. For example, remembering the precise occasion in which the desired book was found in a store. While there is a great amount of evidence on the effects of PE, learning rate, and prediction outcome on incremental learning, its effects on the formation of new episodic memories are still not fully understood.

The study of the relationship between prior expectations and learning has benefited from the use of computational models. Reinforcement learning models in particular have been used, due to their ability to confer a precise mechanistic role to PE in learning and map it to its neural substrates. In fact, it has been shown that firing patterns of mesencephalic dopamine neurons and also Blood-Oxygen-Level Dependent (BOLD) signal change in the striatum resemble the PE signal used in reinforcement learning models (Daw, 2011; McClure et al., 2003; Schultz et al., 1997). This dopaminergic-dependent PE is thought to inform future predictions by indicating deviations between observed and predicted outcomes (Daw and Tobler, 2014; Niv and Schoenbaum, 2008; Rangel et al., 2008; Rushworth and Behrens, 2008), thus encouraging the repetition of actions that are better than expected (positive PE) and discouraging the repetition of actions that are worse than expected (negative PE, Schultz, 2016; Steinberg et al., 2013). An emerging line of research examines the relation between prediction errors and long term memory formation as a potential mechanism that explains interactions between memory and learning. These studies are motivated by both anatomical considerations that the hippocampus receives dopaminergic input (e.g., Lisman and Grace, 2005) and functional findings on interactions between hippocampus and striatum (e.g., Poldrack et al., 2001).

Several studies have manipulated the amount of the reward participants expected and received, linking the obtained reward PE experienced at the time of item presentation or immediately after presentation to the subsequent episodic recognition of those items (Rouhani et al., 2018; Rouhani and Niv, 2021; Jang et al., 2019; De Loof et al., 2018). Some studies have found improved memory for surprising outcomes, namely better memory for items associated to both better- and worse-than-expected outcomes (Rouhani et al., 2018; Rouhani and Niv, 2021). By contrast, some other studies found that better-than-expected

outcomes, compared to worse-than-expected ones, led to improved later recognition (Jang et al., 2019; De Loof et al., 2018; Rouhani and Niv, 2021, Experiment 2 ). Therefore, mixed evidence has been gathered concerning the effects of reward PE on episodic memory, depending on the sign of PE (or put differently, accuracy of prediction).

The aforementioned studies using computational models manipulated expectations by using rewards. Since in everyday life learning does not occur always in the presence of explicit rewards, it is crucial to consider the mechanistic effects of PE *per se*, in contexts in which no explicit information about the reward is conveyed. A different way of looking at the relationship between expectations and memory may involve generating them through associations between a context (e.g., going to a book store to look for a book) triggering some expectations (e.g., more or less strong belief about the presence of the book), and a matched or unmatched outcome (e.g., finding or not finding the book). In addition, the studies cited above have looked at the effects of PE in conditions in which participants were actively learning novel associations. However, familiar situations in which the environmental structure has been somewhat internalized are more common in real life.

Therefore, the present study pre-trained participants to form expectations of varying strengths, which were subsequently matched or mismatched to render PE of different strengths and relate it to episodic memory performance. Specifically, we designed a task in which participants learned probabilistic associations between contexts and object categories. After a learning phase in which the expectations were established, participants were presented with the contexts and had to predict the category of trial-unique objects that appeared at the end of each trial. In order to generate different levels of PE strength, we quantified expectations by using gradually different contingencies. The probability of a certain category following a context was systematically manipulated so that for some contexts expectations were stronger than for others. Findings reported in a separate publication using the same data as the current study (Ortiz-Tudela et al., 2021) showed that such context manipulation affected recognition memory. Specifically, memory was found to be better for contexts characterized by weaker expectations, compared to contexts in which expectations were higher. In the present study, we fitted computational models to the data from the learning and encoding phases to perform novel analyses. More specifically, we used a reinforcement learning model to derive learning rates and trial-level PE experienced during the presentation of unique object images (i.e., encoding phase) and related them to the likelihood of subsequently recognizing the items in a following surprise recognition memory test. The computationally derived PE reflected in a quantitative and gradual manner how unexpected the presentation of an object category in a given context was to the participants. The procedure thus allowed to test whether episodic memory for the objects was related to learning

rate and PE experienced in situations where no explicit reward was involved.

We reasoned that since increased learning rate results in higher weights on more recent events (Daw and Tobler, 2014), it should also bias the system towards increased encoding, resulting in better episodic memory. In relation to PE, we hypothesised that if unpredicted events *per se* improve memory encoding, we would observe a positive relationship between PE and memory encoding, so that memory performance would be better for the more unpredicted events. Contrarily, if the effect of PE on memory is modulated based on whether or not the prediction was correct, we would observe an interaction between PE and choice outcome so that PE improves memory for better than expected outcomes, while impairs memory for worse than expected outcomes. In two experiments, we showed that the likelihood of correctly recognizing an item scaled with PE and was modulated by prediction outcome at encoding. Specifically, when participants correctly predicted the object category stronger PE led to improved memory encoding, whereas for incorrect predictions stronger PE led to impaired encoding.

























## 1 Results

In both experiments, participants performed a task in which they were asked to predict the category of trial-unique objects that would follow a specific context. Instructions were given to indicate that each context was predictive of one object category, but there was no indication about which category and with which contingency (see Figure 1). Participants learned the associations between contexts and object categories during a learning phase (phase 1), in which they received feedback on every trial. In a subsequent encoding phase (phase 2), a new set of never-seen-before objects (but belonging to the same object categories as the ones in phase 1) was introduced and participants were asked to continue doing the same task as in the previous phase. Importantly, participants no longer received explicit feedback on their performance. Finally, they completed a recognition memory test, where they were asked to recognize the objects presented in phase 2 among several distractors.



















In Experiment 1, participants were presented with six contexts. Each of the contexts was predictive of the object categories following specific contingencies (Figure 1): in half of the contexts, one of the three object categories was presented 80 % of the times, and the remaining two object-categories 10 % of the times; in the other half of the contexts, all the three object categories were equally likely. In Experiment 2, two instead of three object categories were used. In addition, the object-context contingencies were 90-10 %, 50-50 %, and 70-30 %, respectively. This manipulation allowed us to sample more points along the PE continuum. To reach the desired contingencies for each condition, filler objects from the

same object categories were introduced and repeated several times, increasing the number of the trials especially in the strong prior conditions. Recognition memory for these objects was not tested.

a)

Scene Categories	Object Categories		
	Instruments	Household objects	Fruits/Vegetables
	 .80	 .10	
	 .10	 .80	
	 .10	 .10	
	 .33	 .33	
	 .33	 .33	
	 .33	 .33	

b)

Scene Categories	Object Categories	
	Instruments	Household objects
	 .80	 .10
	 .10	 .80
	 .10	 .10
	 .33	 .33
	 .33	 .33
	 .33	 .33

c)

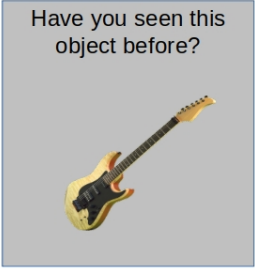
Phase1. Contingency learning



Phase2. Encoding



Phase3. Memory test



**Figure 1: Illustration of the Methods.** Illustration of the scene/object-category contingencies for a) Experiment 1 and b) Experiment 2. Note that each object is representative of its object category and that different, unique items from each category were used on every trial. c), illustration of the three phases of the study.

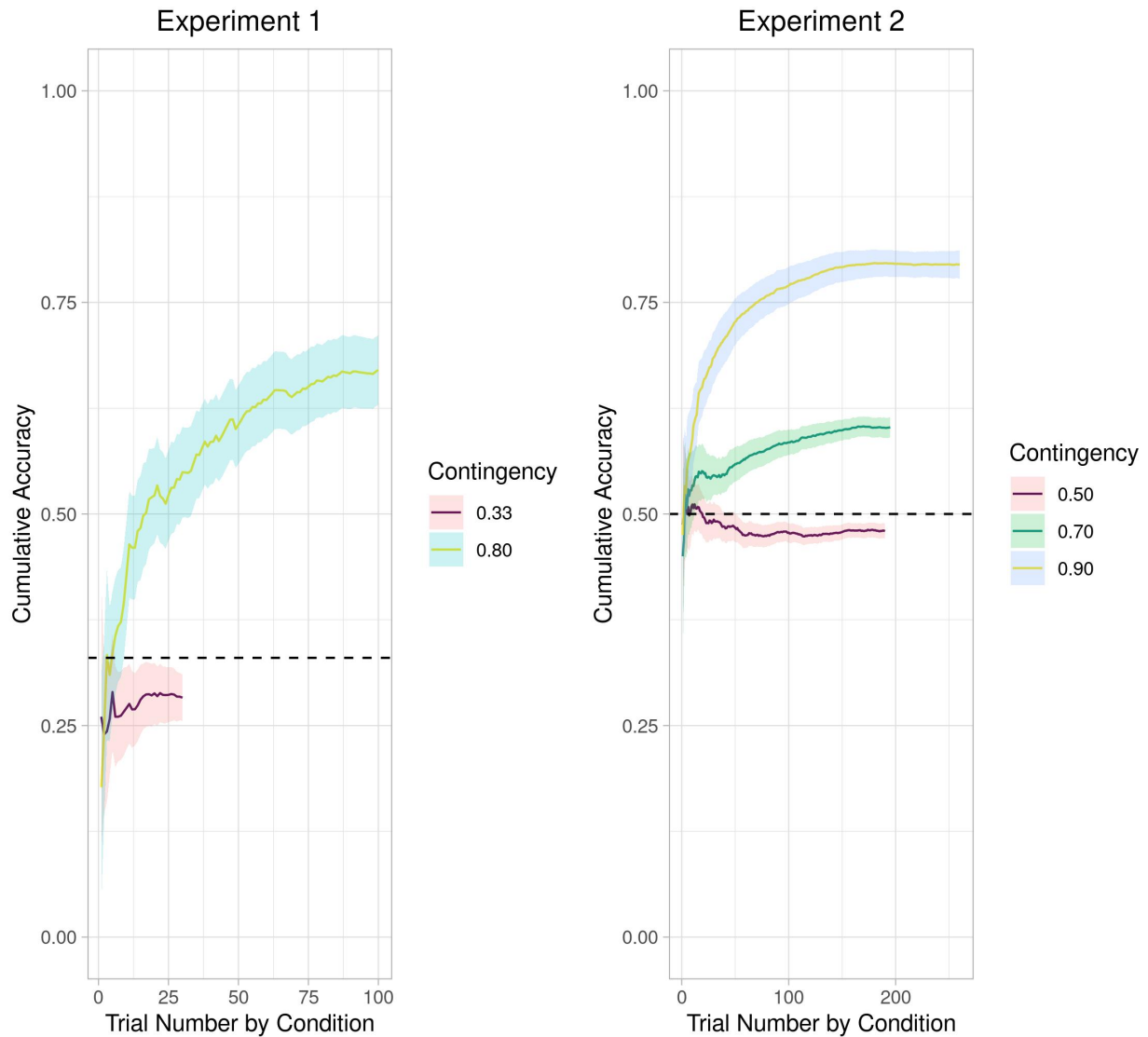
### 1.1 Learning Performance

In both Experiment 1 and Experiment 2, learning performance during the contingency learning and encoding phases showed that participants understood the task correctly and were able to learn to predict the object category that was more likely to be presented for each context. Participants' cumulative accuracy tended to approximate the true probability for each context and experiment, as shown in Figure 2.

### 1.2 Computational Models

In order to derive learning rate and trial-level PE at contingency learning and encoding phases, variants of a reinforcement learning model (Sutton and Barto, 2018) were fitted to data from both learning and encoding phases pooled together. The use of reinforcement





**Figure 2: Participants’ learning performance.** Participants’ learning performance at the learning and encoding phase for Experiment 1 and Experiment 2. Trial number by contingency condition is represented in the x axes, while cumulative accuracy is shown on the y axes. The different colours represent the contingency conditions. The dashed lines indicate chance level while the shadows represent standard error. Note that the number of trials differs across condition, as filler objects were used for stronger contingency conditions, compared to weaker ones to reach the desired contingency.

learning models allowed to capture the process of establishing prior expectations while learning the object-category contingencies for the different contexts. In reinforcement learning models, an agent is assumed to learn values of context-category associations by adding the current expected value to a learning rate  $\alpha$  multiplied by the PE. The learning rate  $\alpha$  is a value between 0 and 1 that determines the influence of the current prediction error on the expected values. It represents the extent to which evidence from the current trial is used to update the expectations: Higher learning rate weights more the present evidence and the extent to which it deviates from the value estimates, while lower learning rate weights more the estimated values, and thus the past trials. We fitted four different reinforcement learning models that made different assumptions on how participants learned the context/object-category associations (see Methods): (1) an instructive model with a learning rate  $\alpha$  (fixed across participants) that decreases across trials (dLRI), and updates the expected values by increasing the value of the object category presented on a given trial and decreasing the



values of the categories not presented, regardless of the choices made by the participants; (2) an instructive model with a decreasing learning rate  $\alpha$  that was free to vary between individuals (dflRI); (3) an instructive model with a free constant learning rate  $\alpha$  (flRI); (4) an evaluative model with a free constant learning rate (flRE), where the expected values were updated depending on the accuracy of participants' actions, increasing for correct predictions and decreasing for incorrect ones.

The dLRI considers how the expected values should be updated optimally, since it is derived from a Bayesian formulation of the task (see Methods and Supplemental Material). In this optimal Bayesian formulation of learning, prediction error is assumed to have its maximal influence on learning in the early trials, and decreases as a function of the inverse of the number of the trials. In this model, the only parameter that was estimated was the 'inverse temperature'  $\beta$ , which regulates the stochasticity/determinism trade-off in selecting the action depending on the expected values: Higher values of  $\beta$  represent more probable preference of the higher context/object-category associations, while lower values also consider low-strength associations, producing more noisy choices.

In addition to the  $\beta$  parameter, the instructive model with the free decreasing learning rate estimates a learning rate  $\alpha$  that was free to vary across individuals and decreased inversely proportional to the number of the trials. The instructive flRI and evaluative flRE models estimated a learning rate  $\alpha$  that was constant throughout the learning and encoding phases, and free to vary across individuals in order to capture participants' distinct learning rates. These models make different assumptions on how participants used information to update the expected values. In fact, while evaluative free-learning rate model (flRE) assumes that participants use the feedback received (correct vs incorrect) to update only the category chosen, the instructive free-learning rate model (flRI) implies that on each trial participants update all the associations by strengthening the one between the context and the category presented, while lowering the associations with that context and the categories that were not presented at that trial.

Prior to fitting the models to participants' data, we ensured that the models could distinguish among different parameter values and also generate qualitatively different data (see 'Parameter Recovery' and 'Model Recovery' in Methods and Supplemental Material). Then, the three models were fit to participant's data, and the parameters of best fit were estimated as the parameters that maximized the likelihood of participants' choices.

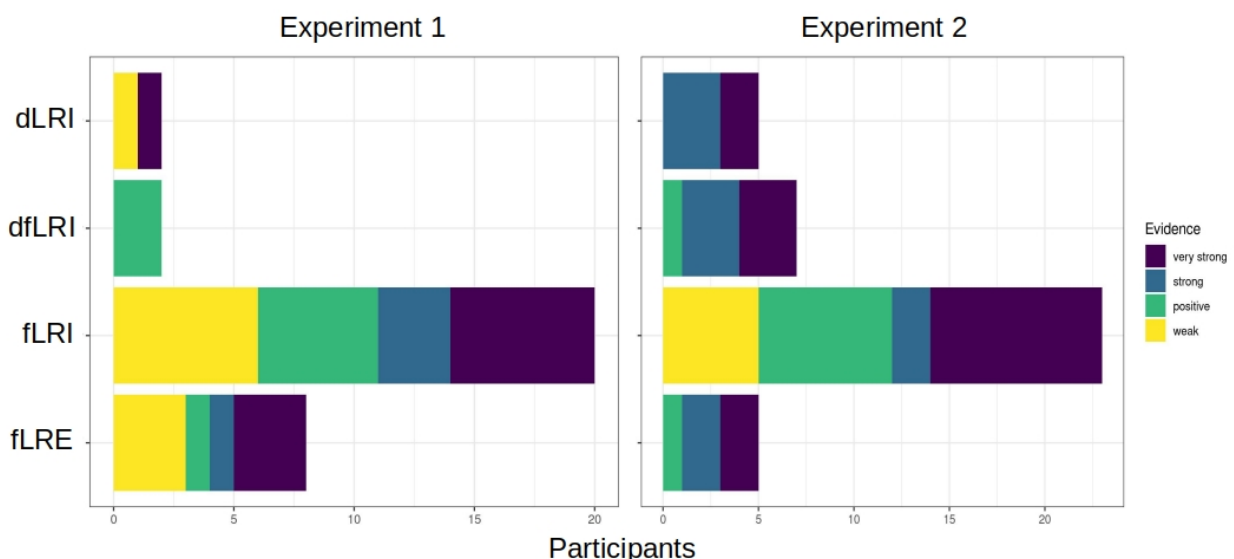
### 1.2.1 Model Comparison

In addition to calculating log-likelihood, we calculated the Bayesian information criterion (BIC) for each model and for each subject, by multiplying the maximum likelihood (i.e. the

**Table 1:** Model Comparison. BIC values and standard errors for each model for Experiment 1 and experiment 2. *Best(N)* and *Very strong(N)* refer to the number of participants for which the model was the best fit and for which there was very strong evidence, respectively.

Model/Experiment	BIC ( <i>se</i> )	Best(N)	Very Strong (N)
<b>Experiment 1</b>			
dLRI	289.3(2.5)	2	1
dfLRI	277.0(2.0)	2	0
fLRI	266.4(1.7)	20	6
fLRE	271.4(2.7)	8	3
<b>Experiment 2</b>			
dLRI	801.2(4.3)	5	2
dfLRI	793.1(2.9)	7	3
fLRI	783.1(3.7)	23	9
fLRE	774.3(2.9)	5	2

likelihood for the parameters of best fit) by the number of free parameters in the model. This approach penalizes models with more parameters. We then marked the number of participants for which each model was the best fit, as well as the evidence for it, computed as the BIC difference between the best and the second best model. Results are shown in Figure 3. Table 1 show BIC values and the the number of participants for which a model was the best fit, as well as the number of participants for which there was strong evidence, for both Experiment 1 and Experiment 2.



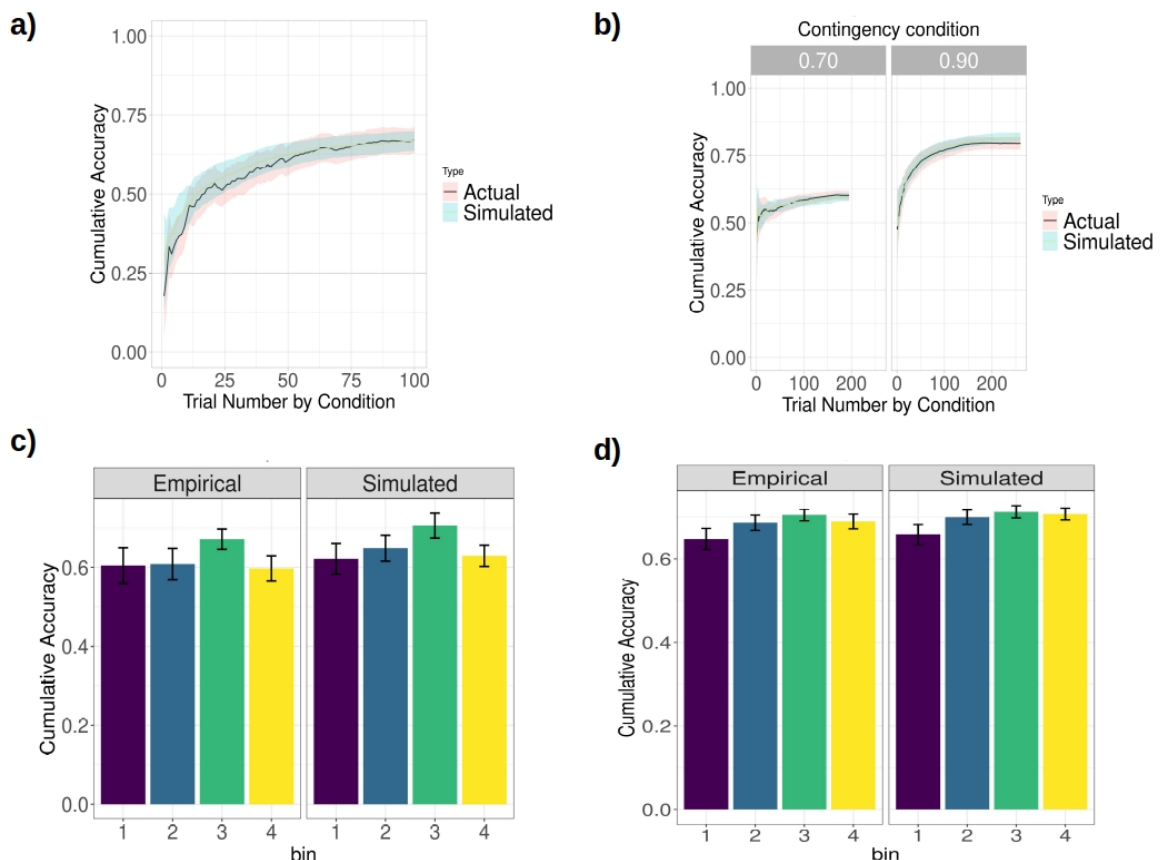
**Figure 3: Model Comparison.** Results of model comparison for the two experiments. Evidence for the best model for each participants is shown.

Model comparison established that the instructive model with the free learning rate (fLRI) explained participants behaviour better than the other models. The overall BIC of fLRI was the smallest (indicating better fit), and the number of participants for which it was the best model were 20 over a total of 32 participants for Experiment 1, and 23 over a total of 40 participants for Experiment 2. In addition, there was very strong evidence for it being the best model for 6 participants in Experiment 1 and 9 participants in Experiment

2. These results indicated that participants learning processes deviate from the behaviour of an optimal Bayesian observer and that are best described by using individual learning rates that are constant across trials. In addition, model comparison shows that most participants used the category information of the object presented at the end of each trial to update all the context/object-category associations, and not only the associations related to the chosen object category.

### 1.2.2 Model Validation

We then validated the winning model by looking at the ability of the model to generate performance which was qualitatively similar to participants' actual behaviour. For each actual participant, we simulated data on the learning and encoding tasks using the best fitting model (fLRI) and its estimation of best fitting parameters. The model simulations included the actual task structure (number of trials and contingencies). In order to evaluate model's simulations, we calculated cumulative accuracy for the data generated by the model and compared it to participants' actual cumulative accuracy. Figures 4a and 4b show that the simulated models capture participants' behaviour. Data simulated from the dLRI and the fLRI models can be found in the Supplemental Material (Figure S4 and Figure S5)



**Figure 4: Simulated vs Empirical Data.** Simulated data (red line) and actual data (green line) overlapped, for a) Experiment 1 and b) Experiment 2. Cumulative accuracy on the learning task for simulated and actual data at different learning rate levels for c) Experiment 1 and d) Experiment 2. Cumulative accuracy was binned.

In addition, to check whether the model captured participants' differences in learning rate, we compared cumulative accuracy for simulated and empirical data at different

values of learning rate  $\alpha$ . For simulated and empirical data, quartiles for  $\alpha$  were calculated (zeroth, first, second, third, and fourth quartile), and cumulative accuracy was aggregated for the data points between one quartile and the previous one. Cumulative accuracy for the four bins created is shown in Figures 4c and Figures 4d, as a function of order of type of data (empirical vs simulated), and experiment (first vs second experiment). In both experiments, the simulated data mirrored the pattern of the actual data, showing that the model was capable of capturing observed effects of individual differences in learning rate on cumulative accuracy.

### 1.3 Recognition Memory Results

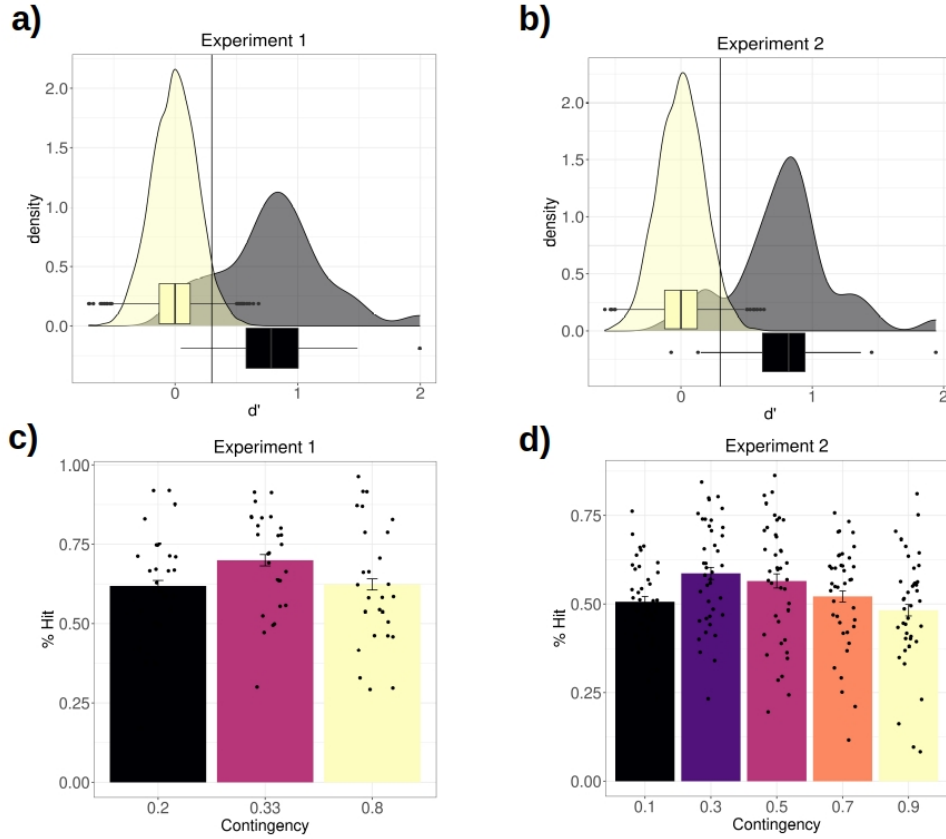
To evaluate overall memory performance for each participant, a  $d'$  score was calculated from the hits (responding "old" to old items) and false alarms (responding "old" to new items), an index which indicates participants' ability to discriminate between old and new items. In order to exclude participants who did not perform the task above chance level, we created a null distribution by generating 5000 random permutations of the trial labels. We then excluded participants whose performance was below the 95 % percentile of the null distribution (see Figures 5a and 5b). Five participants from experiment 1 and five from experiment 2 with overall  $d'$  score below the obtained threshold were excluded from further analyses. After the exclusion, the final  $d'$  was  $d' = 0.93$ ,  $t(26) = 13.7$ ,  $p < .001$  for experiment 1, and  $d' = 0.90$ ,  $t(34) = 16.8$ ,  $p < .001$  for experiment 2, indicating that participants were overall able to discriminate previously presented old items from new distractors.

Figures 5c and 5d show recognition accuracy as a function of contingency condition. Participants' recognition accuracy was higher for weaker contingency conditions in both Experiment 1 and 2.

### 1.4 Memory as a function of model-derived PE

The fLRI model was fit to participant data with the estimated best fit parameters in order to derive trial-level PE during the encoding phase. Figure shows how model-derived PE was computed and then linked to recognition accuracy. Participants' expected values for each object category were computed on each trial, and used to derive trial-level PE. Higher PE levels were generated when a category presented was not expected, as reflected by its corresponding expected value being smaller. By contrast, lower PE was generated by trials in which the category presented was characterized by a high expected value. The PE derived at encoding was then linked to the likelihood of successfully recognize an image as "old" in the subsequent recognition memory test by using a logistic regression model.

As the two experiments were conceptually identical, the analyses were run on data

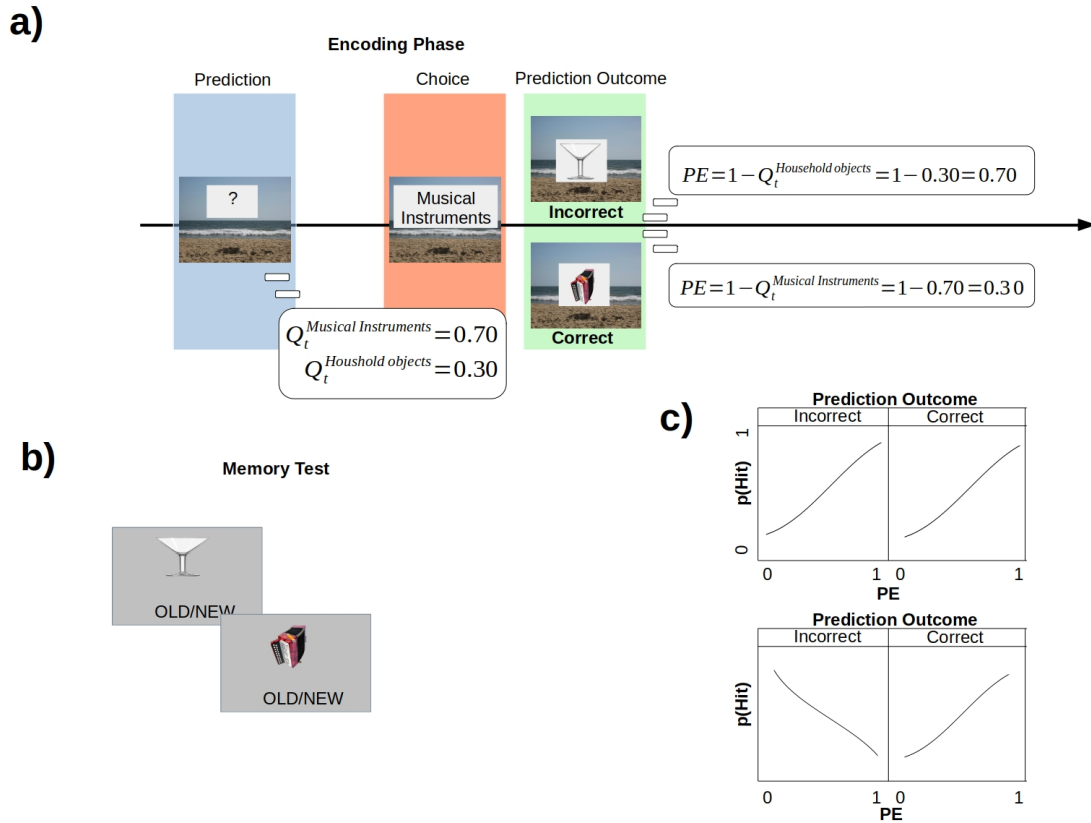


**Figure 5: Recognition accuracy.** a) distributions of  $d'$  by experiment. Participants'  $d'$  distributions (on the right side of both graphs) overlaid with the null distribution created by generating 5000 random permutations (on the left side), for experiment 1 and 2. The vertical black line represents the 95 % percentile of the null distribution. Participants' whose  $d'$  fell below that threshold were excluded from further analyses. b) recognition accuracy as a function of contingency condition, for Experiment 1 and 2.

collapsed across them. Distribution of PE by contingency collapsed across experiments is shown in Figure 7a.

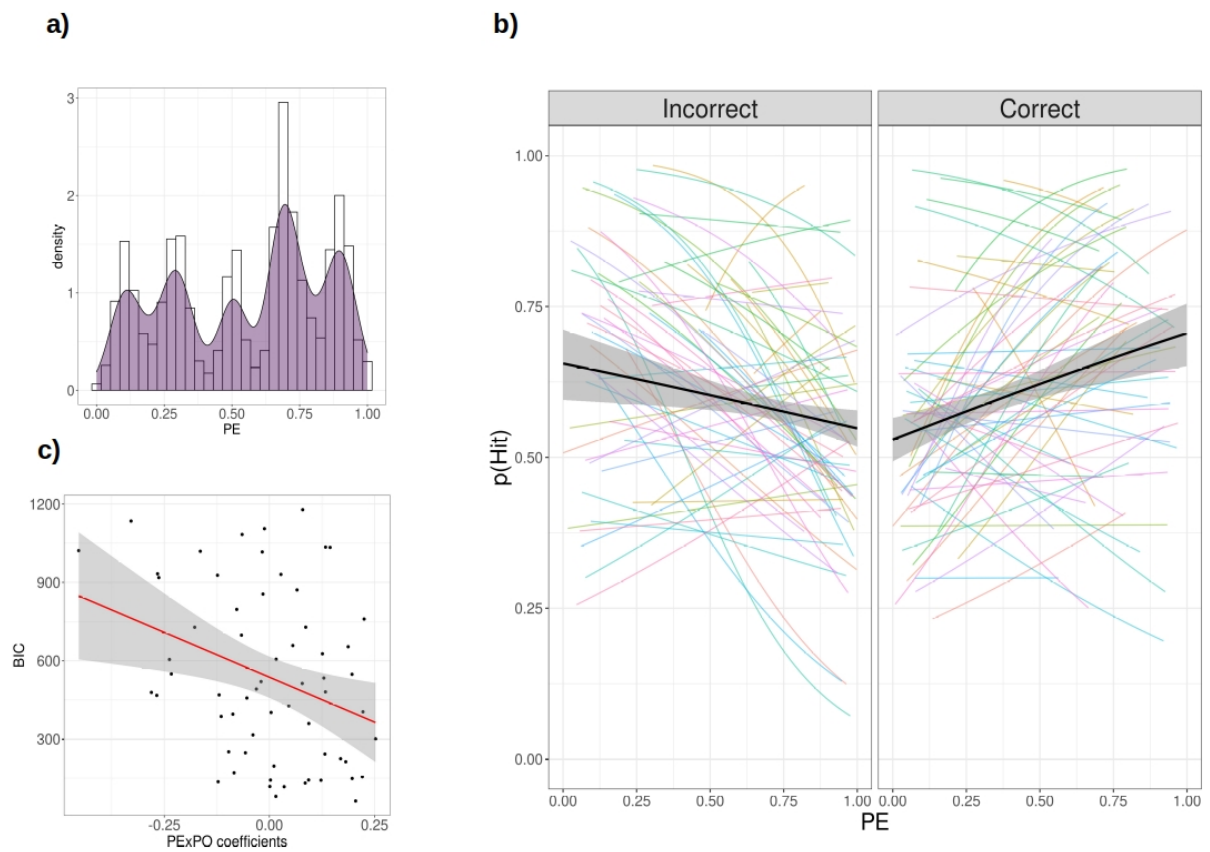
We then tested whether model-derived PE was related to recognition memory. Plots of the observed values (see Figure 7b) revealed a relationship between PE and memory modulated by prediction outcome. In order to statistically test for the significance of this relationship, we used a generalized linear-mixed model where participants were treated as random effects (see Methods). In the model, PE, learning rate, and prediction outcome, as well as their interactions, were added as fixed effects. In addition, random slopes for PE and prediction outcome, and their interaction, were also added to the model. Finally, experiment (1 or 2) was added as a fixed term (along with interactions) to model any differences between the two experiments. Results of the analysis are presented in Table 2. In terms of comparability between experiments, there was a main effect of experiment,  $\chi^2_{(1)} = 9.91$ , with participants performing worse at the recognition test in Experiment 2, compared to Experiment 1,  $\beta = -0.63$ ,  $p = .002$ , OR = 0.53. Importantly, all the interactions including the experiment number were also non significant, all  $ps > .204$ , showing that the effects of interest did not differ across the two experiments.

Next, we turned to the effects of key interest involving learning rate, PE, and prediction outcome, as well as their interactions. First, results showed that the main effects of



**Figure 6: Model-derived PE .** a) the model computed different expected values  $Q$  at each trial. Participants' category choices could be correct or incorrect. The PE depended on the expected value of the category presented at the end of each trial. b) items presented during the encoding phase were showed again to participants. Participants were asked to recognized those objects, among distractors. The probability of recognizing old object as "old" (p(hit)) was related to the PE error computed during the encoding phase in a logistic regression model. c) two hypothetical relationship between PE and recognition memory. On the top, linear relationship between PE and recognition memory, regardless of prediction outcome. On the bottom, prediction outcome modulates the PE-memory interplay.

learning rate, as well as all the interactions including learning rate, were not significant, all  $ps > .171$ . These results suggest that the estimated learning rates did not affect overall memory accuracy and that also it did not affect the effects of PE, prediction outcome, and memory. Second, the main effects of PE and prediction outcome did not reach significance, all  $ps > 0.316$ , suggesting that the overall effect of PE and prediction outcome was not significant. Importantly, there was a significant interaction between PE and prediction outcome,  $\chi^2_{(1)} = 26.14$ ,  $p < .001$ . In order to break down this interaction, we analysed the effect of PE on recognition memory separately for correct and incorrect predictions.



**Figure 7: Computationally-derived PE and Memory.** a) density plot of model-derived PE for merged data from Experiment 1 and Experiment 2. b) spaghetti plot of the observed relationship between PE and recognition memory as a function of prediction outcome. Each coloured line represent one participant, while the logistic regression line across participants is depicted in black. c) Scatterplot showing the relationship between participants' PE-by-prediction outcome interaction effects and participants' BIC. Lower BIC represents better model fit.



**Table 2:** Results of the Main Analysis.

<i>Fixed Effects</i>	$\beta$ ( <i>se</i> )	95% CI		<i>z</i>	<i>p</i>	OR
Intercept	0.66 (0.10)	0.46	0.85	26.54	<.001	1.91
PE	-0.12(0.19)	-0.64	0.22	-0.12	.521	0.88
Prediction Outcome <sup>1</sup>	0.10(0.10)	-0.08	0.29	1.00	.316	1.11
Learning Rate	0.68(2.30)	-3.74	5.19	0.30	.767	1.98
Experiment <sup>2</sup>	- 0.63(0.20)	-1.00	-0.22	-3.15	.002	0.53
PE x Prediction Outcome	1.87(0.36)	1.18	2.53	5.11	<.001	6.47
PE x Learning Rate	1.96(3.67)	- 3.26	6.83	0.53	.593	7.11
Prediction Outcome x Learning Rate	2.91(2.13)	-0.50	6.34	1.37	.171	18.40
PE x Experiment	-0.09(0.38)	-0.85	0.63	-0.25	.801	0.91
Prediction Outcome x Experiment	0.19(0.20)	-0.20	0.60	0.94	.348	1.21
Learning Rate x Experiment	4.0(4.58)	-4.83	13.15	0.87	.381	55.03
PE x Prediction Outcome x Learning Rate	0.90(5.84)	-5.75	6.80	0.15	.877	2.47
PE x Prediction Outcome x Experiment	-0.93(0.74)	-2.28	0.47	-1.27	.204	0.393
PE x Learning Rate x Experiment	-1.95(7.32)	-10.62	7.03	-0.27	.790	0.141
Prediction Outcome x Learning Rate x Experiment	3.53(4.23)	-3.00	10.39	0.83	.404	34.23
PE x Prediction Outcome x Learning Rate x Experiment	5.17(10.96)	-2.09	12.93	0.47	.637	175.88
<i>Random Effects</i>	Variance	St. Dev.				
Subjects (Intercept)	0.30	0.55				
Prediction Outcome	0.1	0.03				
PE	0.03	0.17				
Prediction Outcome X PE	0.03	0.18				

Note. <sup>1</sup> Prediction Outcome contrasts have been set to 0.5 and -0.5 for correct and incorrect prediction, respectively.

<sup>2</sup>Experiment contrasts were set to 0.5 and -0.5 for Experiment 1 and Experiment 2, respectively.

Results showed a significant positive linear relationship between PE and recognition memory for correct prediction outcome,  $\beta = 0.80$ ,  $p < .001$ ,  $OR = 2.22$ , and a significant negative linear relationship between PE and prediction outcome for incorrect prediction outcome,  $\beta = -0.78$ ,  $p < .001$ ,  $OR = 0.46$ . These results suggest that the effect of PE on memory encoding is different depending on prediction outcome.

We ran additional analyses with hit rate binned by aggregating it between the quartiles for PE for each participant. Results from these analyses did not change the overall pattern presented in the previous analyses (see Supplementary Material and Figure S7). We then explored whether participants’ learning behaviour in the encoding phase was linked to the strength of the effect of PE on subsequent recognition. We thus correlated participants’ PE-by-prediction outcome interaction coefficients to participants’ BIC. The BIC represents how well the model (instructive model with free learning rate) reflected participants’ learning behaviour, with lower values indicating better fit. A scatterplot of the relationship is shown in Figure 7c. There was a significant negative correlation between the PE-by-prediction outcome interaction coefficients and the BIC,  $r(60) = -.33$ ,  $p = .009$ , indicating that the closer participants’ learning behaviour was to the reinforcement learning model, the stronger the PE-by-prediction outcome effect on memory encoding.

## 2 Discussion

Previous literature has provided mixed evidence on the effects of PE on memory formation, with studies on reward PE using reinforcement learning models producing contrasting results (Jang et al., 2019; De Loof et al., 2018; Rouhani et al., 2018; Rouhani and Niv, 2021). We explored the effects of PE on memory for the first time in a paradigm which did not include an explicit manipulation of the reward and conditions in which participants had already established prior expectations. In the task used, associations between context and object categories were first learned by participants and then used to predict the category of trial-unique upcoming objects. In two experiments, we used a reinforcement learning model to derive trial-by-trial PE generated by expectations of different strength and analyzed its effect on memory encoding. We showed that the outcome of participants’ predictions was a modulator of the effects of PE on memory encoding. Precisely, when a prediction turns out to be correct, higher PE was related to better memory; conversely, when a prediction turns out to be incorrect, lower PE was related to better memory. These results reveal a computationally specific effect of PE, highlighting the crucial modulating role of prediction outcome.

Our findings are in line with studies on reward PE showing better memory for

better than expected outcomes (De Loof et al., 2018; Jang et al., 2019), a pattern suggested to be related to dopaminergic activity promoting hippocampal plasticity and memory formation (Bethus et al., 2010; Rosen et al., 2015). It is well known in computational neuroscience that the neurotransmitter dopamine is responsible for a PE signal that drives plasticity in the striatum, facilitating repetitions of actions with better-than-expected outcomes (Daw and Tobler, 2014; Niv and Schoenbaum, 2008). Dopamine is also known to enhance long term potentiation in the hippocampus (Lemon and Manahan-Vaughan, 2006), and this modulatory effect might be responsible for the prioritization of relevant information in memory.

It has also been shown that the effect of dopamine on the hippocampus can be bidirectional: Higher levels of dopamine cause phasic firing in the hippocampus which results in increased activation, while lower levels of dopamine produce tonic firing and inhibit hippocampal activation (Rosen et al., 2015). In the present study, this dopaminergic effect may have been driven by the difference between the expectations and the outcome of the prediction. More specifically, a correct prediction might nevertheless provide a PE signal under conditions of weak expectations, which increases the release of dopamine in the striatum, promoting hippocampal activation, and resulting in better encoding. This idea is supported by evidence showing that increased striatum-hippocampus connectivity during learning in some conditions supported enhanced memory encoding (Davidow et al., 2016). Conversely, an incorrect prediction in conditions of strong expectations might correspond to a negative PE, which suppress the release of dopamine in the striatum and activation in the hippocampus, resulting in impaired encoding. More studies investigating the connectivity between hippocampus and striatum at these different conditions are needed to provide support for these hypotheses.

In the present study, PE was experienced at the time of the presentation of the to-be-remembered items. The object presented provided feedback to participants on whether or not their predictions were correct. Previous studies from De Loof et al. (2018) and Jang et al. (2019) found effects of reward PE experienced during item presentation on memory encoding that are consistent with our results. Importantly, Jang and colleagues showed effects of reward PE to be elicited at item presentation, but not at the time of the presentation of the feedback, when the objects were no longer shown. Therefore, our results provide additional support to the view that PE has to be elicited during object presentation in order to have solid effects on memory encoding.

Results from the current study are in contrast with previous findings showing a positive relationship between unsigned reward PE and memory (Rouhani et al., 2018; Rouhani and Niv, 2021). In the present study, PE represented how unexpected the presentation of a category was, and thus was equivalent to unsigned PE. In contrast to the findings from

Rouhani and colleagues, our results showed that the overall effect of PE, independent of the outcome of the prediction, was not significant. One possible explanation for this discrepancy is the task that these studies used at encoding. Rouhani et al. (2018; 2021) presented participants with scenes that could be predictive of the future rewards. After participants made their predictions, the images were presented together with the reward received, which could be either better or worse than expected, thus generating a reward PE. In a first study (Rouhani et al., 2018), they showed a positive effect of unsigned prediction error on memory encoding, which thus improved for both better- and worse-than-expected outcomes. However, in this study it was not clear whether the effect was due to PE occurring before or after feedback presentation, as the images were presented even before the presentation of the feedback. In a second study (Rouhani and Niv, 2021), the authors manipulated reward PE before and during feedback delivery separately, finding an effect of signed reward PE for images presented before feedback delivery and an effect of unsigned reward PE for items presented during feedback presentation. The discrepancy between these findings and our findings could be due to the different methodologies used to elicit PE. In fact, the reward PE experienced during feedback delivery in the study by Rouhani and colleagues (Rouhani and Niv, 2021) was driven by a specific condition in which participants could win or lose money. As a consequence, the effects observed might have been triggered by arousal-related emotional processes linked to the activity of the amygdala (Watanabe et al., 2019), which are known to enhance memory (Mather and Sutherland, 2011).

Furthermore, the utility of the information presented may have played a role in the results of the current study. Events can be processed differently depending on whether they are useful to predict events in similar contexts. In fact, while positive outcome when prior expectations are weak may inform individuals that the choice is important for future similar contexts, negative outcome in contexts where prior expectations are strong may be taken as a random event not being particularly useful for future predictions. Therefore, events that are important to guide future behaviour may be remembered better, while those that are deviant from established expectations may be discarded as exceptions, unless if they happen repeatedly. It has been shown that dopaminergic neurons react similarly to both rewards and to cues that indicate potentially important information (Bromberg-Martin and Hikosaka, 2011). This effect may reflect in a strengthened representation of items that carry more information for predicting future events, in line with views considering reinforcement learning not only as a reward-seeking system, but also as an information-seeking system (Bromberg-Martin and Hikosaka, 2011; Niv and Chan, 2011). More studies are needed to disentangle the effects of reward-seeking and information-seeking processes of PE on memory.

It is important to note that mismatched information might have been discarded as

not helpful for the future because the task used in both experiments included contingencies that were established before the encoding of the events and never changed during the course of the tasks. Participants underwent extensive learning phases in which they established their expectations for the different contexts prior to the encoding task. This settings is not frequent in previous computational modeling studies, although it is more common in real life, where in most situations individuals are likely to find themselves in situations they are typically familiar with. In conditions in which expectations are established and known to be stable, deviant information may be taken as rare event of chance. On the contrary, it is possible that mismatched information would be more valued during the learning of the priors or in conditions where the changes are more unexpected. Evidence showing different behavioural and neurophysiological correlates of expected and unexpected uncertainty is in line with this view (Yu and Dayan, 2005). For example, in tasks in which the contingencies are stable, encountering low-probability trials is quite expected, and the value of those stimuli in predicting future events is suppressed via cholinergic neurotransmission (Witte et al., 1997). In contrast, in tasks in which the probabilistic structure of the environment changes unexpectedly, encountering low-probability stimuli boosts learning through the release of norepinephrine (Yu and Dayan (2005)), an effect that might also result in increased episodic memory encoding.

To characterise the contingency-learning process we fit four different reinforcement learning models to participants' data: An optimal model with a fixed-decreasing learning rate, a model with a free-decreasing learning rate, a free-learning rate model considering the outcome of participants' choice (evaluative model), and a free-learning rate outcome-free model considering the information given on each trial (instructive model). Model comparison showed that participants' learning processes did not conform to the normative behaviour of a Bayesian model, which prescribes that the optimal way of learning in this task entails decreasing the learning rate proportionally to the inverse of the number of the trials. In fact, participants data was best explained by models estimating individual, constant learning rates. In addition, participants' learning data were best explained by a model in which the context-category associations were learned by increasing the strength of the associations of the category presented on a given trial and decreasing the strength of the associations of the categories not presented, regardless of participants' choice and its outcome. This result suggests that prediction outcome might not be important for PE-driven incremental learning of associations, while it is crucial for modulating the influence of PE on the encoding of item identity.

Our results also showed that the estimated learning rate did not affect overall recognition memory, in contrast to our prediction. This can be explained by the fact that

the estimated learning rate was fixed throughout the experiment. As the learning rate has been suggested to depend on the level of uncertainty (e.g., Behrens et al., 2007), in our paradigm in which the contingencies were kept fixed, it makes sense that the learning either remained constant or potentially decrease over time if task length would have been longer. Furthermore, a learning rate was estimated for each individual in our study, akin to an individual difference measure. It is possible that the potential effect of learning rate on memory is rather a within-person process, observable only in paradigms in which learning rate changes (for example, when environmental contingencies change). It is also important to note that our interest focused on conditions where the contingencies were already learned, while its effect during the learning of the contingencies were not addressed. However, the effects of learning rate might vary considerably in conditions in which participants learn the contingencies. Future studies are needed to examine the dynamic relationships between learning rate and memory over time.

Interestingly, our results showed that the PE by prediction outcome interaction correlated with an index of reinforcement learning model fit, suggesting that this effect was stronger for participants whose behaviour was more similar to a reinforcement learning agent. In the present study, we selected the best fitting model as the model that fitted best to most participants. However, this is not an optimal solution, as for several participants other models fitted best. Therefore, future work should consider allowing different best fitting models for different individuals, for example by treating the model as a random factor. In conclusion, the current study provides novel evidence on the effects of PE on memory encoding. We show that the effects of PE on memory are modulated by whether or not a prediction is correct, suggesting a dependency between hippocampal and striatal dopaminergic systems, hence informing future studies exploring the interactions between learning and memory.

## 3 Methods

### 3.1 Participants

As experiment 1 and 2 were conceptual replication with similar design, both experiments are described together in the following. Differences between the experiments are pointed out.

In experiment 1, thirty-two young adults (20 female; mean age = 22.59 years,  $sd = 3.18$ ) were recruited through advertisements placed at the Goethe University campi in Frankfurt am Main. In exchange of participation, participants received either course credits or a monetary reimbursement of 8 €/hour. In experiment 2, 40 participants (19 female; mean age = 24.87,  $sd = 4.64$ ) were recruited through the Prolific platform <https://www.prolific.co/>. All participants had normal or corrected-to-normal vision and no history of psychological or

neurological disorders. All participants gave written informed consent prior to participation. The study was approved by the ethics committee of the Goethe University Frankfurt am Main.

## 3.2 Materials

For a more detailed description of the materials and methods used, please refer to the original publication (Ortiz-Tudela et al., 2021). For experiment 1, six coloured scene categories depicting real world outdoor locations taken from the ECOS database (<https://sites.google.com/view/ecosdata>) were used as contexts (see Figure 1). The selected scene categories were beach, mountain, road, desert, savannah, and seabed. As objects, 192 coloured images depicting real world objects were collected from an online search and were used as target objects. The images selected included the same number of objects for three different object-categories: musical instruments, fruits/vegetables, and household objects. All images were subjected to creative commons licensing and are available at <https://github.com/ortiztud/premup>. For experiment 2, the number and types of the scene categories were the same as experiment 1. However, the objects categories were reduced and only two were used: musical instruments and household objects.

## 3.3 Design and procedure

In experiment 1, participants completed the learning, encoding, and retrieval phases in one session, while in Experiment 2 participants completed the learning phase in a first session and the encoding and retrieval phase approximately 24 hours later. In addition, in the second session of experiment 2 participants worked on an extra reminder block of contingency learning before the encoding phase. In Experiment 1, stimulus presentation and recording of the responses was done using MATLAB’s Psychtoolbox (Brainard, n.d.) in a 60 Hz monitor (resolution: 1680 x 1050, full HD). Experiment 2 was moved online due to the COVID-19 pandemic, and some necessary changes were implemented. Stimulus presentation and response collection were programmed in PsychoPy v2021.1.4 and hosted online in Pavlovia (<https://pavlovia.org>). At the beginning of each session, the experimenter met the participant in a virtual room using an online video-conferencing tool, during which the appropriateness of the testing setup was assessed with a brief set of questions about the participant’s overall well-being, about the physical room in which the task would be performed and about the computer that would be used. Experimenters ensured that all participants were sitting in a quiet room, used a laptop or a desktop computer and were encouraged to minimize distractions as much as possible during the session. At the end of the session, the experimenter met the participant again and ask them about any unforeseen



event or situation that might have come up during the completion of the task. Finally, to maximize engagement, self-administered breaks were included after every 40 trials during the contingency learning and the encoding phases.

**Contingency learning phase.** Participants were presented with the scene contexts and were instructed to learn which object category was more likely to belong to each of the scene contexts; they were told that some contexts were easier to learn than others, but the exact contingencies were not explicitly given. A fixation cross at the center of the screen marked the beginning of each trial and lasted for 500 ms. After that, a scene image including a rectangular white patch with a question mark was presented. They were then asked to make a prediction about the object category that they thought they would encounter in that context. Three response alternatives were given for Experiment 1 (i.e. musical instruments, fruits/vegetables, and household objects) and only two for Experiment 2 (i.e. musical instruments and household objects). Category reminders were placed at the bottom of the screen and participants could choose among them by pressing one of three arrow keys in Experiment 1 (left arrow, down arrow, right arrow) and two arrow keys in Experiment 2 (left arrow, right arrow). The selected category was highlighted with a yellow frame. After 2 seconds from scene onset, the question mark within the white patch was replaced by an object and the coloured frame changed colour to indicate correct or incorrect response. Specifically, red frame indicated incorrect responses, green frame indicated correct responses. Object and feedback were shown on the screen for 1 second. Participants were told to use the feedback to learn the contingencies over trials.

The frequency to which an object category was encountered in the given scene contexts was manipulated to create different prior strengths. The prior strengths were "Flat" and "Strong" in Experiment 1, and "Flat", "Weak", and "Strong" in Experiment 2. In Experiment 1, the "Strong" prior condition consisted on three scene contexts in which one of the three object categories was frequently presented 80 % of the trials, while the other two were equally presented 10 % of the trials each. Conversely, the "Flat" prior condition consisted of three scene contexts in which the object categories were all three equally probable, being presented each 33 % of trials. In Experiment 2, the "Strong" prior condition was composed by two context scenes where one of the two object categories was presented 90 % of the trials, while the other object category was presented on 10 % of the trials. Two more scene contexts belonged to the "Weak" prior condition, in which the more frequently presented object category was shown on 70 % of the trials, while the other object category appeared on 30 % of the trials. Finally, the "Flat" prior condition included two scene contexts in which both object categories were equally likely to be presented, appearing each one

on 50 % of the trials. In order to achieve the desired contingencies without proportionally increasing the number of individual objects used, different objects were repeated a different number of times depending on its category and on the contexts in which they were shown. The association of each object category to each scene category was counterbalanced across participants so that across the entire sample, every object category was paired with every scene category.

**Encoding Phase.** The encoding Phase in Experiment 1 and 2 was similar to the learning phase, with only minor changes introduced. The explicit feedback represented by the coloured squared surrounding the object was removed in this phase. In addition, a new set of objects was used, and each of these objects was presented only once. To equate the number of objects in each critical cell for our analysis, we selected a fixed number of objects ( $n=20$ ) for each PE condition, and these were presented only once. Then, to achieve the desired contingencies for each scene category, we used filler objects from the same object categories and repeated them 7 times. Filler trials were not considered for recognition memory.

Similarly to the contingency learning phase, participants’ task was to predict which object category followed a scene context that was presented on every trial. The contingencies between object categories and scenes were the same as the previous learning phase.

**Retrieval Phase.** In the object recognition test, all the objects from encoding phase together with 192 new objects were used. In Experiment 1, hit rate was calculated on a sample of half of the 192 objects (96 trials), as half of the trials were selected for the immediate recognition session which is the object of the analysis of the current study. The rest of the trials were selected for a delayed recognition test, which was added to explore the potential modulating factor of consolidation on the interplay between PE and memory and it is not considered in the current study. In experiment 2, all the 192 old objects were considered for hit rate calculation. Trials started with a fixation cross for 500ms, and objects were presented in isolation at the center of the screen. Participants were required to make old/new judgements. All the responses in the retrieval phase were self-paced and not time-constrained, and the display stayed unaltered until participants made a response. After that, a new trial was then presented.

### 3.4 Computational Models

We fitted participants’ contingency learning and encoding data with computational models. The models considered are all different version of a standard Rescorla-Wagner model (or Q-learning) (Sutton and Barto, 1998; Daw, 2011). For each scene category, the model estimates a trial-level variable  $Q$  for each object category included in the experiments (three in

experiment 1 and two in experiment 2). These  $Q$  values reflect the strength of participants' belief that a certain object category (for example, "Instruments") will be presented in a specific context (for example, "beach"). Since we have  $N$  object categories for each  $n$  context, the estimates  $\hat{Q}$  of the probabilities can be represented by the following  $j$ -by- $c$  matrix:

$$\begin{bmatrix} \hat{Q}^{1,1}, & \hat{Q}^{1,2}, & \dots \\ \hat{Q}^{2,1}, & \hat{Q}^{2,2}, & \dots \\ \dots, & \dots, & \hat{Q}^{j,c} \end{bmatrix} \quad (1)$$

Where  $\hat{Q}^{1,1}$  represent the expected value  $Q$  for category  $j=1$  in context  $c=1$ . For all the models considered in this study, the estimated values are stored in a category  $j$  by context  $c$  matrix as this one, and initialize as  $\hat{Q}^{j,c} = 0.33$  in experiment 1, and  $\hat{Q}^{j,c} = 0.5$  in experiment 2.

**Decreasing Learning Rate Instructive Model (dLRI)** First, to provide a normative Bayesian solution, we used a Dirichlet-multinomial model that we re-formulated to a delta-rule model (see XX). This model learned according to prediction errors scaled by a learning rate to produce an update of the estimated category probabilities. The learning rate dynamically decreased across trials so that the model learned more rapidly in the beginning of the task, and more slowly on later trials. Formally, the model sequentially updated the category probabilities for each context according to

$$\hat{Q}_{t+1}^{j,c} = \hat{Q}_t^{j,c} + \frac{1}{t} \delta_t, \quad (2)$$

where  $\hat{Q}_{t+1}^{j,c}$  denotes the estimate of the probability of category  $j$  in context  $c$ ,  $j$  at the next trial  $t+1$ . This estimate is based on the current estimate of the category probabilities ( $\hat{Q}_{t,j}^{j,c}$ ) and the prediction error  $\delta$ , calculated as:

$$\delta = r_t^j - \hat{Q}_t^{j,c} \quad (3)$$

where the feedback  $r_t^j$  represents an array of  $N$ -by- $j$  elements, in which each element refers to a category  $j$ , defined as following:

$$r_t^j = \begin{cases} 1 & \text{if } j = j_t \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

The values of the array are 1 if category  $j$  is present on trial  $n$ , and 0 if it is not. Therefore the model is assuming that a value estimate for an object category that appears on a trial incrementally increases as a result of a prediction error until  $\hat{Q}_t^{j,c}$  reaches its asymptote of 1.

Conversely, the value estimates of categories that are not presented on trial  $t$  decrease as a result of a negative prediction error, unless  $Q_t^{j,c}$  for those categories has already a value of 0. Therefore, this model only uses instructive feedback, which indicates what is the correct choice, independently of participants' action. The learning rate  $1/t =: \alpha$  indicates to which degree the prediction error influences the updated estimate of the category probabilities. Given that the learning rate in our case directly depends of the number of completed trials  $t$  for a context  $c$ , it continuously decays as a function of trials. This principle ensures that the influence of prediction errors is stronger at the beginning of the task. For more information about the formalization of the optimal Bayesian model, see Supplemental Material.

**Decreasing Free Learning Rate Instructive Model (dfLRI)** The dLRI shows how an optimal agent should update the expected values. However, participants' behaviour may be far from optimal. For this reason, the dfLRI allows each participant to have its own learning rate  $\alpha$ , which decreases as a function of the trial number, similarly as in the dLRI:

$$Q_{t+1}^{j,c} = Q_t^{j,c} + \alpha \frac{1}{t} \delta_t, \quad (5)$$

where  $\hat{Q}_{t+1}^{j,c}$  and  $\delta_t$  are estimated as in the previous model (dLRI).

**Free Learning Rate Instructive Model (fLRI)** This model estimates learning rate by participant, as in the previous dfLRI model. However, the present model consider the learning rate  $\alpha$  to remain constant throughout the learning and encoding phases. The expected values are thus updated according to the following rule:

$$Q_{t+1}^{j,c} = Q_t^{j,c} + \alpha \delta_t, \quad (6)$$

while  $\delta$  is the same as in equation 3. Also, note that this model uses the same instructive feedback as in equation 4.

**Free Learning Rate Evaluative Model (fLRE)** This model still allows participants to have a fixed learning rate  $\alpha$ . However, this model assumes that the feedback depends on the actions that participants take. The expected values are thus updated as follows:

$$Q_{t+1}^{j,c} = \begin{cases} \alpha \delta_t & \text{if } a_t = j_t \\ Q_t & \text{otherwise} \end{cases} \quad (7)$$

where  $a_t$  is the object-category selected by participants on a given trial,  $\delta_t$  is calculated as in equation 3, and  $r_t$  is 1 if the choice is the correct one, and 0 otherwise.

**Table 3:** Priors for the parameters

Parameter	Priors
$\alpha$	$\sim U(0, 1)$
$\beta$	$\sim \exp(1)$

### 3.4.1 Action Selection

The expected  $Q$ -values computed through the models listed above were translated into choice probabilities by implementing a softmax rule as follows:

$$P_t^{j,c} = \frac{\exp(\beta Q_t^{j,c})}{\sum_{j=1}^j (\exp(\beta Q_t^c))}, \quad (8)$$

where  $P_t^{j,c}$  represents the probability of choosing a specific object-category  $j$  for a defined scene category  $c$ . The inverse temperature parameter  $\beta$  is another free parameter that modulates the stochasticity of the choice, with higher values meaning more deterministic actions and lower values more stochastic choices.

### 3.4.2 Parameter Recovery

Before fitting the models to participants' data, a parameter recovery procedure was run for both Experiment 1 and Experiment 2, in order to check whether the fitting procedure for each model gave meaningful parameters and to find potential parameters boundaries. *Surrogate* data with randomly sampled known parameters were simulated, then the models were fit to the simulated data (see [Wilson and Collins, 2019](#)). The priors from which the simulation parameters were sampled are shown in [table 3](#). In order to fit the data, we used maximum likelihood estimation (see section). Because the models are designed to reflect participants' learning, only the strong (Experiment 1) and strong and weak (Experiment 2) conditions were simulated. High correlation between simulated and fitted data indicates that the model successfully recovered the parameters that were used to generate the data. First attempts to recover the parameters allowed to set the boundaries for inverse temperature parameters. Plots of the parameter recovery are shown in [figure ??](#).

### 3.4.3 Model Recovery

Besides parameter recovery, another procedure to evaluate the reliability of the model is model recovery ([Wilson and Collins, 2019](#)). The aim of model recovery is to determine that a model, among several ones, can successfully be indicated to be the one to have generated the data. To achieve this, data of the four different models were simulated (with randomly sampled parameters) and then fit to each of the models. The models were then compared to determine which one fitted the data best. The method used to assess the fit of the models was the Bayesian Information Criterion,  $BIC$ , which incorporates a penalty for the number

of parameters:

$$BIC = -2\log\hat{LL} + k_m\log(T), \quad (9)$$

where  $\hat{LL}$  is the log-likelihood value when the model is fitted with the best fitting parameter, and  $K_m$  is the number of parameters in the model  $m$ . Lower values of  $BIC$  mean better fit. The comparison between the models for each set of generated data was repeated 100 times to generate the confusion matrices shown in figure [S3](#).

### 3.5 Parameter Estimation and Model Comparison

The models were finally fit to participants' data in order to estimate the parameters. The parameters of best fit for each model were estimated through maximum likelihood estimation. This procedure allowed to find the parameters  $\theta$  that maximize the likelihood of the data given the parameters  $p(d_{1:t}|\theta, m)$ . The probability of the whole dataset  $d$  is calculated as the product of the choice probabilities  $p(c_t|d_{1:t-1}, \theta, m)$ . As the product of the choice probabilities is often a very small number, it is common practice to use the log-likelihood instead, which is the sum of the log of the choice probabilities [Daw \(2011\)](#); [Wilson and Collins \(2019\)](#):

$$LL = \sum_{t=1}^n \log p(c_t|d_{1:t-1}, \theta, m) \quad (10)$$

where  $p(c_t|d_{1:t-1}, \theta, m)$  is the probability of each single choice given the parameter  $\theta$ , the model  $m$ , and all the data up to that point.

The search over the full set of free parameters was optimized through the package *optim* in R, which was fed with the negative log likelihood and a set of starting points randomly selected from the priors shown in Table [3](#). The  $\alpha$  parameter was constraint between 0 and 1, while the  $\beta$  parameter between 0 and 10, as parameter recovery shown that for values that exceeded 10 the model could not distinguish between different beta parameters. Because the optimizer may find a local rather than a global, we run the search for the best parameters five times, starting from different points, and then used the best winning parameters among the five iterations, i.e. the parameters that minimized the log-likelihood. After estimating the parameters, a  $BIC$  value(see equation [9](#)) was computed for each model and for each participant using the parameters of best fit.

To compare the fit of the models, we calculated the average  $BIC$  across all subjects for each model, then counted the number of participants for which each model was the best fit. In addition, we used the model evidence of the best model within each participant: Following [Raftery \(1995\)](#) and [Gluth et al. \(2017\)](#), model evidence was defined as "weak", "positive", "strong", or "very strong" depending on the  $BIC$  difference between the best and

the second best model for each participant. Precisely, evidence was "weak" when the *BIC* difference between best and second best model was below 2, "positive" when it was between 2 and 6, "strong" when it was between 6 and 10, and "very strong" when it was above 10.

### 3.6 Statistical Analysis

In order to test the statistical significance of the effects of interest, we used linear mixed-effect models and generalized linear mixed-effect models, implemented in R through the package *lme4* (Bates et al., 2014). Because our main outcome variable (memory) is binary, we used the logit link function in the binomial family to fit the models to accuracy data. Participants were modelled as random intercepts, while the explanatory variables and their interactions were modelled as both fixed and random effects. The resulting generalized linear-mixed effect model can be formalized as the following:

$$p(Hit)_{i,j} = \frac{1}{1 + \exp - (\beta_{0,j} + \beta_{1,j}PE + \beta_{2,j}PO + \beta_{3,j}PE \cdot PO + \epsilon_{i,j})}. \quad (11)$$

The formula represents the probability of remembering an object  $i$  for a participant  $j$ . The intercept  $\beta_{0,j}$  is composed by a common (fixed) intercept for the population (i.e. the average  $p(Hit)$ ) plus a subject-specific random effect.  $\beta_{1,j}PE$  and  $\beta_{2,j}PO$  represent the slopes for PE and prediction outcome, respectively, while  $\beta_{3,j}PE \cdot PO$  refers to their interaction. Each of slope coefficients is formed by a common (fixed) slope for the population level plus a subject-specific random slope. Finally,  $\epsilon_{i,j}$  represents the within-participant residual term. Note that this formula does not include the effects of learning rate, which varied between-participants, and thus was included to the generalized mixed-effect model only as fixed effect and fixed interaction term. The variance-covariance matrix for the random effects was set as unstructured, so that the covariances between the random terms could take any finite positive value. Therefore, we used the maximal random effect structure justified by the design (Barr, 2013). The test of the significance of the parameters was obtained through Wald chi-square test. Effect sizes were reported as  $\exp(\beta)$ , which represent odds ratio, which represents the change in odds. The odds are in turn calculated as follows:

$$Odds = \frac{p(Hit = 1)}{1 - p(hit = 1)} \quad (12)$$

In the analysis of the effect of binned PE, we used a linear mixed-effect model, with hit rate as response variable:

$$HitRate = \frac{hits}{hits + missed}. \quad (13)$$



Binned PE was treated as a categorical variable with four levels. Testing for significance of planned contrasts was corrected for multiple comparison by using Bonferroni correction:

$$p_{corr} = p \cdot k, \quad (14)$$

where  $p$  is the  $p$  value of the comparison and  $k$  is the overall number of comparisons considered in a model.

## References

- Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in psychology*, 4:328.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.
- Behrens, T. E. J., Woolrich, M. W., Walton, M. E., and Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *NATURE NEUROSCIENCE*, 10(9).
- Bethus, I., Tse, D., and Morris, R. G. M. (2010). Dopamine and Memory: Modulation of the Persistence of Memory for Novel Hippocampal NMDA Receptor-Dependent Paired Associates. *Journal of Neuroscience*, 30(5):1610–1618.
- Bromberg-Martin, E. S. and Hikosaka, O. (2011). Lateral habenula neurons signal errors in the prediction of reward information. *Nature neuroscience*, 14(9):1209–1216.
- Davidow, J. Y., Foerde, K., Galván, A., and Shohamy, D. (2016). An Upside to Reward Sensitivity: The Hippocampus Supports Enhanced Reinforcement Learning in Adolescence. *Neuron*, 92(1):93–99.
- Daw, N. D. (2011). Trial-by-trial data analysis using computational models. *Decision Making, Affect, and Learning: Attention and Performance XXIII*, pages 1–26.
- Daw, N. D. and Tobler, P. N. (2014). Value learning through reinforcement: The basics of dopamine and reinforcement learning. In *Neuroeconomics*, pages 283–298.
- De Loof, E., Ergo, K., Naert, L., Janssens, C., Talsma, D., Van Opstal, F., and Verguts, T. (2018). Signed reward prediction errors drive declarative learning. *PLoS One*, 13(1):e0189212.
- Ergo, K., Loof, E. D., and Verguts, T. (2020). Reward Prediction Error and Declarative Memory. *Trends in Cognitive Sciences*, 24(5):388–397.

- Friston, K. (2018). Does predictive coding have a future? *Nature Neuroscience*, 21(8):1019–1021.
- Ghosh, V. E. and Gilboa, A. (2014). What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia*, 53(1):104–114.
- Gluth, S., Hotaling, J. M., and Rieskamp, J. (2017). The attraction effect modulates reward prediction errors and intertemporal choices. *Journal of Neuroscience*, 37(2):371–382.
- Jang, A. I., Nassar, M. R., Dillon, D. G., and Frank, M. J. (2019). Positive reward prediction errors during decision making strengthen memory encoding. *Nature Human Behaviour*, 3(July):719–732.
- Lefebvre, G., Lebreton, M., Meyniel, F., Bourgeois-Gironde, S., and Palminteri, S. (2017). Behavioural and neural characterization of optimistic reinforcement learning. *Nature Human Behaviour*, 1(4):1–9.
- Lemon, N. and Manahan-Vaughan, D. (2006). Dopamine d1/d5 receptors gate the acquisition of novel information through hippocampal long-term potentiation and long-term depression. *Journal of Neuroscience*, 26(29):7723–7729.
- Lisman, J. E. and Grace, A. A. (2005). The hippocampal-VTA loop: Controlling the entry of information into long-term memory. *Neuron*, 46(5):703–713.
- Mather, M. and Sutherland, M. R. (2011). Arousal-biased competition in perception and memory. *Perspectives on psychological science*, 6(2):114–133.
- McClure, S. M., Berns, G. S., and Montague, P. R. (2003). Temporal Prediction Errors in a Passive Learning Task Activate Human Striatum. *Neuron*, 38(2):339–346.
- Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. MIT Press, Cambridge.
- Niv, Y. and Chan, S. (2011). On the value of information and other rewards. *Nature Neuroscience* 2011 14:9, 14(9):1095–1097.
- Niv, Y. and Schoenbaum, G. (2008). Dialogues on prediction errors. *Trends in Cognitive Sciences*, 12(7):265–72.
- Ortiz-Tudela, J., Nolden, S., Pupillo, F., Ehrlich, I., Schommartz, I., Turan, G., and Shing, Y. L. (2021). Not what u expect: Effects of prediction errors on episodic memory. <https://psyarxiv.com/8dwb3/>.
- Poldrack, R. A., Clark, J., Paré-Blagoiev, E., Shohamy, D., Creso Moyano, J., Myers, C., and Gluck, M. (2001). Interactive memory systems in the human brain. *Nature*, 414(November):546–550.

- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological methodology*, pages 111–163.
- Rangel, A., Camerer, C., and Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, 9:545–556.
- Rosen, Z. B., Cheung, S., and Siegelbaum, S. A. (2015). Midbrain dopamine neurons bidirectionally regulate ca3-ca1 synaptic drive. *Nature neuroscience*, 18(12):1763–1771.
- Rouhani, N. and Niv, Y. (2021). Signed and unsigned reward prediction errors dynamically enhance learning and memory. *eLife*, 10(Lc):1–28.
- Rouhani, N., Norman, K. A., and Niv, Y. (2018). Dissociable Effects of Surprising Rewards on Learning and Memory. *Journal of Experimental Psychology: Learning Memory, and Cognition*, (Advance online publication).
- Rushworth, M. F. and Behrens, T. E. J. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nature Neuroscience*, 11(4):389–97.
- Schultz, W. (2016). Dopamine reward prediction error coding. *Dialogues in Clinical Neuroscience*, 18(1):23–32.
- Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward. *Science (New York, N.Y.)*, 275(5306):1593–1599.
- Sharot, T. and Garrett, N. (2016). Forming beliefs: Why valence matters. *Trends in cognitive sciences*, 20(1):25–33.
- Sharot, T., Korn, C. W., and Dolan, R. J. (2011). How unrealistic optimism is maintained in the face of reality. *Nature neuroscience*, 14(11):1475–1479.
- Sharot, T., Riccardi, A. M., Raio, C. M., and Phelps, E. A. (2007). Neural mechanisms mediating optimism bias. *Nature*, 450(7166):102–105.
- Steinberg, E. E., Keiflin, R., Boivin, J. R., Witten, I. B., Deisseroth, K., and Janak, P. H. (2013). A causal link between prediction errors, dopamine neurons and learning. *Nature Neuroscience* 2013 16:7, 16(7):966–973.
- Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT Press, Cambridge.
- Sutton, R. S. and Barto, A. G. (2018). *Reinforcement Learning: An Introduction, Second Edition*, volume 258.

- Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., Witter, M. P., and Morris, R. G. (2007). Schemas and memory consolidation. *Science*, 316(5821):76–82.
- Watanabe, N., Bhanji, J. P., Ohira, H., and Delgado, M. R. (2019). Reward-driven arousal impacts preparation to perform a task via amygdala–caudate mechanisms. *Cerebral Cortex*, 29(7):3010–3022.
- Wilson, R. C. and Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *eLife*, 8:1–33.
- Witte, E., Davidson, M., and Marrocco, R. (1997). Effects of altering brain cholinergic activity on covert orienting of attention: comparison of monkey and human performance. *Psychopharmacology*, 132(4):324–334.
- Yu, A. J. and Dayan, P. (2005). Uncertainty, Neuromodulation, and Attention. *Neuron*, 46(4):681–692.

## Supplemental Materials

### 4 Supplemental Material

#### Dirichlet-multinomial model

We used the Dirichlet-multinomial model to formalize learning as optimal Bayesian updating of the Dirichlet distribution (Murphy, 2012). We apply the Multinomial distribution because the Categorical distribution that we use in our task is a special case of the Multinomial distribution where only one outcome is sampled on each trial. That is, the outcomes  $x_t$  of the task were drawn from a Multinomial distribution

$$\text{Mu}(x|1, \boldsymbol{\theta}) =: \text{Cat}(x|\boldsymbol{\theta}) \quad (\text{S1})$$

where  $p(x = j|\boldsymbol{\theta}) = \theta_j$ . For example, in the weak prior condition,  $\boldsymbol{\theta} = [0.33, 0.33, 0.33]$ . The Dirichlet distribution is the conjugate distribution of the Multinomial distribution and can therefore be utilized as a prior. This distribution is parameterized by the concentration parameters  $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \dots, \alpha_3\}$ .

We use the Dirichlet distribution to model the participants' prior expectations (i.e., at the beginning of the learning phase) about the category probabilities, which is often called pseudo-counts. Here we assume that participants start the task with a flat prior that reflects that all categories are equally likely, which corresponds to  $\boldsymbol{\alpha} = (1, 1, \dots, 1)$ . These values thus indicate the assumption that each category has been pseudo-counted once.

To obtain the posterior, the only operation required is adding the observed data to the prior. In order to obtain an estimate of the category probabilities, we can compute the expected value of the posterior, referred to as the maximum a posteriori (MAP) estimate:

$$\hat{\theta}_k = \frac{N_k + \alpha_k - 1}{N + \alpha_0 - K}. \quad (\text{S2})$$

Under the assumption that  $\boldsymbol{\alpha} = (1, 1, \dots, 1)$ , the MAP estimate is equal to the maximum likelihood (ML) estimate that is based on the empirically observed frequency of the categories:

$$\hat{\theta}_k = \frac{N_k}{N}. \quad (\text{S3})$$

#### 4.1 Delta-rule formulation

We now show how eq. S3 can be translated into the delta rule. Let  $\hat{\theta}_{n,j}$  denote the estimate of the  $j$ th category probability on trial  $n$ , then the estimate of the  $j$ th category at trial  $n+1$ , denoted by  $\hat{\theta}_{n+1,j}$ , can be computed according to

$$\begin{aligned}
\hat{\theta}_{n+1,j} &= \frac{n_j}{n} \\
&= \frac{1}{n} n_j \\
&= \frac{1}{n} \sum_{i=1}^n x_{i,j} \\
&= \frac{1}{n} \left( x_{n,j} + \sum_{i=1}^{n-1} x_{i,j} \right) \\
&= \frac{1}{n} \left( x_{n,j} + (n-1) \frac{1}{n+1} \sum_{i=1}^{n-1} x_{i,j} \right) \tag{S4} \\
&= \frac{1}{n} \left( x_{n,j} + (n-1) \hat{\theta}_{n,j} \right) \\
&= \frac{1}{n} (x_{n,j} + n\hat{\theta}_{n,j} - \hat{\theta}_{n,j}) \\
&= \hat{\theta}_{n,j} + \frac{1}{n} (x_{n,j} - \hat{\theta}_{n,j}) \\
&= \hat{\theta}_{n,j} + \alpha_n \delta_{n,k},
\end{aligned}$$

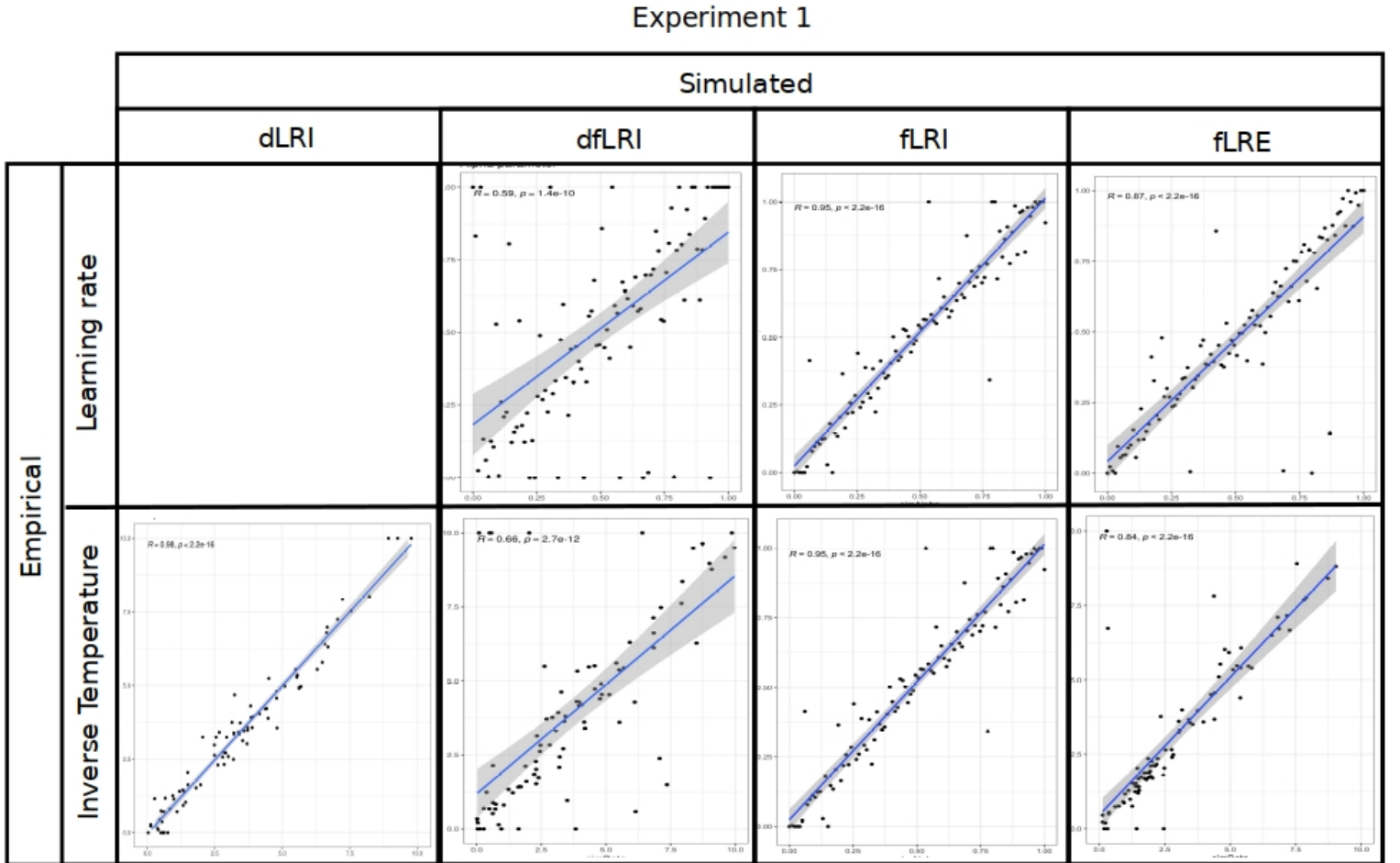
where  $\delta_{n,j} := (x_{n,j} - \hat{\theta}_{n,j})$  corresponds to the prediction error and  $\alpha_n := \frac{1}{n}$  is the learning rate (Sutton and Barto, 1998).

## Parameter Recovery

In order to check whether the models could successfully recover the parameters, *fake* data were first simulated with known parameters. Next, models were fit to the simulated data and parameters of best fit were estimated. Finally, the recovered parameters were compared with the known parameters. Graphs showing the results of parameter recovery are presented in Figure ???. High correlation between the simulated and fitted parameters indicates successful recovery. Note that the inverse temperature parameter was constrained between 0 and 10 as previous parameter recovery attempts showed that recovery was not reliable for parameters above that range.

## Model Recovery

The ability of a model can successfully distinguish between different models, a model recovery procedure was used. Data from the three different models were simulated and then fit to each of the models to determine which model fits best. this procedure was repeated 100 times. The confusion matrices shown in Figure S3 show the results of this procedure. Each cell represents the probability of data simulated by models in the X axis to be best fit by models in the Y axis. Higher probabilities in the diagonal means that the models can successfully recover the models from which the data were generated.



**Figure S1: Parameter Recovery.** Parameter Recovery for Experiment 1.

## Simulated and Actual comparison for dLRI and fLRE models

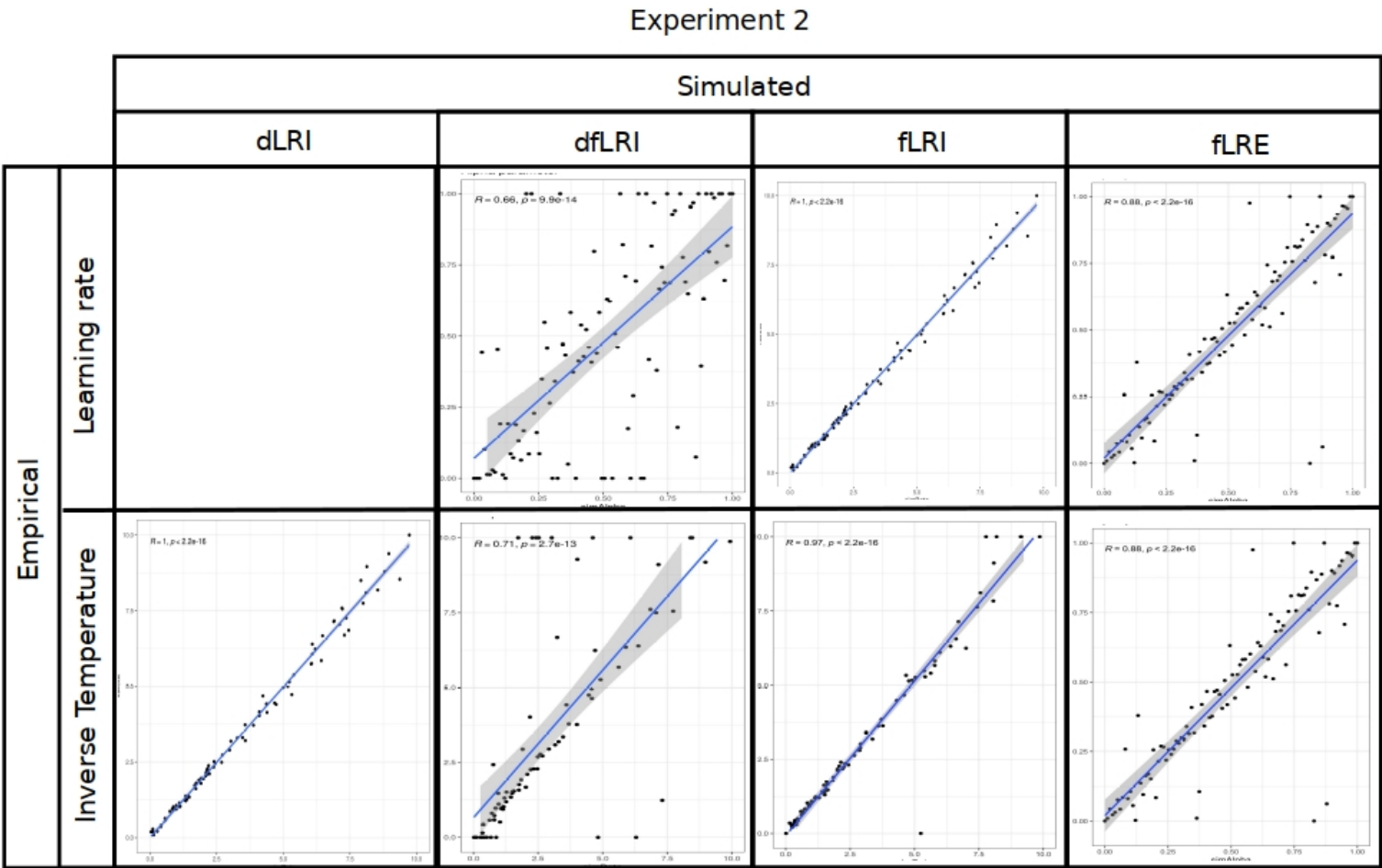
Figures S4 and S5 show a comparison between simulated and actual data for the models instructive model with a decreasing learning rate (dLRI) and the evaluative model with a free learning rate (fLRE).

## Analysis with binned Hit Rate

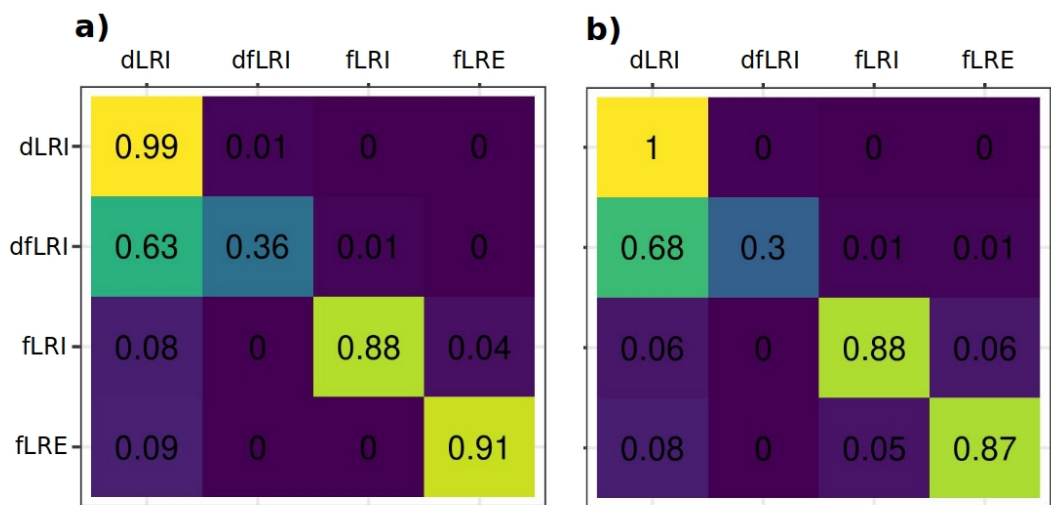
In order to compare memory at different levels of the model-derived PE, we calculated the quartiles for PE for each participant, separately for trials with correct and incorrect prediction outcome. We binned the hit rate by aggregating it between the quartiles, to create four bins which eventually were used as the explanatory variable in our analysis. A graph with the distribution of PE by binned data is shown in Figure S6.

Hit rate as a function of binned PE and prediction outcome is shown in Figure S7. We then tested for the three-way interaction between binned PE, prediction outcome, and experiment, in a linear mixed-effects model, adding participants as random effects. The three-way interaction was not significant,  $\chi^2_{(3)} = 4.23$ ,  $p = .238$ . In addition, the interactions between PE and experiment, and the interaction between prediction outcome and exper-



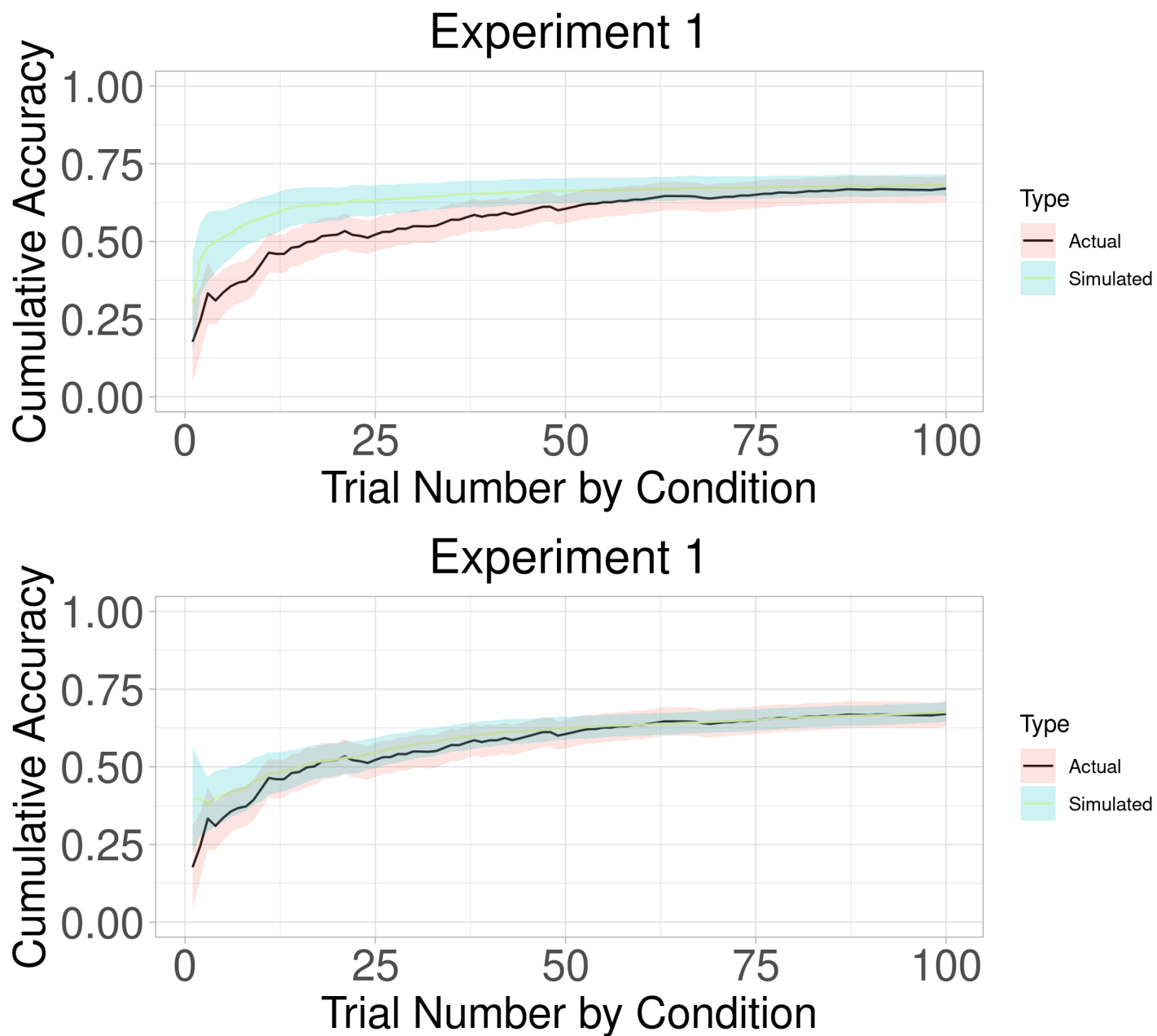


**Figure S2: Parameter Recovery.** Parameter Recovery for Experiment 2.



**Figure S3: Model Recovery.** Confusion Matrices showing model recovery for a) experiment 1 and b) experiment 2. The numbers show the probability of data generate by model X to be best fit by model Y.

iment were not significant ( $\chi^2_{(3)} = 1.68, p = .642, \chi^2_{(3)} = 0.81, p = 0.368$ , respectively). These results suggest that there were not significant differences in the effects of PE and in the interaction between PE and prediction outcome between the two experiments. By

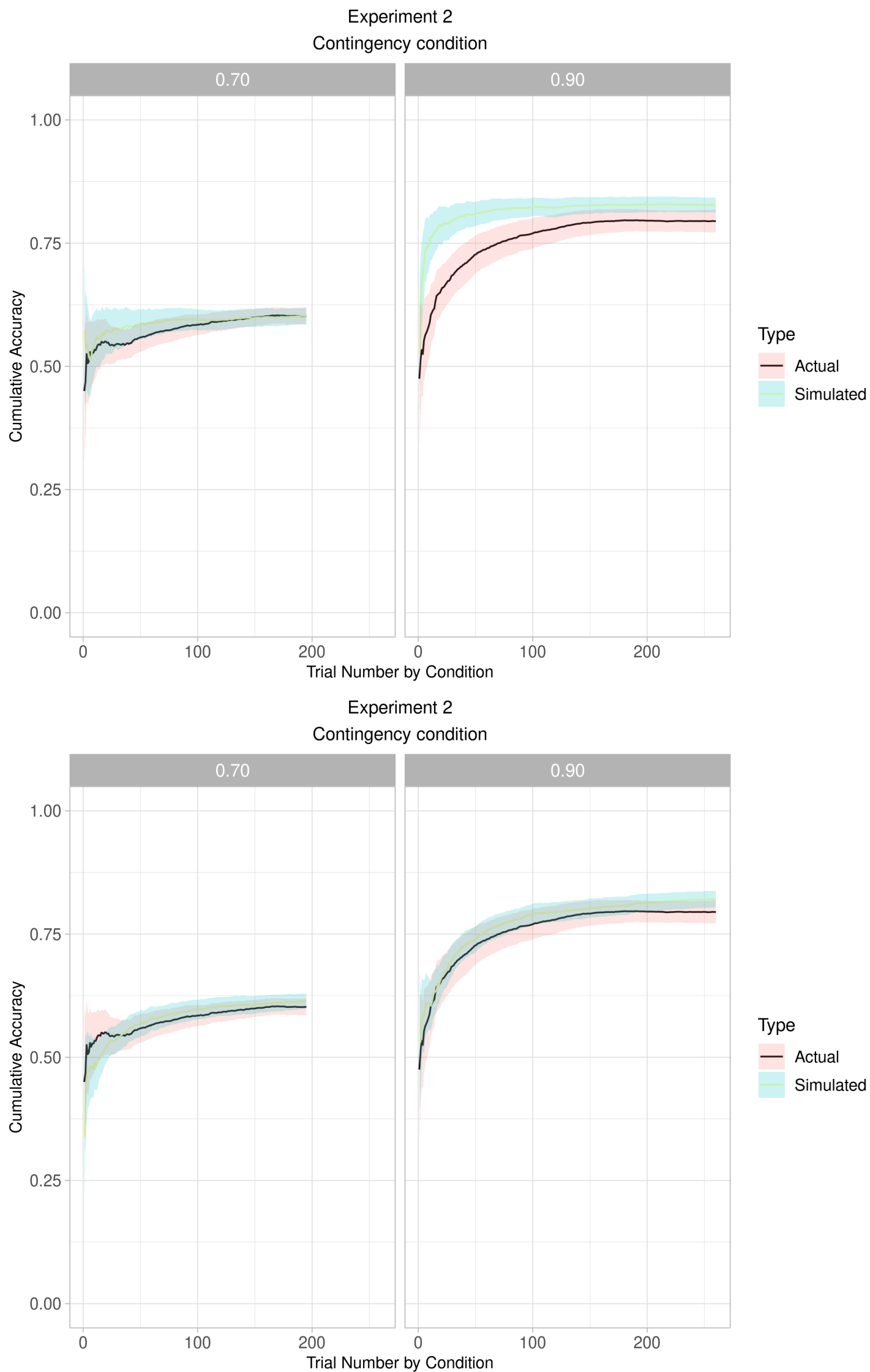


**Figure S4: Simulated vs Empirical Data for the Experiment 1 for the dLRI model (left) and fLRE (right).** Simulated data (red line) and actual data (green line) overlapped, for experiment 1, for weak priors condition and strong prior condition.

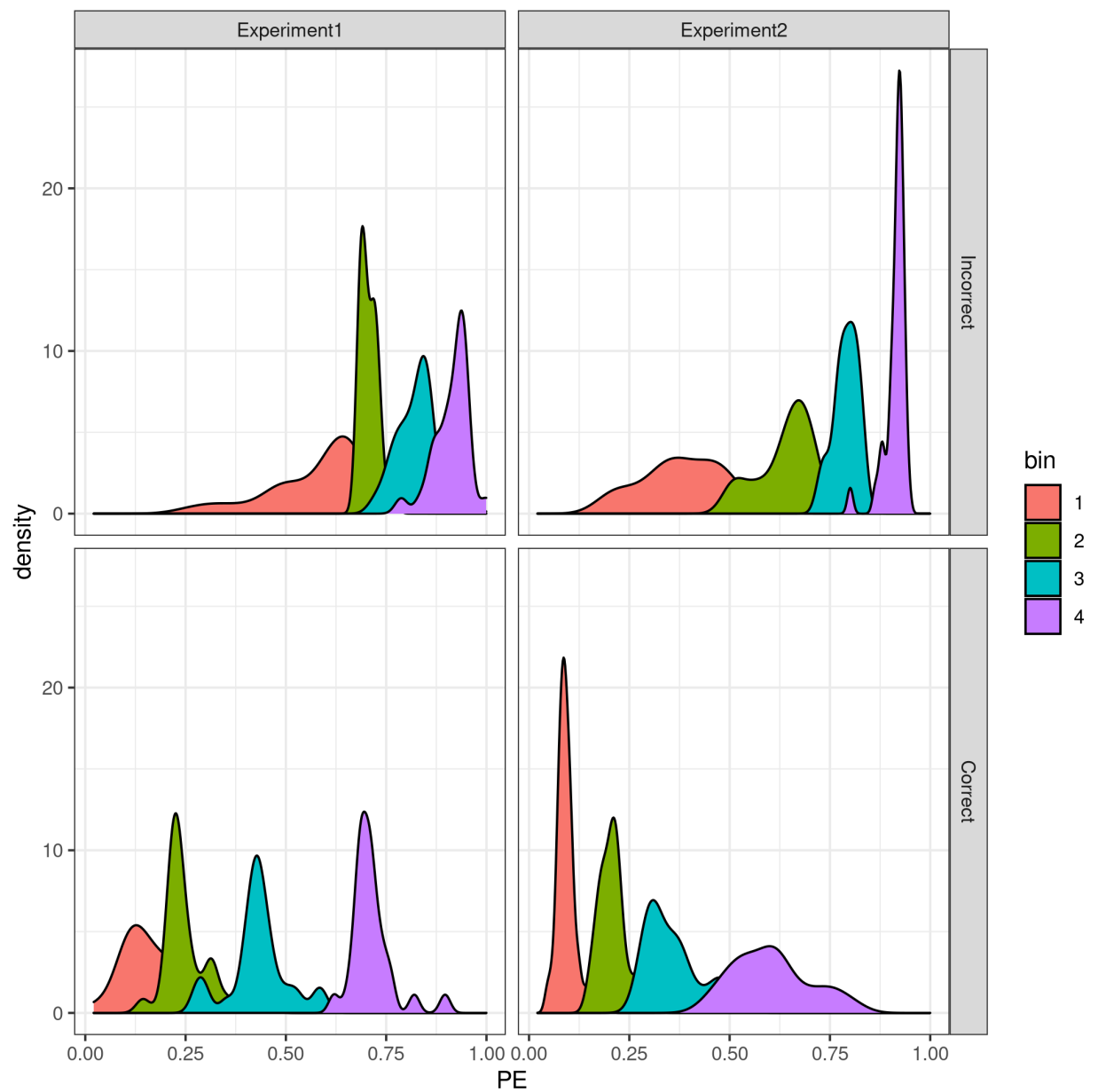
contrast, there was a main effect of experiment,  $\chi^2_{(3)} = 7.70$ ,  $p = .005$ , showing that overall participants' performance was significantly worse in Experiment 2, compared to Experiment 1. Importantly, there was also a significant interaction between PE and prediction outcome,  $\chi^2_{(3)} = 14.09$ ,  $p = .003$ .

To break down the interaction, the effect of PE on recognition was analyzed separately for correct and incorrect prediction outcomes. We compared each bin with the first one, to test whether increasingly higher PE significantly affected memory encoding. Results showed that for incorrect prediction outcomes, the difference between the first and the second was significant,  $\beta = -0.0557$ ,  $p_{corr} = .040$ , OR = 1.06. In addition, the comparisons between the third and the first, and the fourth and the first, were both significant, ( $p_{corr} < .001$ ). For correct prediction outcomes, the comparison between first and second quantile, and first and third quantile did not reach significance ( $p_{corr} > .114$ ), whereas the comparison between the fourth and the first quantile was significant,  $\beta = 0.081$ ,  $p_{corr} = .009$ , OR = 1.08. These

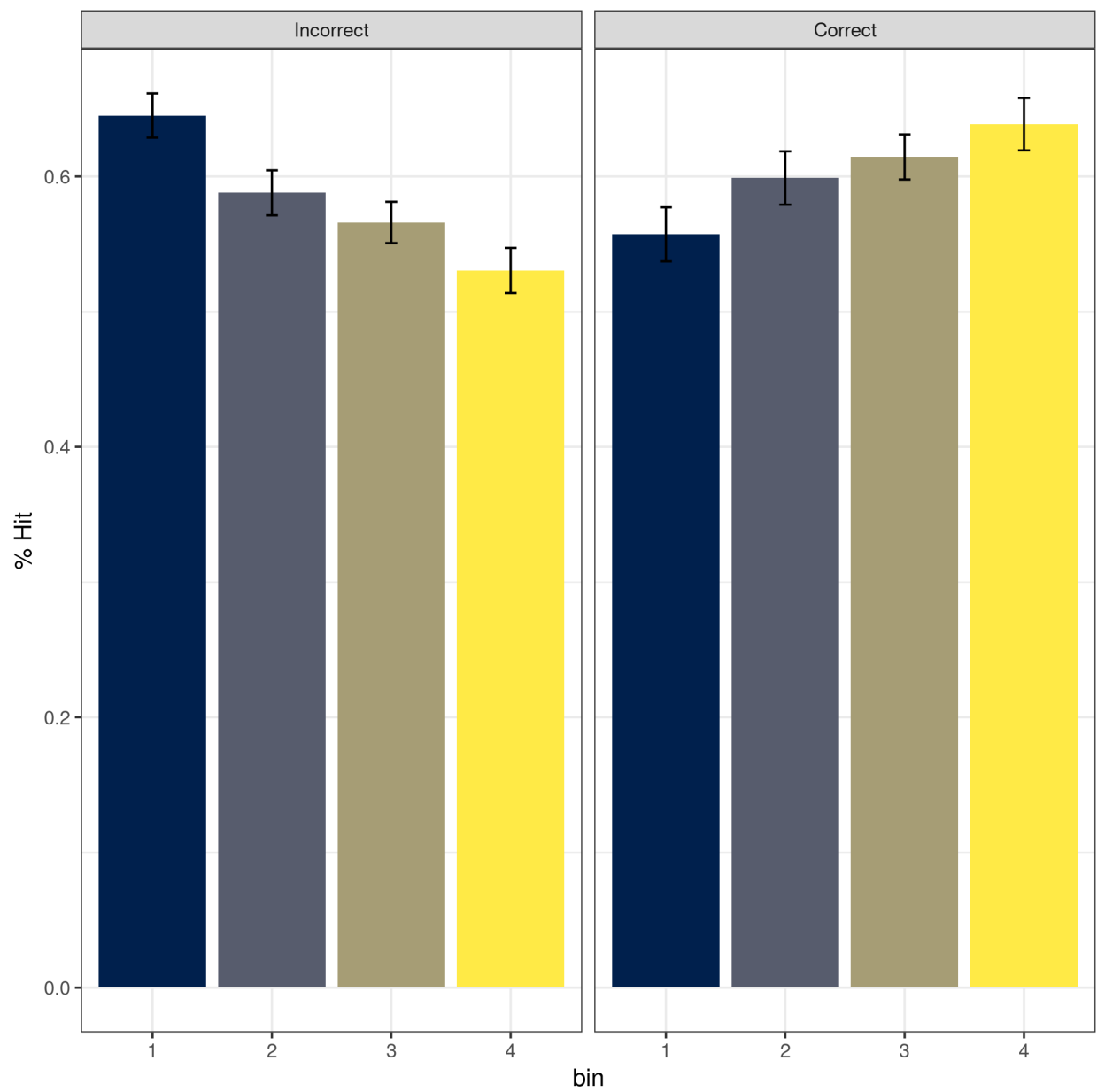
results suggest that while a PE error generated by incorrect prediction impairs memory even when prior expectations are not very strong, for correct predictions a higher PE is needed in order to observe benefits for memory encoding.



**Figure S5: Simulated vs Empirical Data for the Experiment 2 for the dLRI (left) and fLRE (right).** Simulated data (red line) and actual data (green line) overlapped, for experiment 2, for weak priors condition and strong prior condition.



**Figure S6: Distribution of binned PE.** Distribution of PE after binning it, as a function of prediction outcome and experiment.



**Figure S7: Hit Rate by Binned PE.** Effect of binned PE on hit rate as a function of prediction outcome.