Prediction error and Memory:
Insights from computational models

Pupillo, Ortiz-Tudela, Bruckner, & Shing

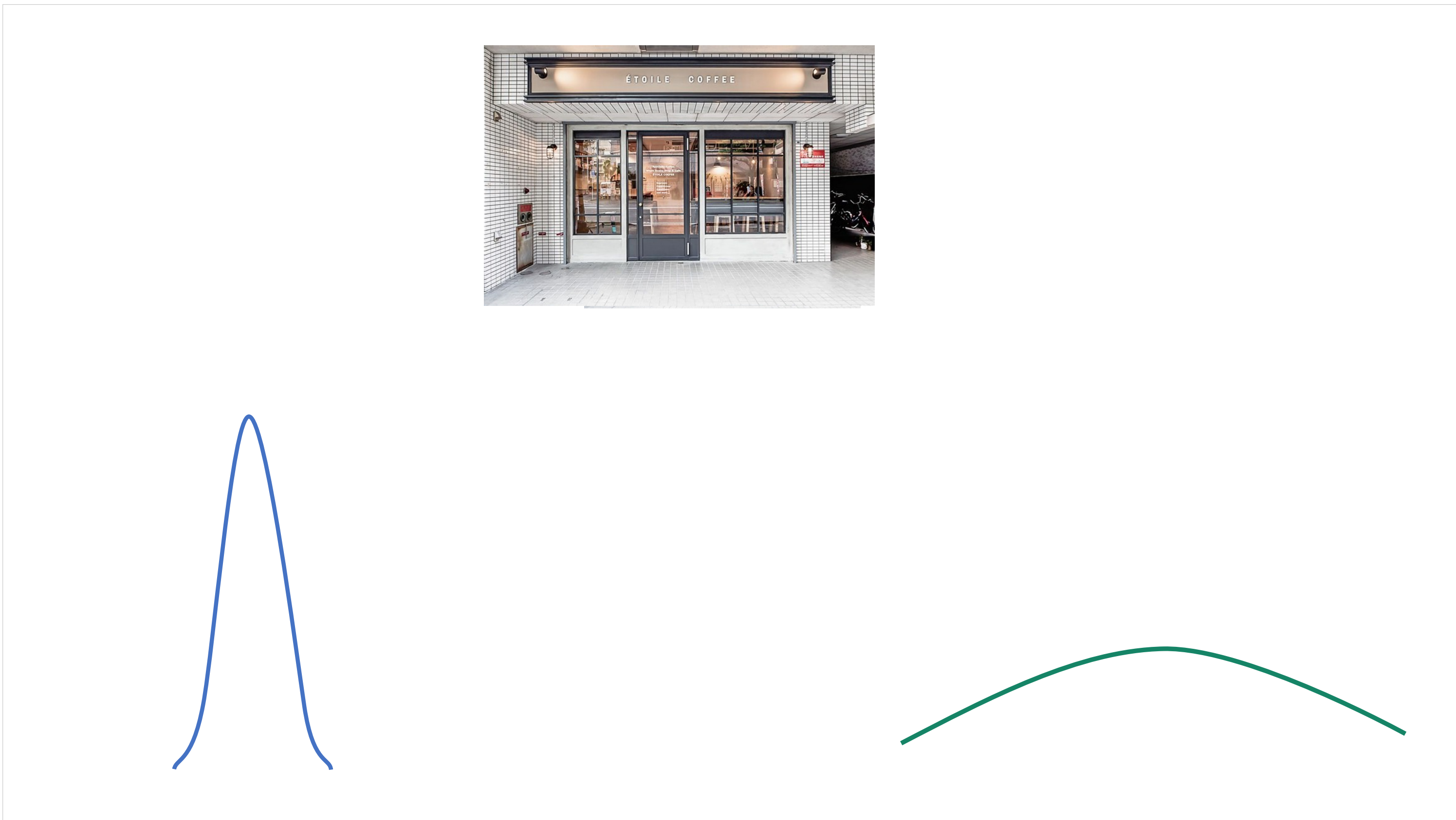
GOETHE UNIVERSITÄT FRANKFURT AM MAIN

LISCO

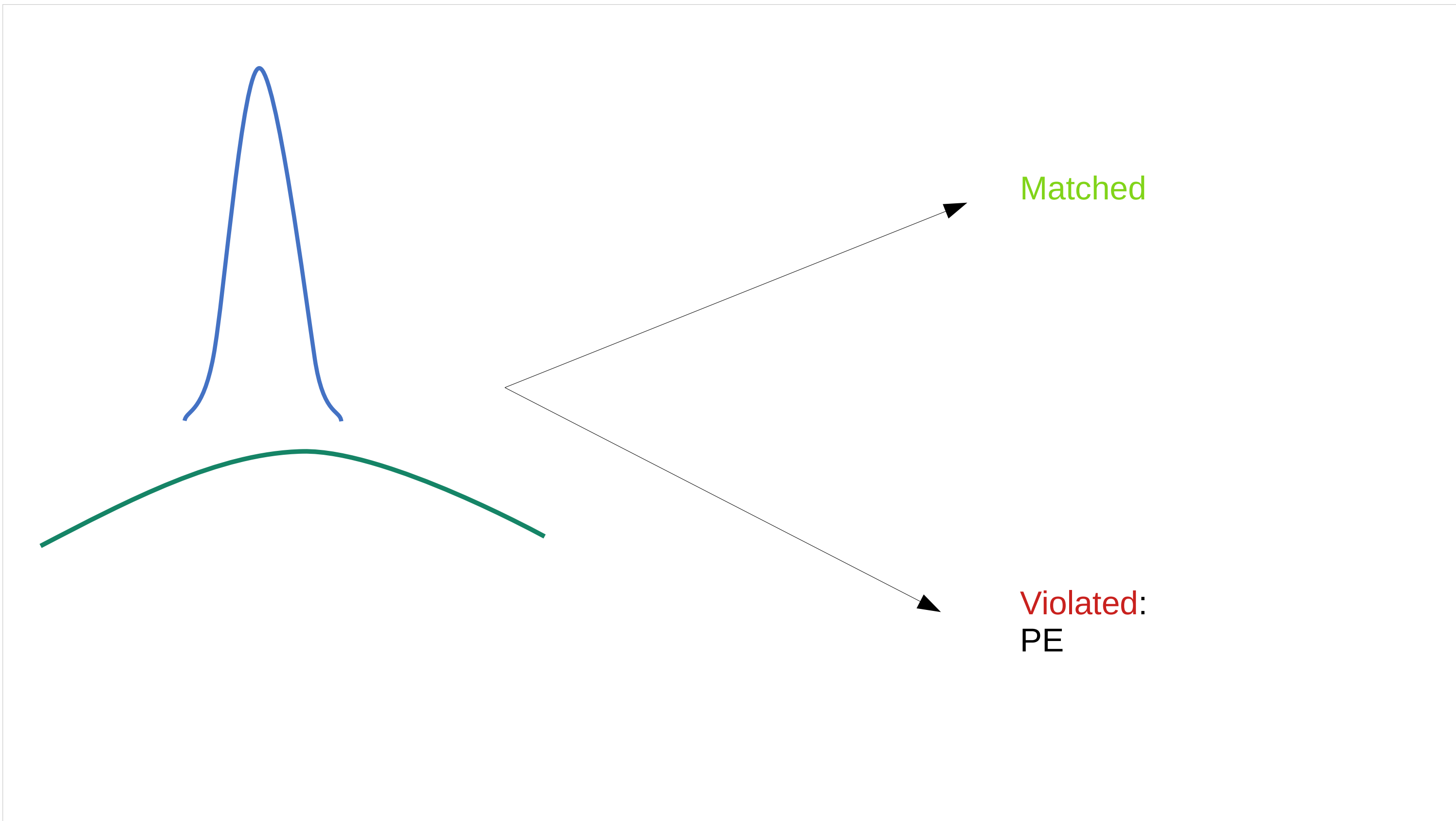
Size 40 header; size 20 body

The brain is a prediction machine, which tries to generate Predictions based on the extraction of regularities in the environment.

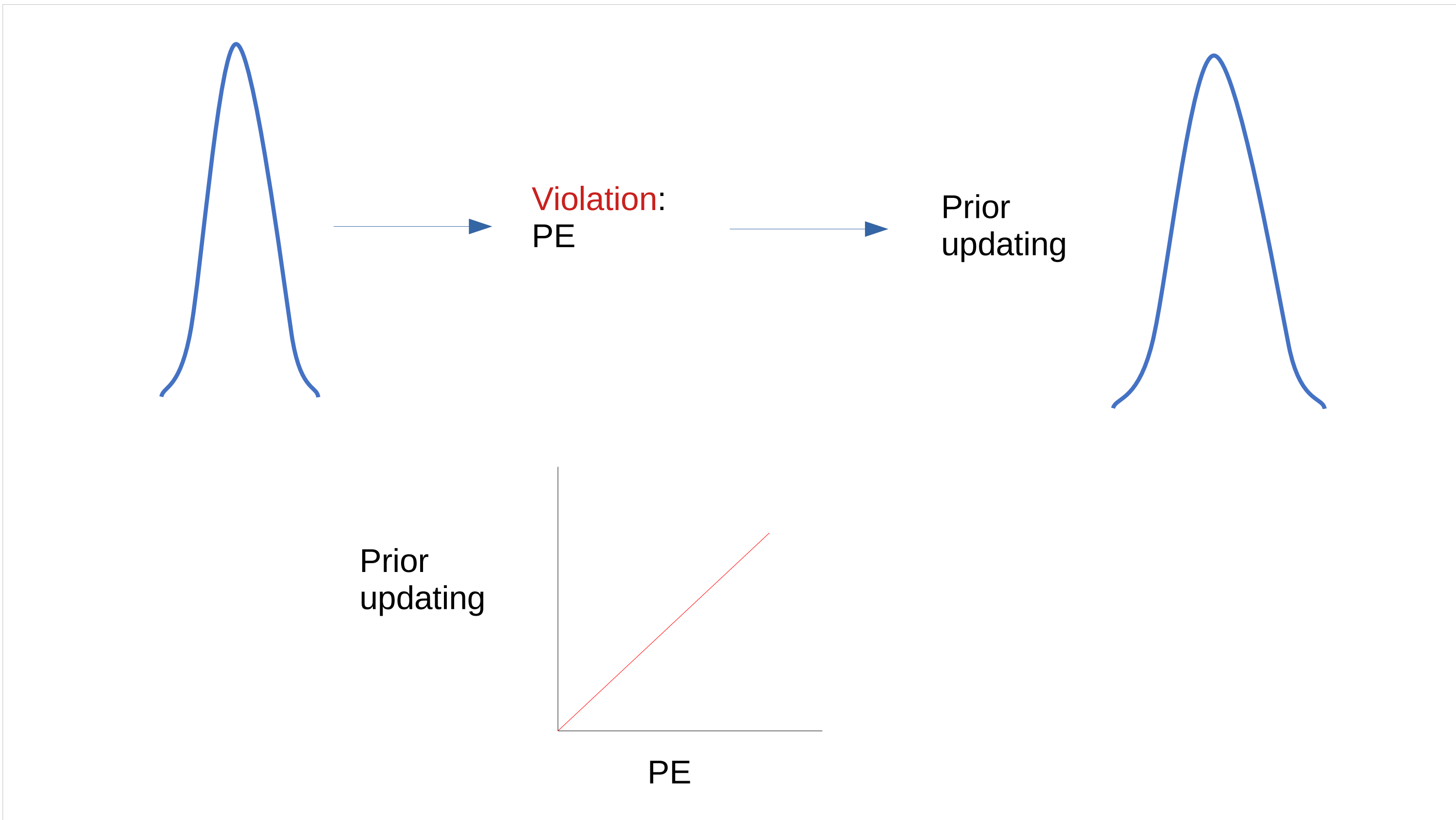
- Through repeated experiences we develop predictions of what will happen in a specific context.
- For example, we may go every day to the same cafe for our morning coffee. As a consequence, we develop a strong expectation of what is going to happen: the taste of the coffee, the furniture, the employees, etc.
- But prior expectations could also not be that strong: a cafe could be constantly restructuring and changing the staff, so that our predictions would be rather weak.



So we may have either a strong prior, with strong expectations of what it is going to happen in one context, and a weak prior, with weaker expectations.

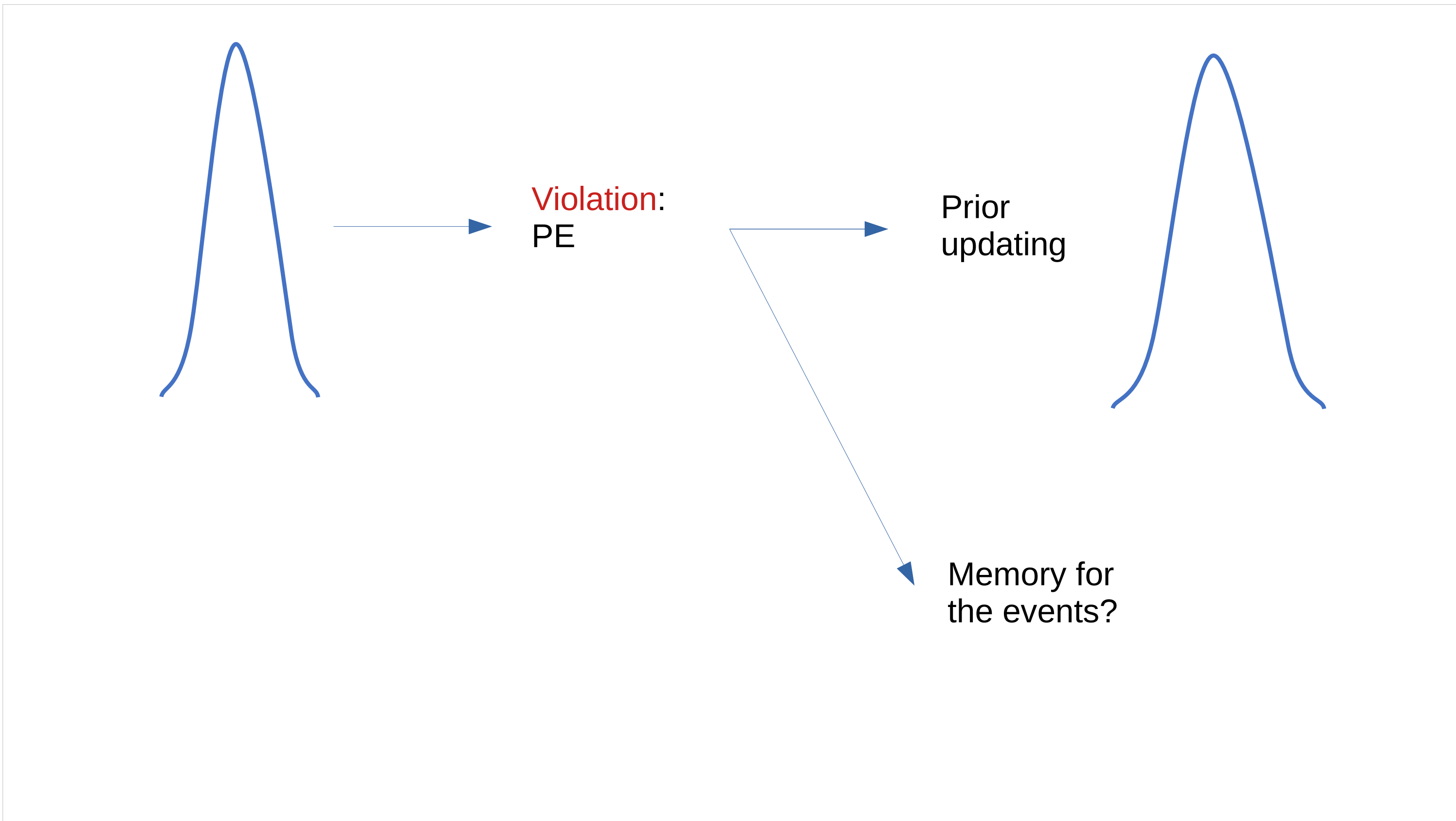


These expectations can be either matched violated, causing prediction errors of different degrees.



Prediction error is responsible for prior updating: changes in the expectations to accommodate the error.

There is a linear relationship between PE strength and prior updating.



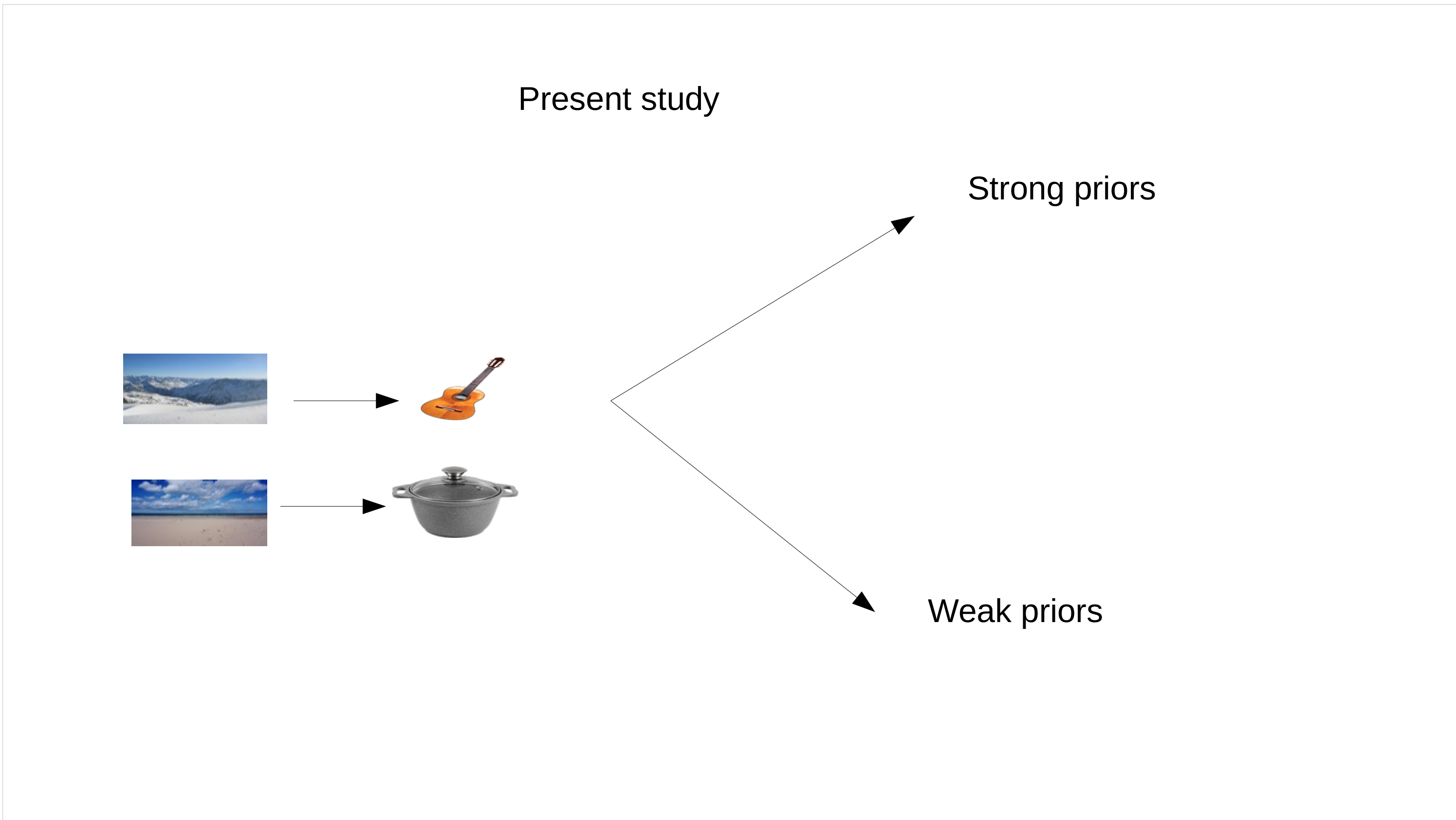
But what happens to the events that cause prediction error? Will they be better or worse remembered?

Contrasting evidence

Information is better recognized when is congruent with previous schema (e.g., Bein et al., 2015; Brod & Shing, 2019; Craik and Tulving 1975)

Information is better recognized when it is incongruent with previous schema (e.g., Greve et al., 2017; Kafkas & Montaldi, 2018)

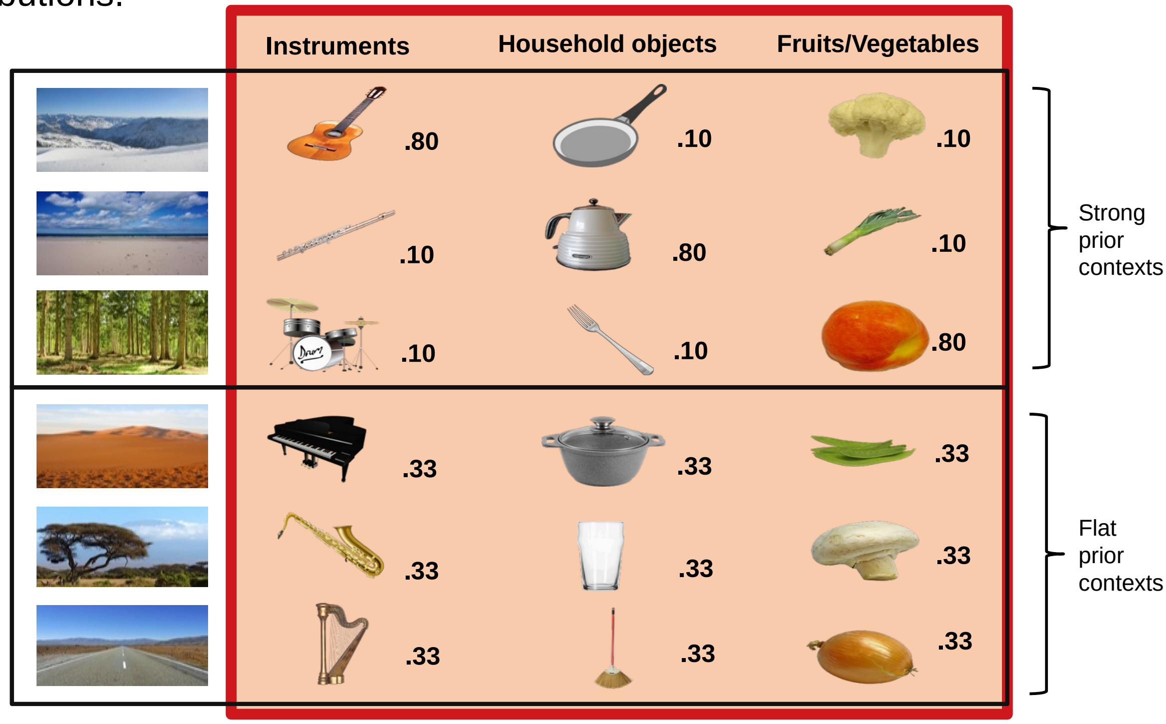
So far, 2:30 min ca.



The aim of the present study was to create different levels of priors that were later violated (PE).

That was achieved by inviting participants to learn associations between scenes and object categories. Different contingencies, to render different prior strength.

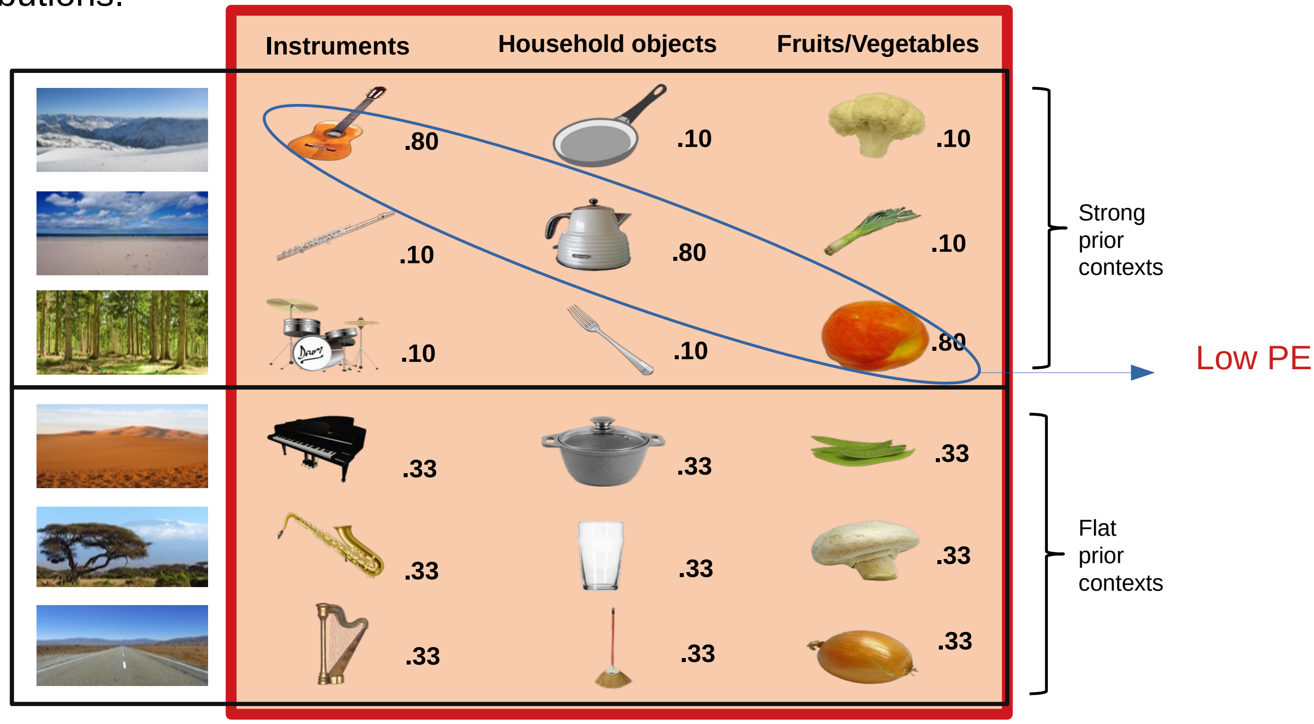
Three object **types** are presented embedded in each context following certain probability distributions.



There are three object categories, each category associated with a certain contingency with each of 6 scenes.

3 scenes belong to strong prior condition, 3 to weak.

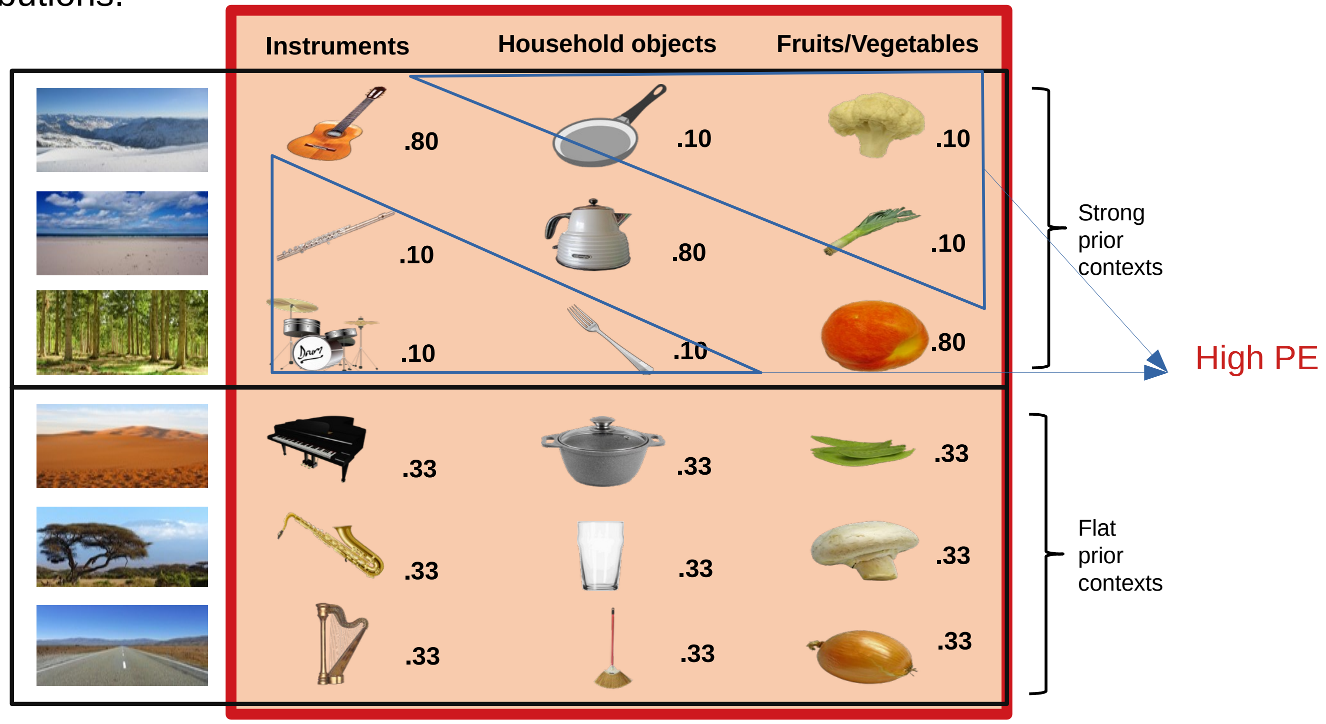
Three object **types** are presented embedded in each context following certain probability distributions.



The design creates three experimentally-manipulated levels of PE: low, high, and medium.

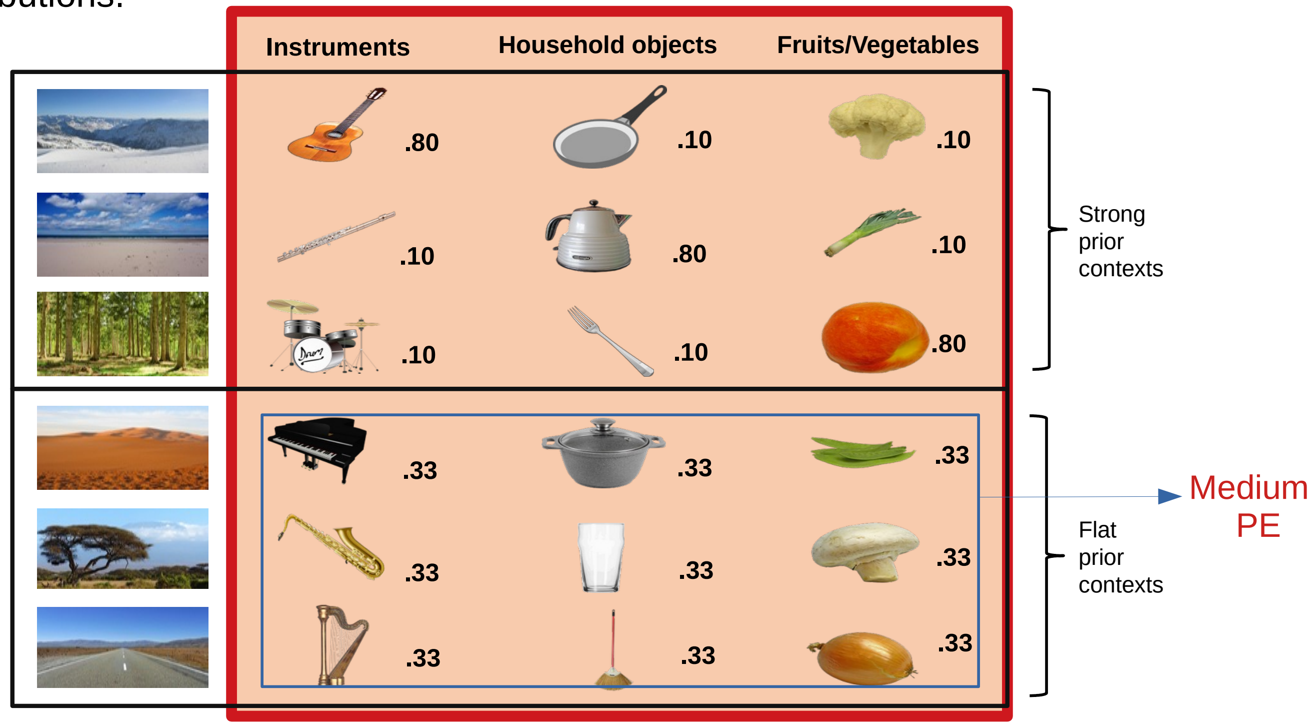
Low PE refers to object categories presented more frequently in the strong prior contexts;

Three object **types** are presented embedded in each context following certain probability distributions.

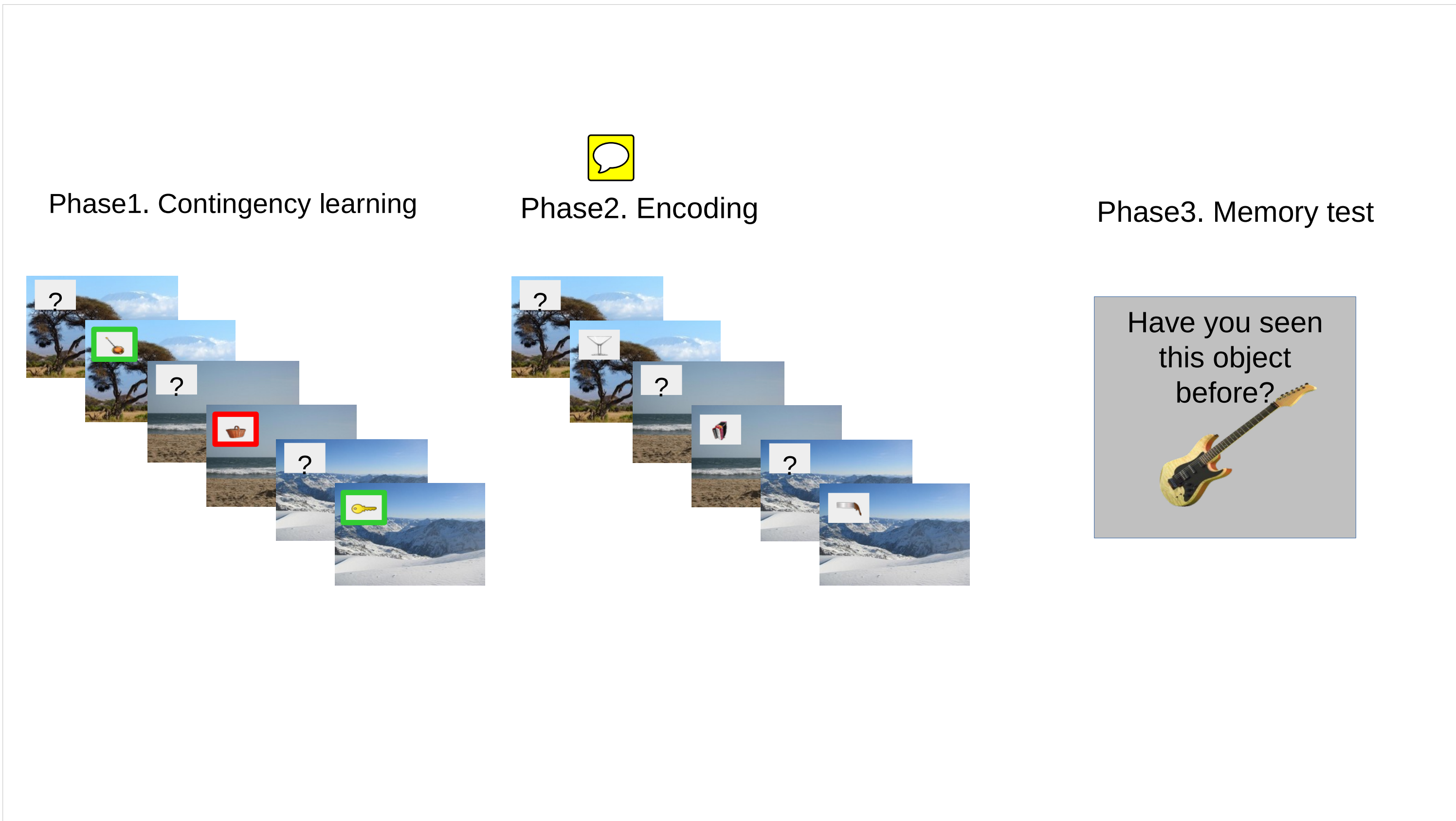


High PE to categories presented less frequently in strong prior contexts

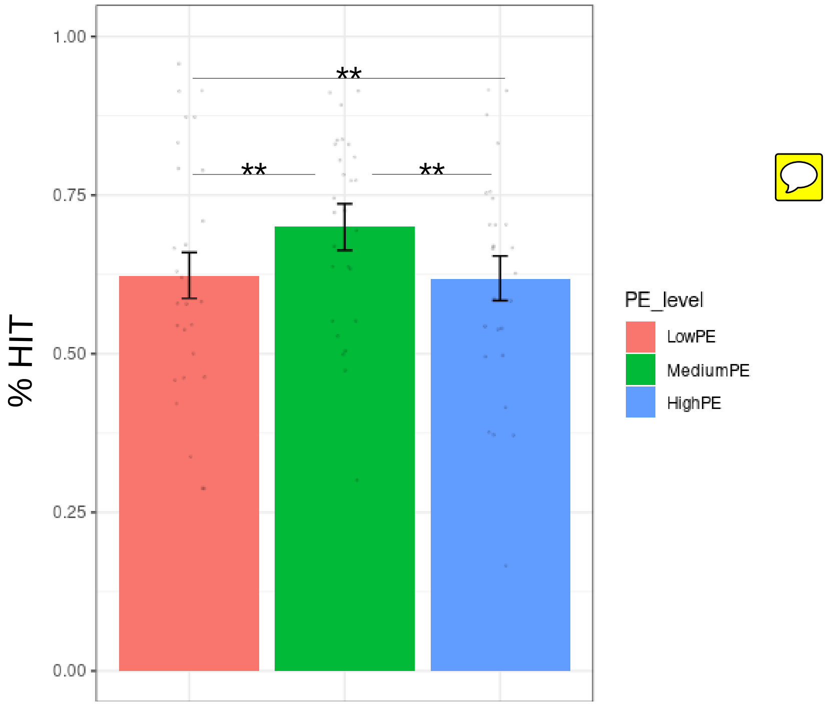
Three object **types** are presented embedded in each context following certain probability distributions.



Medium PE refers to object categories presented in the flat prior contexts



Briefly explain the paradigm.(5:30 – 6 ca so far).



Behavioural results: better memory for objects presented in the flat prior context (medium PE)

Reinforcement learning model

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

$Q^{c,j}$	Value of object category *J* for context *C*
α	Learning rate
δ_t	Prediction error

Simple RL model based on Rescorla-wagner rule.

Q is the expected strength of the association of a certain category with a context, given the history of the associations.

The learning rate regulates the extend to which evidence on a trial is used to update choice

Prediction error: the extent to which the evidence deviates from the expectations.

δ_t Prediction error

$$\delta_t = r_t - Q_t$$

$$r_t = 1\ ,\ 0$$

Observation - based - depending on the object category shown

Choice-based – depending on the feedback received (correct vs incorrect)

Prediction error can come from two sources:

The outcome of the choice or

The object category presented, irrespective of the choice.

Observation-based PE:
It depends on the object category displayed, regardless of participants' choice. It is >=0, inversely proportional to the expected value.

	Instruments	Household Objects	Fruits or Vegetables
$Q_t =$	0.66	0.00	0.00
$R_t =$	0	0	1
$\delta_t = r_t - Q_t$	0-0.66 = -0.66	0-0.00 = 0.00	1-0.00 = 1.00
$Q_{t+1} + \alpha \cdot \delta$	0.66+(0.3·(-0.66))=0.46	0.0+(0.3· 0.00)=0.0	0.00+(0.3·1)=0.30

Example of how the two ways of computing prediction error differ
In this case, observation-based is shown.

Choice-Based PE:
If participant predicted Instrument, but another object category, the PE for the category chosen is < = 0. The stronger the belief, the more negative it will be.

	Instruments	Household Objects	Fruits or Vegetables
Q_t =	0.66	0.00	0.00
R_t =	0	0	1
$\delta_t = r_t - Q_t$	0-0.66 = -0.66	0-0.00 = 0.00	1-0.00 = 1.00
$Q_{t+1} + \alpha \cdot \delta$	$0.66+(0.3\cdot(-0.66))=0.46$	$0.0+(0.3\cdot 0.00)=0.0$	$0.00+(0.3\cdot 1)=0.30$

Now choice based

Choice-Based PE:
If participant predicted Instrument, but another object category, the PE for the category chosen is < = 0. The stronger the belief, the more negative it will be.

Observation-based PE:
It depends on the object category displayed, regardless of participants' choice. It is >=0, inversely proportional to the expected value.

	Instruments	Household Objects	Fruits or Vegetables
$Q_t =$	0.33	0.00	0.00
$R_t =$	1	0	0
$\delta_t = r_t - Q_t$	1-0.33 = 0.66	0-0.00 = 0.00	0-0.00 = 0.00
$Q_{t+1} + \alpha \cdot \delta$	0.33+(0.3·(0.66))=0.53	0.0+(0.3· 0.00)=0.0	0.00+(0.3·0)=0.00

When Participants' prediction is correct, the two PEs are the same.

In this example, the participant predicted musical instrument, and a musical instrument was presented.

$$Q_{t+1}^{(c,j)} = Q_t^{(c,j)} + \alpha \cdot (rOBS - Q_t) \qquad Q_{t+1}^{(c,j)} = Q_t^{(c,j)} + \alpha \cdot (rCB - Q_t)$$

A softmax function was then used to transform the values into probabilities:

$$P_j(c = I | Q_t^c) = \frac{\exp(\beta \cdot Q_t^c)}{\sum_{i=1}^{N_c} \exp(\beta \cdot Q_t^c)}$$

We fitted the two parallel models to participants' data, one with choice-based and one with observation-based PE, to derive the two different PEs.

Softmax distribution was the same for both

$$Q_{t+1}^{(c,j)} = Q_t^{(c,j)} + \alpha \cdot (rOBS - Q_t) \qquad Q_{t+1}^{(c,j)} = Q_t^{(c,j)} + \alpha \cdot (rCB - Q_t)$$

A softmax function was then used to transform the values into probabilities:

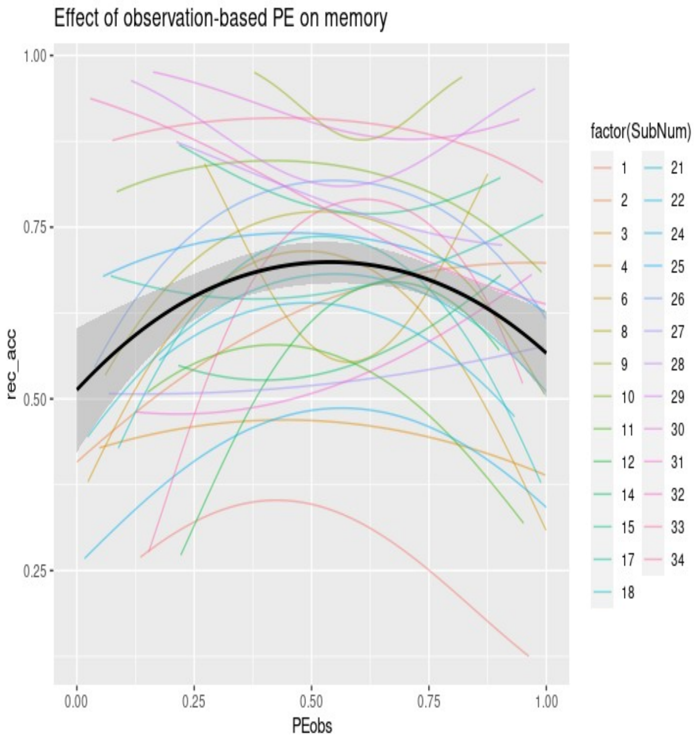
$$P_j(c = I | Q_t^c) = \frac{\exp(\beta \cdot Q_t^c)}{\sum_{i=1}^{N_c} \exp(\beta \cdot Q_t^c)}$$

Parameter estimation:
Alpha and beta values were estimated at the participant level through ML estimation

$$p(d_{1:T} | \theta_m, m) \qquad LL = logp(d_{1:T} | \theta_m, m)$$

Parameter estimation – maybe it is good to leave this out if there is not enough time.

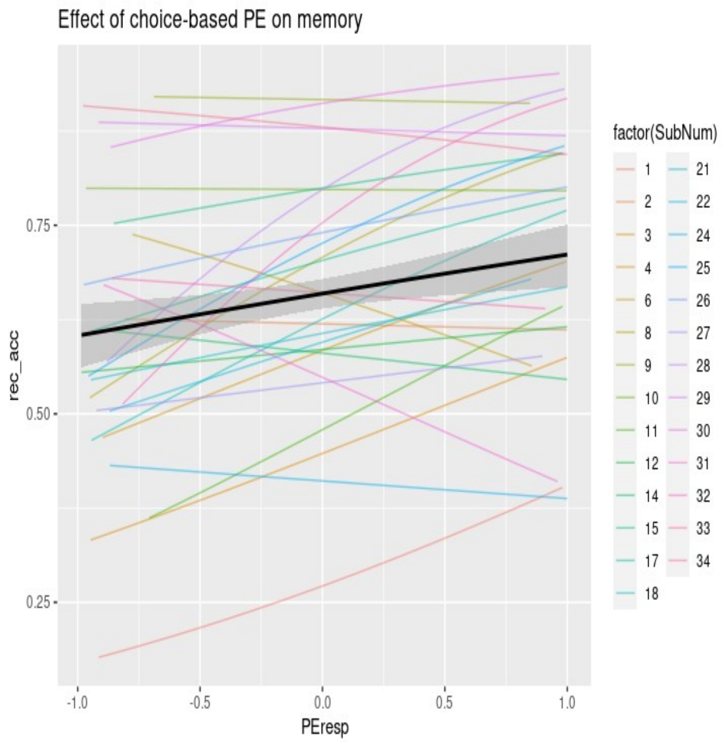
Quadratic effect: β = -6.43, p = 0.006



27 participants: 6 excluded for recognition performance below chance level.

Significant quadratic effect. This effect recapitulates what observed with PE conditions.

Linear effect: β = 0.28, *p* = 0.003



This positive linear relationship between choice-based PE and memory shows that stronger prediction error (as an outcome of weak prior) when participants correctly predicted the category led to enhanced memory. In contrast, when participants incorrectly predicted the object category, stronger prediction error (as an outcome of strong prior) led to impaired memory.

Conclusions

- Important moderating role of prediction accuracy and of considering PE in relation to task participants are doing.
- Similarity between signed PE effects on memory and signed PE effects on prior updating: dopaminergic activity?
- Strong prediction error in the context of positive feedback might indicate the utility of that item for future predictions.
- Strong prediction error in the context of negative feedback in a task where contingencies are stable may informs participants that that choice is not informative.

There is s similarity between effect of signed PE on memory and signed PE effects on prior updating. This might suggest that the memory effects observed rely on phasic activity of dopamine neurons in the midbrain and striatum, which project to hippocampus.

- About the utility, Items that were successfully predicted when priors where low might be prioritised because they can be helpful to predict in the future (e.g., Jang et al., 2019, Rohuani et al., 2021).
- In the case of negative feedback, the stronger the participant' prior at some point, the more is the likelihood that that item will be not important, and will not be useful for future choice·

Thank you!

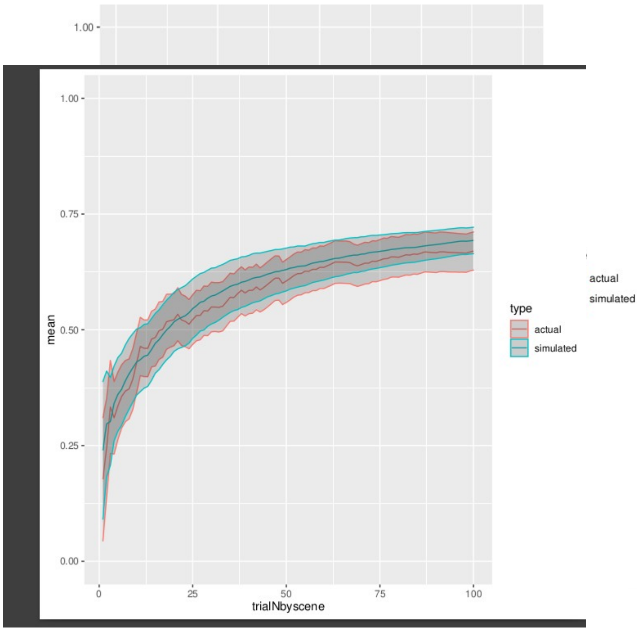
GOETHE UNIVERSITÄT FRANKFURT AM MAIN

erc European Research Council

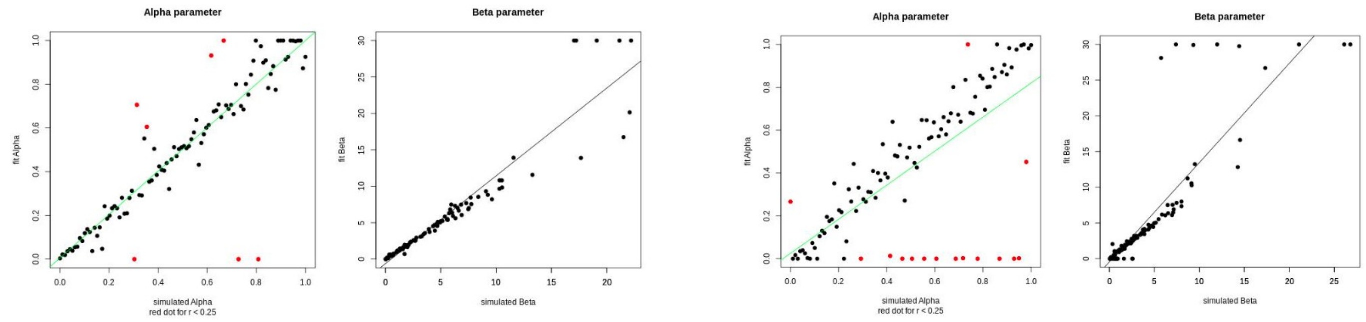
LISCO

There is similarity between effect of signed PE on memory and signed PE effects on prior updating. This might suggest that the memory effects rely on phasic activity of dopamine neurons in the midbrain and striatum, which project to hippocampus.

- About the utility, Items that were successfully predicted when priors where low might be prioritised because they can be helpful to predict in the future (e.g., Jang et al., 2019, Rohuani et al., 2021). Therefore, the choice might be important for future similar context
- In the case of negative feedback, the stronger the participant' prior at some point, the more is the likelihood that that item will be not important, and will not be useful for future choice.



Validation: cumulative accuracy for simulated vs empirical



Parameter recovery was successful for both models. Beta parameter was constrained between 0 and 10.