PREMUP: Reinforcement learning models

$$Q_{t+1} = Q_t + \alpha \cdot \delta_t$$

The value of each category with reference to a scene is updated to the current value plus the δt, the prediction error, multiplied by the learning rate α

α = the extent to which reinforcement
on each trial is used to update choice

$$\delta_t = r_t - Q_t$$

r^o *:* 1 = shown; 0 = not shown

PREMUP: Reinforcement learning models

$Q_t(c_t, j_t)$ thus refers to how strongly a category *c* is associated with a scene *j* at trial *t.*

$$Q_{t+1}(I_t, M_t) = Q_t(I_t, M_t) + \alpha \cdot \delta_t$$

$$\delta_t = r_t - Q_t \qquad r_t = 1 , 0$$

α = the extent to which reinforcement on each trial is used to update choice

Reinforcement learning models: Observational

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

Reinforcement learning models: Feedback-based

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

Reinforcement learning models: observational

	Instruments	Household Objects	Fruits or Vegetables
Q_t =	0.95	0.20	0.20
R_t =	1	0	0
$\delta_t = r_t - Q_t$	1-0.66 = 0.34	0-0.00 = 0.00	0-0.00 = 0.00
$Q_{t+1} + \alpha \cdot \delta$	0.66+(0.3·0.34)=0.7	0.0+(0.3· 0.00)=0.00	0.00+(0.3·0.00)=0.00

Reinforcement learning models: feedback-based

	Instruments	Household Objects	Fruits or Vegetables
Q_t =	0.66	0.00	0.00
R_t =	1	0	0
$\delta_t = r_t - Q_t$	1-0.66 = 0.34	0-0.00 = 0.00	0-0.00 = 0.00
$Q_{t+1} + \alpha \cdot \delta$	0.66+(0.3·0.34)=0.7	0.0+(0.3· 0.00)=0.00	0.00+(0.3·0.00)=0.00

Reinforcement learning models: observational

	Instruments	Household Objects	Fruits or Vegetables
Q_t =	0.66	0.30	0.20
R_t =	0	0	1
$\delta_t = r_t - Q_t$	0-0.66 = -0.66	0-0.30 = -0.30	1-0.20 = 0.80
$Q_{t+1} + \alpha \cdot \delta$	$0.66+(0.3\cdot(-0.66))=0.46$	$0.3+(0.3\cdot(-0.30)=0.21$	$0.2+(0.2\cdot0.8)=0.36$

Reinforcement learning models: feedback-based

	Instruments	Household Objects	Fruits or Vegetables
Q_t =	0.66	0.30	0.20
R_t =	0	0	0
$\delta_t = r_t - Q_t$	0-0.66 = -0.66	0-0.30 = -0.30	0-0.20 = -0.20
$Q_{t+1} + \alpha \cdot \delta$	$0.66+(0.3\cdot(-0.66))=0.46$	$0.3+(0.3\cdot(-0.30)=0.21$	$0.2+0.3\cdot(-0.2)=0.14$

Reinforcement learning models: observational

Choice-Based PE:
If participant predicted Instrument, but another object category is shown, 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	0.30	0.20
R_t =	0	0	1
$\delta_t = r_t - Q_t$	0-0.66 = -0.66	0-0.30 = -0.30	1-0.20 = 0.80
$Q_{t+1} + \alpha \cdot \delta$	0.66+(0.3·(-0.66))=0.46	0.3+(0.3· (-0.30)=0.21	0.2+(0.2·0.8)=0.36

Reinforcement learning models: observational

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	0.30	0.20
R_t =	0	0	1
$\delta_t = r_t - Q_t$	0-0.66 = -0.66	0-0.30 = -0.30	1-0.20 = 0.80
$Q_{t+1} + \alpha \cdot \delta$	$0.66+(0.3\cdot(-0.66))=0.46$	$0.3+(0.3\cdot(-0.30)=0.21$	$0.2+(0.2\cdot0.8)=0.36$

Reinforcement learning models: observational

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	0.30	0.20
R_t =	0	0	1
$\delta_t = r_t - Q_t$	0-0.66 = -0.66	0-0.30 = -0.30	1-0.20 = 0.80
$Q_{t+1} + \alpha \cdot \delta$	$0.66+(0.3\cdot(-0.66))=0.46$	$0.3+(0.3\cdot(-0.30)=0.21$	$0.2+(0.2\cdot0.8)=0.36$

Reinforcement learning models: feedback-based

Choice-Based PE:
Same as for the observational model.

	Instruments	Household Objects	Fruits or Vegetables
Q_t =	0.66	0.00	0.20
R_t =	0	0	0
$\delta_t = r_t - Q_t$	0-0.66 = -0.66	0-0.00 = 0.00	0-0.20 = -0.20
$Q_{t+1} + \alpha \cdot \delta$	0.66+(0.3·(-0.66))=0.46	0.0+(0.3· 0.00)=0.0	0.2+0.3·(-0.2)=0.14

Reinforcement learning models: feedback-based

Observation-based PE:
It depends on the object category displayed, In case of incorrect choice, it is always negative,and it is inversely proportional to the expected value.

	Instruments	Household Objects	Fruits or Vegetables
Q_t =	0.66	0.00	0.20
R_t =	0	0	0
$\delta_t = r_t - Q_t$	0-0.66 = -0.66	0-0.00 = 0.00	0-0.20 = -0.20
$Q_{t+1} + \alpha \cdot \delta$	0.66+(0.3·(-0.66))=0.46	0.0+(0.3· 0.00)=0.0	0.2+0.3·(-0.2)=0.14

Reinforcement learning models: feedback-based

Observation-based PE:
It depends on the object category displayed, In case of incorrect choice, it is always negative,and it is inversely proportional to the expected value.

	Instruments	Household Objects	Fruits or Vegetables
Q_t =	0.20	0.00	0.80
R_t =	0	0	0
$\delta_t = r_t - Q_t$	0-0.20 = -0.20	0-0.00 = 0.00	0-0.80 = -0.80
$Q_{t+1} + \alpha \cdot \delta$	0.2+0.3·(-0.2)=0.14	0.0+(0.3· 0.00)=0.0	0.8+0.3·(-0.8)=0.56

PREMUP: Reinforcement learning models

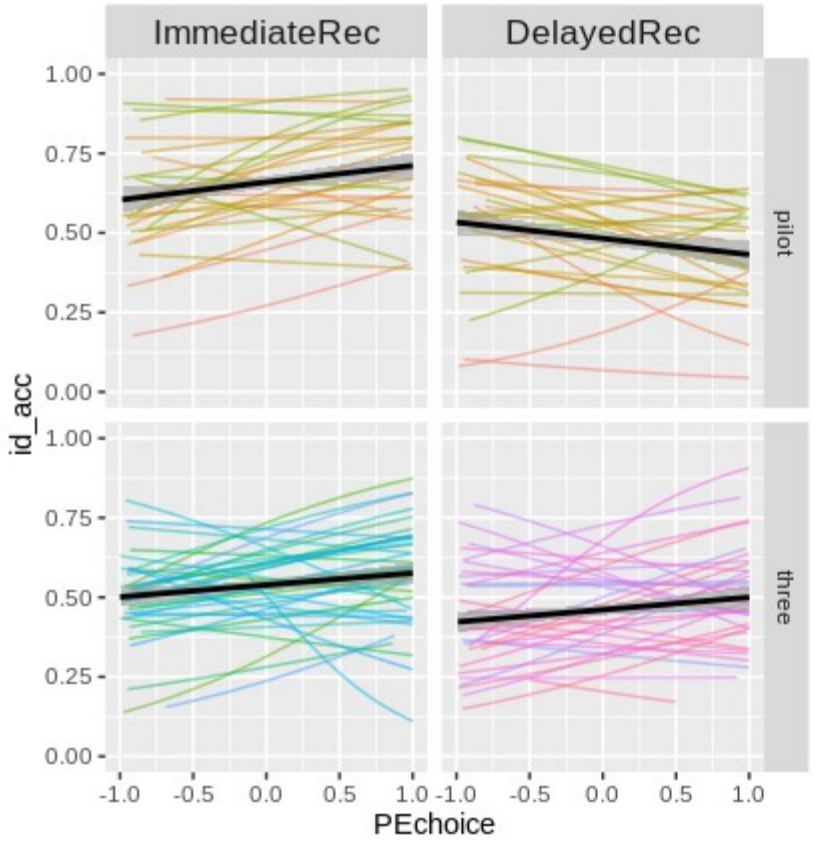
- Learning rates that best accounted for behavioural data were estimated for each participant through maximum likelihood estimation, which consists in finding the values that maximize the likelihood of choice probability, given the model and the parameter.

- Models with the best fitting learning rate for each participant were fitted to the data to derive trial-level PE.

PREMUP: Reinforcement learning models

Choice PE model

Choice based-PE effect on memory



Observational PE model

Observational-PE effect on memory