Classical Rescorla-Wagner models on reversal learning can be enhanced by inclusion of a metacognitive index

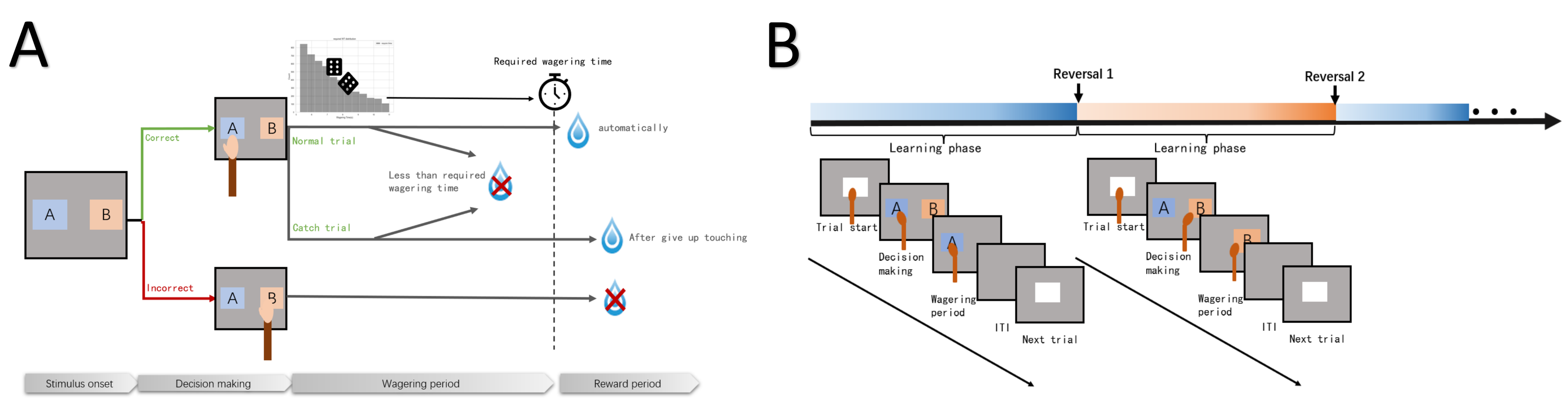
Zhiyong JIN[1, 3, #], Chenxi ZHAI[1, 3, #], Yudian CAI[2, 3], Xuanlong ZHU[1, 3], Alireza SOLTANI[4], Sze Chai KWOK[1, 3, *]

[1] Duke Kunshan University, Kunshan, China.
[2] East China University of Science and Technology, Shanghai, China.
[3] East China Normal University, Shanghai, China.
[4] Dartmouth College, Hanover, United States.

Introduction

Learning is a complex process. Decades of learning research have looked into myriad factors that affect learning including effects of rewards, intrinsic motivation, probabilistic structure of the external environment and so on. However, the role of the agent's subjective state, such as metacognitive markers like confidence, has not been systemically considered. We aim to elucidate the influence of confidence on learning processes and provide a quantification of the interactive relationship between the two.

Method & Data Analysis



Task	Trial structure	Reverse condition
1	5 catch trials after reversal, others are normal trials	Reverse every 30 trials
2	5 catch trials after reversal, others are normal trials	Reverse after reach a random required accuracy in last 20 trials
3	All trials are catch trials	Reverse after reach a random required accuracy in last 20 trials

Task	Animal name	Number of session	Number of reversal	Number of trial
1	Saturn	24sessions	81	3149
1	Mars	24sessions	71	2785
2	Uranus	21sessions	97	2990
2	Neptune	21sessions	86	2545
3	Uranus	26sessions	121	4902
3	Neptune	27sessions	131	4161

Fig. 1 Task paradigm of the reversal learning task in macaque monkeys. A: Illustration of a single trial, monkeys need to put their hands on the touch screen and hold until they reach a certain duration. (exponential distribution) B: Reverse routine of one block, three tasks are divided by trial structure and reverse condition, offering different environment uncertainties.

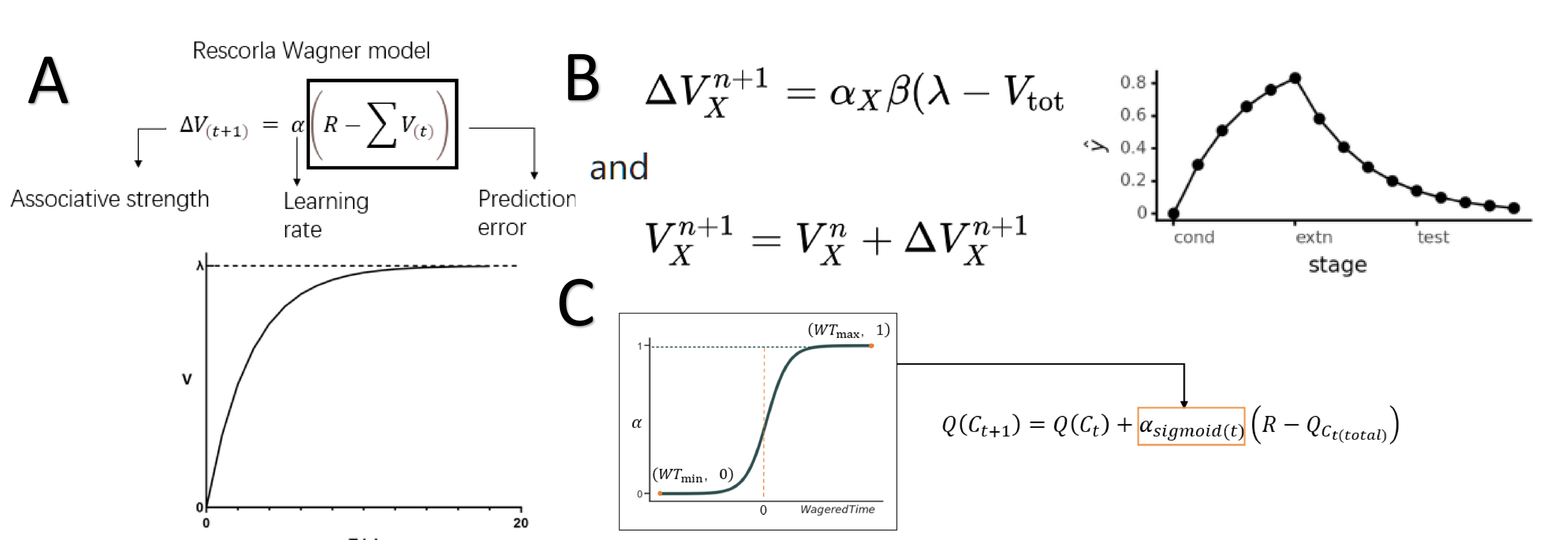
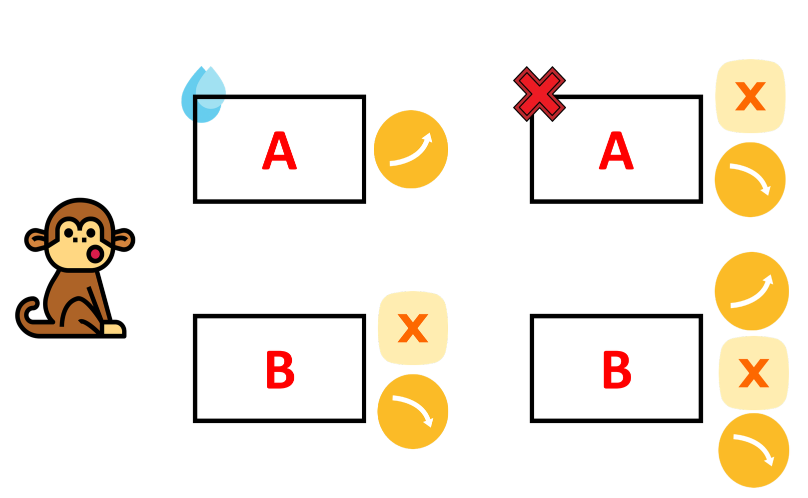


Fig. 2 Mathematical representation of the Rescorla-Wagner Model (RW model) and the curve illustrating the transformation of confidence level to the learning rate. A. Schematic of the RW model, where the curve represents the changing association strength between stimuli and rewards across trials. B. Mathematical expression of the RW model (left panel) and a graph depicting the process of association strengthening and weakening (right panel), with both V and y representing the association strength. C. The transformation function between WT (representing confidence) and alpha (representing learning rate), where we employ a sigmoid function to convert the monkey's confidence from the previous trial into the learning rate for the current trial, demonstrating the impact of confidence on the learning process as depicted by the RW model through changes in LL, AIC, and BIC.



Outcome	VA	VB
Reward + A	Increase	No changes
		Decrease
	Decrease	Increase
No-reward +A		No changes
	Increase	Decrease

Fig. 3 Potential forms of the RW model. We contemplate constructing multiple RW models, thus necessitating consideration of all possible directions of change between rewards and stimuli. Taking stimulus A as an example, assuming monkeys receive a reward or no reward upon choosing A, we can construct a total of 12 models based on the potential changes in VA and VB (increase, no change, or decrease). Furthermore, based on the number of alphas, we have established a total of 30 basic RW models.

Results

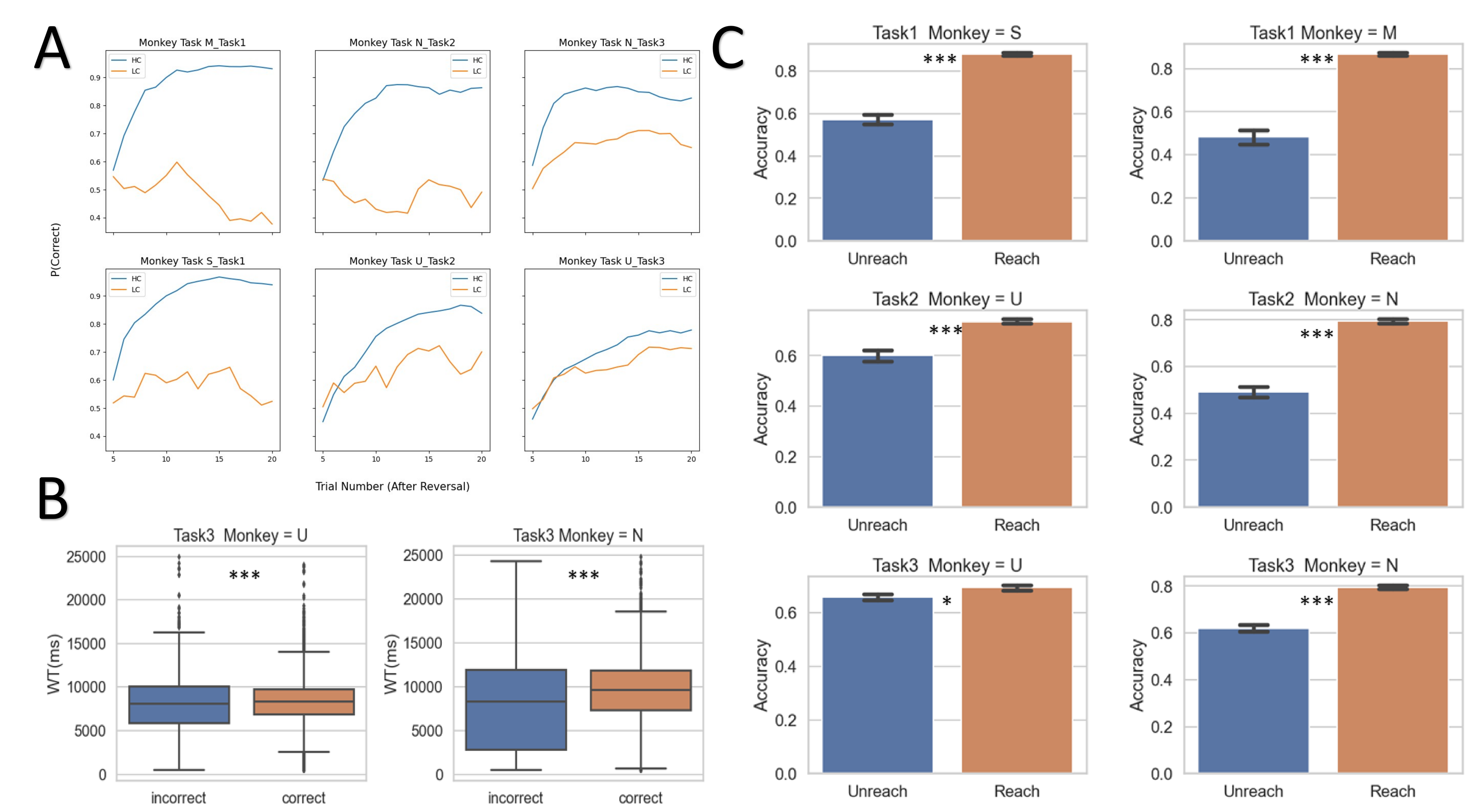


Fig. 4 The time wagering confidence paradigm effectively reflects the confidence of macaques in the current trial during a reversal learning task. A. After the reversal, the probability of receiving a reward at different levels of confidence (represented by wagered time, WT) diverges over the course of trials, with higher confidence trials showing improved accuracy compared to lower confidence trials across different tasks and monkeys. B. The distinction in WT between correct and incorrect trials, with the mean WT of correct trials being significantly greater than that of incorrect trials. C. Trials reaching the required time (high confidence) consistently exhibit higher accuracy than those not reaching the required time (low confidence). (***: P<0.001; *: P<0.05)

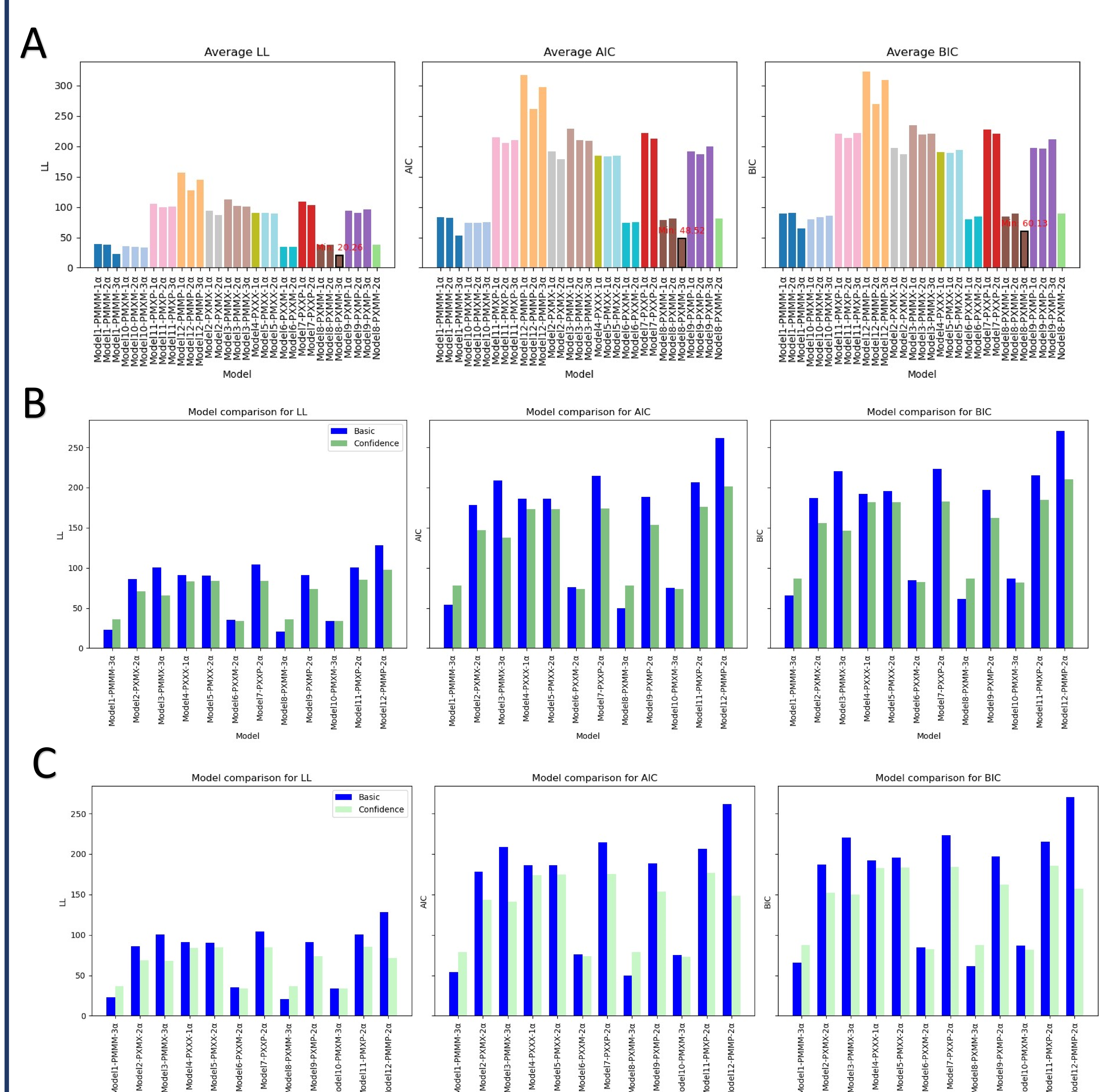


Fig. 5 Comparison of RW model performance. A. Among the 30 basic models constructed based on the relationship between V and reward, as well as the type of alpha, we compare their performance using Negative Loglikelihood (LL), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) as metrics to assess model fit, with lower values indicating better fit. From these models, we select the best-performing model from each as the winning model. B. We transform the previous trial's WT (representing confidence) using a sigmoid function to derive alpha as the learning rate for the current trial, and the fitting results suggest that the confidence-learning rate mapping enhances model performance, where the basic model refers to the optimal fitted model selected in the previous process. C. We further transformed the WT from the last two trials, and the results still favour the model with confidence-learning rate mapping over the basic model.

Conclusion

- We evaluated the feasibility of the time wagering confidence paradigm in macaques' reversal learning task, and the results revealed the paradigm's viability and broad applicability.
- We establish a behavioural model which captures how trial-wise confidence modulates macaques' learning efficiency, suggesting the inclusion of metacognitive attributes to existing learning models would be beneficial.

References & Funders

- Trepka et al. Nature Communications 2021; Lak et al. Neuron. 2014

- This research is sponsored by the Kunshan Municipal Government research funding Project Number: 23KKSGR017 and 24KKSGR017 (to SCK).

- Correspondence:sk695@duke.edu (SCK); alireza.soltani@dartmouth.edu (AS) zj72@duke.edu (JZY); cz194@duke.edu (ZCX)