Why some simple probabilistic rules are difficult to learn: A hypothesis diffusion model

Jianbo Chen (jianbo@pku.edu.cn)
Academy for Advanced Interdisciplinary Studies and Center for Life Sciences, Peking University
Beijing 100871, China

Tianyuan Teng (tianyuan.teng@tuebingen.mpg.de)
Max Planck Institute for Biological Cybernetics
Tübingen 72070, Germany

Yuxiu Shao (shaoyx@bnu.edu.cn)
School of System Science, Beijing Normal University
Beijing 100875, China

Hang Zhang (hang.zhang@pku.edu.cn)
School of Psychological and Cognitive Sciences, Peking University
Beijing 100871, China

Abstract

People are known for their ability to learn probabilistic rewarding rules of the environment in both laboratory and real-world settings. However, in a series of experiments, we found that human participants failed to learn simple non-linear combinatory rules (e.g., the XOR rule of 11→1, 00→1, 01→0, 10→0 with a probability of 0.8), even after 320 trials, despite the small feature space for possible rules (i.e., including only two binary dimensions). This contrasted with the rapid learning of repetition or alternation rules in the same experiments. To explain why probabilistic XOR rules are difficult to learn, we propose a computational model that views rule learning as a progressively evolving hypothesis testing process. This hypothesis diffusion model assumes that (1) the weights assigned to different hypotheses diffuse across a network connecting hypotheses that can be transformed into each other through a single operation, and (2) the diffusion process is evidence-driven. The model successfully reproduces the behavioral patterns observed under each rule condition. Moreover, the model parameters estimated from a single rule condition can predict the observed differences in learning performance across different conditions.

Keywords: probabilistic learning; hypothesis testing; hidden Markov network; representation learning

Introduction

Humans can learn reward-predictive features in stochastic environments with multiple linearly-additive dimensions (Farashahi et al., 2017; Niv et al., 2015; Song et al., 2022) and can also learn deterministic reward rules that non-linearly combine multiple dimensions (Cohen et al., 2021; Cohen & Schneidman, 2013). However, learning is poor when the rules are both stochastic and non-linearly combinatory (Wang & Soltani, 2023). In this study, we conducted three experiments to test participants' learning capacity on stochastic reward rules in a 2-by-2 small feature space, finding striking failures in non-linear combinatory rules that cannot be attributed to the combinatorial explosion of feature space. Nor can it be explained by previous reinforcement learning (RL) models, unless distinctively different initial biases were assumed for different rule conditions. Inspired by the hidden Markov model of strategy switching (Ashwood et al., 2022), we proposed a hypothesis diffusion model based on evidence-driven stochastic flows of rule weights in a sparsely connected network of hypotheses, which provides a unified explanation for human performance in all conditions.

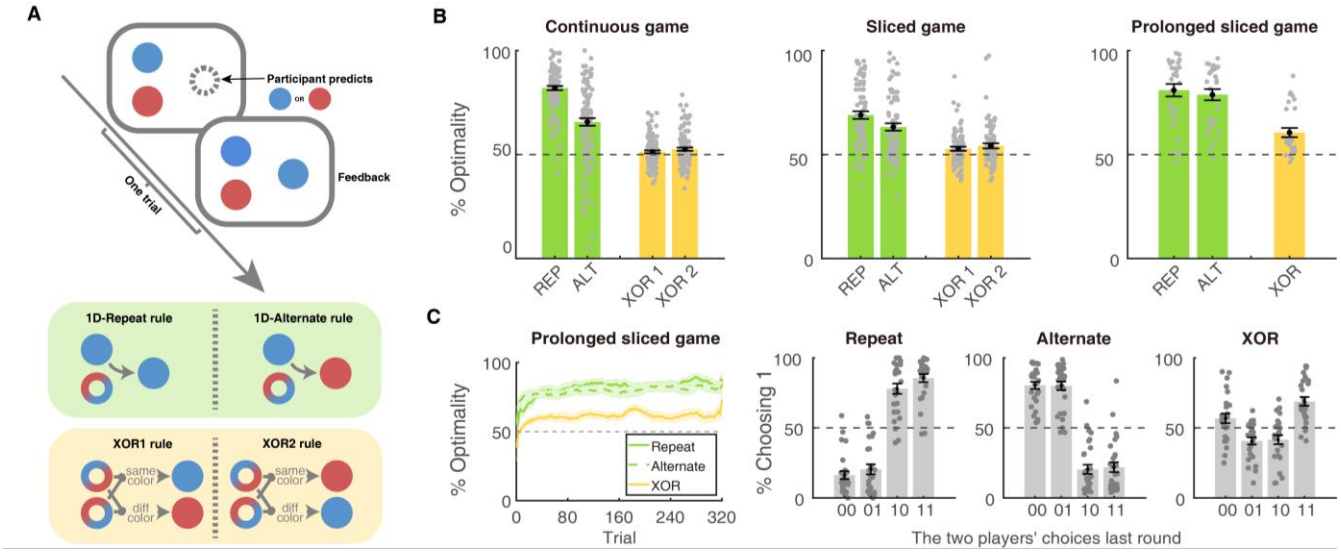

Figure 1: Experimental setting and behavioral results. **A**, Procedure and the decision rules. **B**, Performance in each experiment. **C**, Detailed patterns in the prolonged sliced game. Left: Learning curves showing early performance plateaus. Right: Participants' predictions following the 4 different states (e.g., 01: tail (0) and head (1) observed).

Behavioral experiments and results

In a 320-trial adapted matching-pennies game, participants predicted a target player's ("alien") round 2 choice based on the two players' round 1 choices. After each prediction, the actual round 2 choices were shown as feedback (Fig. 1A). The alien followed one of three probabilistic rules (p=0.8): "**Repeat**" (same choice as round 1), "**Alternate**" (switch choice), or "**XOR**" (head if same choice in round 1, tail otherwise, or vice versa). Exp. 1 ("**continuous game**"): Participants played as the alien's opponent; previous trial's round 2 became current trial's round 1. Exp. 2 ("**sliced game**"): Participants watched independent game slices. In both, participants completed four 80-trial blocks of repeat,

alternate, and two XOR rules. Exp. 3 ("**prolonged sliced game**"): Each participant completed one 320-trial block of a single rule. Exp. 1–3 had respectively 110, 76, and 86 valid participants from Prolific. Due to space constraints, only Exp. 3 statistics are reported.

Failure in learning probabilistic XOR rules Learning probabilistic XOR rules was challenging (Fig. 1B), as indicated by the proportion of optimality (choosing the alien's likely choice). Performance varied across three conditions ($F(2, 83) = 17.45$, $p < .001$, $\eta_p^2 = .30$), with the XOR performance near chance and significantly lower than the Repeat and Alternate conditions (post-hoc tests, ps < .001 with Bonferroni correction), suggesting difficulty learning the XOR rule.

Preference for "all matching" under XOR rule In all conditions, participants were more likely to choose predictions that agreed with the condition's rule (GLMM, ps <. 001). However, in the XOR condition, when the alien and human made the same choice in round 1, participants also tended to repeat the choice (i.e., 11→1 or 00→0, ps <. 05), leading to behavioral asymmetry in addition to following the XOR rule (Fig. 1C).

Computational modeling

The hypothesis diffusion (HD) model (Fig. 2A) assumes that hypotheses transformable into each other through a single operation are connected. The agent assigns weights to different hypotheses and makes decisions by weighing their predictions. After receiving round-2 feedback, each hypothesis' weight is redistributed among itself and connected hypotheses through an evidence-driven, biased, and conservative diffusion process. As an alternative model, we constructed an RL model that updates the weights for the Repeat, Alternate, and XOR rules after each feedback.

We fit the models separately to each participant's choice data in Exp. 3. The fits of both the HD and RL models agreed with participants' behavioral patterns (Fig. 2BC). However, the HD model outperformed the RL model as well as a logistic regression model in goodness-of-fit in all conditions (Fig. 2D).

Moreover, with model parameters estimated in one single rule condition, the HD model can cross-predict participants' performances in all three rule conditions (Fig. 2E, upper), while the RL model failed to cross-predict (Fig. 2E, lower). Examining model parameters fitted from different conditions reveals that the RL model relied on heterogeneous initial bias parameters to obtain across-condition differences, which amounts to a mere description of the data pattern instead of reflecting the underlying learning mechanics. In contrast, the HD parameters were homogeneous across conditions. That is, its performance differences across conditions are emergent properties from the hypothesis-diffusion learning process.

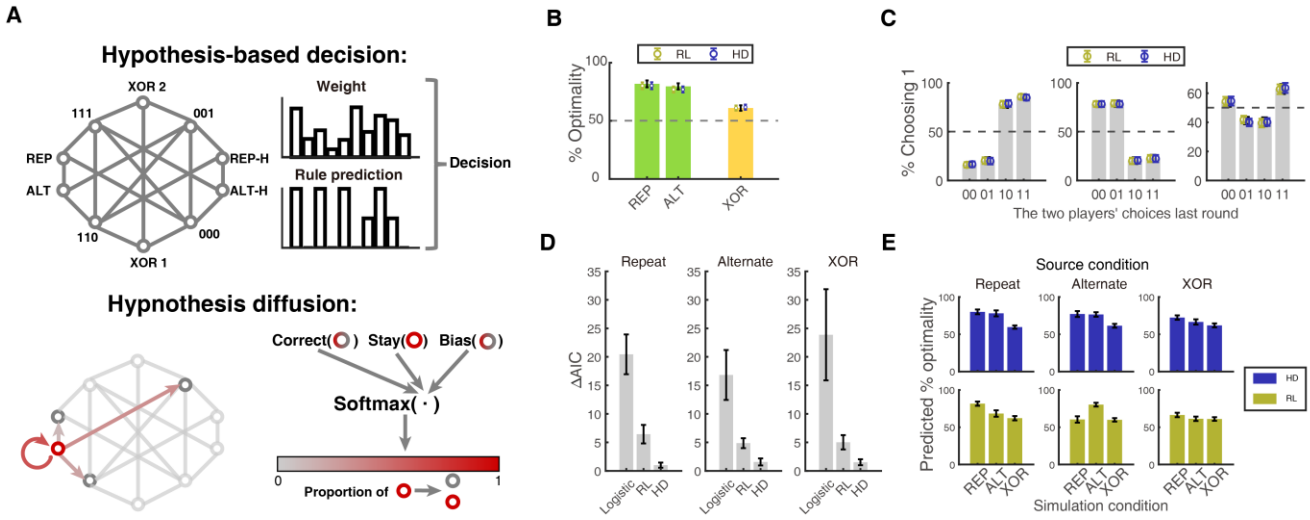

Figure 2: **A**, Hypothesis diffusion (HD) model: connected hypotheses (upper left) weighted for decision-making (upper right) and weights redistributed based on prediction accuracy, bias, and source (lower). **B** & **C**, Data vs. model fits for RL and HD. **D**, The HD model outperformed the RL and logistic regression models in goodness-of-fit (smaller AIC). **E**, The HD model parameters estimated from every single rule condition can cross-predict task performance in all other conditions (upper), while the RL model failed to do so (lower).

Acknowledgements

HZ was partly supported by the National Natural Science Foundation of China (32171095), National Science and Technology Innovation 2030 Major Program (2022ZD0204803), and funding from Peking-Tsinghua Center for Life Sciences.

References

Ashwood, Z. C., Roy, N. A., Stone, I. R., The International Brain Laboratory, Urai, A. E., Churchland, A. K., Pouget, A., & Pillow, J. W. (2022). Mice alternate between discrete strategies during perceptual decision-making. Nature Neuroscience, 25(2), 201–212.

Cohen, Y., & Schneidman, E. (2013). High-order feature-based mixture models of classification learning predict individual learning curves and enable personalized teaching. Proceedings of the National Academy of Sciences, 110(2), 684–689.

Cohen, Y., Schneidman, E., & Paz, R. (2021). The geometry of neuronal representations during rule learning reveals complementary roles of cingulate cortex and putamen. Neuron, 109(5), 839-851.e9.

Farashahi, S., Rowe, K., Aslami, Z., Lee, D., & Soltani, A. (2017). Feature-based learning improves adaptability without compromising precision. Nature Communications, 8(1), 1768.

Niv, Y., Daniel, R., Geana, A., Gershman, S. J., Leong, Y. C., Radulescu, A., & Wilson, R. C. (2015). Reinforcement Learning in Multidimensional Environments Relies on Attention Mechanisms. The Journal of Neuroscience, 35(21), 8145–8157.

Song, M., Baah, P. A., Cai, M. B., & Niv, Y. (2022). Humans combine value learning and hypothesis testing strategically in multi-dimensional probabilistic reward learning. PLOS Computational Biology, 18(11), e1010699.

Wang, M. C., & Soltani, A. (2023). Contributions of attention to learning in multi-dimensional reward environments [Preprint]. Neuroscience.