Learning of predictable rules mixed with random reinforcement in humans

Yuhao Jin (yj2525@columbia.edu)
Department of Biological Sciences, Columbia University 1212 Amsterdam Avenue
New York, NY 10027 USA

Greg Jensen (belarius@gmail.com)
Department of Psychology, Reed College 3203 SE Woodstock Blvd
Portland, OR 97202 USA

Jacqueline Gottlieb (jg2141@columbia.edu)
Department of Neuroscience, Columbia University 3227 Broadway
New York, NY 10027 USA

Vincent Ferrera (vpf3@cumc.columbia.edu)
Department of Neuroscience, Columbia University 3227 Broadway
New York, NY 10027 USA

Abstract:

Humans as well as animals are constantly learning novel predictable relationships to better adapt to the environment. However, such "learnable" patterns are often intermixed with noisy "unlearnable" randomness. It is not known if, when they are presented simultaneously, humans are capable of differentiating them, so that more energy can be invested in learnable rules. Here, we exposed humans with two pictorial sets: a "learnable" set in which the stimuli were implicitly ordered and the correct response was always to choose the higher-rank stimulus, and an "unlearnable" set in which stimuli were unordered and feedback was random regardless of the choice. The behavior patterns under the two sets were extremely polarized: Some participants ordered the stimuli in neither set (non-learners). Others ordered the stimuli in both sets, learning the correct order from the learnable set while behaving as though some ordering also existed from the unlearnable set, consistent with our previous finding from monkey behavior. Only when subjective ordering of the unlearnable set was strongly discouraged did many participants start to behave differently toward the two sets. Our results suggest that under the neutral condition humans did not differentiate well between real (learnable) patterns as opposed to random reinforcement, which contributes to deeper understanding of multi-rule learning and the formation of persistent superstitious biases.

Keywords: learnability; randomness; superstitious learning; transitive inference;

Introduction

Humans and animals have been tested in a wide range of learning tasks from stimulus-outcome association to complex hierarchical strategy planning. However, in natural environments, systematic rules may be intermixed with random unpredictable feedback. Therefore, organisms not only are learning but also need to manage "what to learn". Such "learning to learn" problems have been explored during serval past studies (Faraut et al., 2016; Piray et al., 2021; Ten et al., 2021; Simoen et al., 2024) but it is not well known whether humans are able to differentiate truly learnable rules from random or noisy relations. The ability to discriminate rule-based from random reinforcement would enable the prioritization of resources on the learnable problems and potentially speed up learning.

Here, we examined this question in the context of a "transitive inference" (TI) task that tested the ability to infer the ordinal relationships among a set of pictorial stimuli that had a hidden order (**Fig 1B**). The task is well-suited to our question because it has been extensively studied in a wide range of species and easy to be manipulated (Jensen et al., 2019). In our current task, human participants viewed pairs of pictures drawn from an ordered, learnable set (**L**, **Fig. 1A, top**) where the correct answer was to choose the higher ranked stimulus. On the other hand, to introduce

random associations, in randomly interleaved trials, they viewed pairs from an unlearnable image set (**U**, **Fig. 1A, bottom**) that had no hidden order, and their choices were reinforced randomly at the same overall rate as responses to the learnable list.

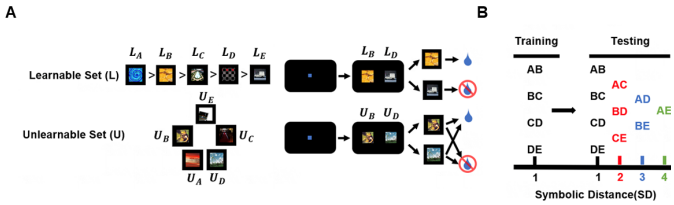

Figure 1: Task paradigm

To better understand how subjects approached the U trials, each session used one of two different reward schedules for the U trials. Under the "preference neutral" schedule (**PN**), the percentage of correctness for U pairs was equated to that for L pairs by dynamically adjusting it to match the mean performance for L pairs on the preceding 10 L trials. Under the second, "preference discouraging" schedule (**PD**), the probability to be correct for each U stimulus was inversely related to how recently the subject had selected it, thus discouraged repeated choice of any specific U stimulus, and yielded maximal performance if differences in U stimuli preferences were minimized and each U stimulus was selected equally.

Result

We first examined the participant's performance in the L set. Interestingly, 40% of the subjects failed to learn the task under either schedule (green), with the average accuracy at 52% (**Fig 2A, left**). The proportion of non-learners was much higher than what's reported from the past studies (10%, Jensen et al., 2021).

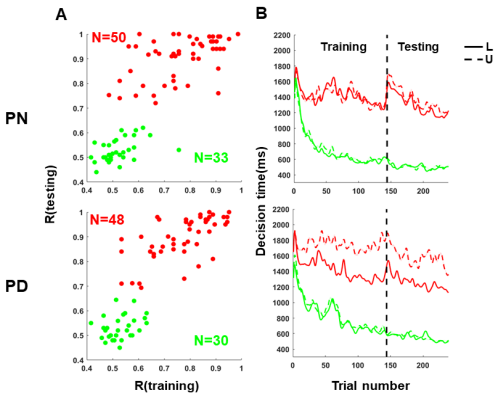

Figure 2: Overall performance and decision time

For the participants who learned the task (red), we estimated how frequently they chose each U stimuli using a model-based subjective ordering analysis (SOA). The analysis fit choices in each session based on the assumption that subjects represented stimuli along a continuum with some uncertainty about stimulus positions (**Fig. 3A**). The analysis produced a *z*-score indicating the relative rank of each stimulus. A stronger *z*-score gradient indicated stronger preference, more consistent choices and less overlap between inferred stimulus ranks. Expectedly, for the people who were considered to have learned the correct order from the L set, the gradients (slopes) over the *z*-scores were significantly higher than would be expected from a baseline of random responding (**Fig. 3B**, red and pink). Surprisingly, under PN schedule, the *z*-score gradients for U sets also displayed slopes that were significantly steeper than baseline for most of the people (**Fig. 3B** top, red). This suggests that subjects displayed consistent preferences among U stimuli, despite receiving rewards that were independent of stimulus. It is noteworthy that when such preferences are actively discouraged in the PD schedule, a large group of people displayed the slopes for the U set insignificantly from the baseline (**Fig. 3B** bottom, purple), indicating a decoupling of the behavior pattern between the L (ordered response) and U sets (random selection). This decoupling also showed up in the decision time (**Fig. 2B)** where there were much longer response times in the U trials compared to L trials in the PD schedule but equally long in the PN schedule. Such decoupled effects are not seen in our previous study that shows monkeys resisted the PD schedule by still showing consistent subjective orderings (Jin et al., 2022). Indeed, in the PD schedules, there is still about 35% of the active learners who retained their subjective orderings (**Fig. 3B** bottom, red), albeit the strength is significantly weaker than that in PN schedule.

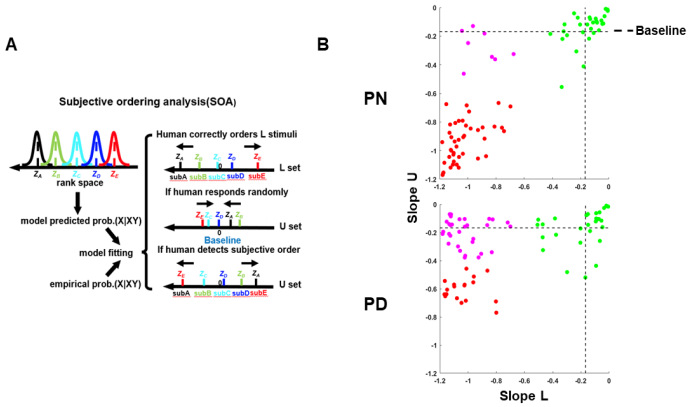

Figure 3: Human's subjective preference to the U stimuli under both schedules

To further delve into their behaviors, after the main task, humans were asked to put both L and U stimuli into the order they believed was correct, and also reported the confidence level (0-100) of their ordering (**Fig. 4A)**. Consistent with the aforementioned results, in the PN schedule, under both L and U sets, the self-reported ordering aligned well with the model-fitted Z scores, indicating that participants indeed showed ordered preferences for the U stimuli when there was none. Furthermore, the strength of the ordering also significantly correlated with the confidence level, showing that humans were more confident when such "decisional illusions" were stronger. On the contrary, in the PD schedule, the alignment between the reported and model-fitted ranks of the U stimuli was totally broken. Similarly, there was no correlation between the confidence and the slopes, further suggesting that humans treated L and U sets in a more divergent manner in PD compared to PN schedule (**Fig. 4B)**.

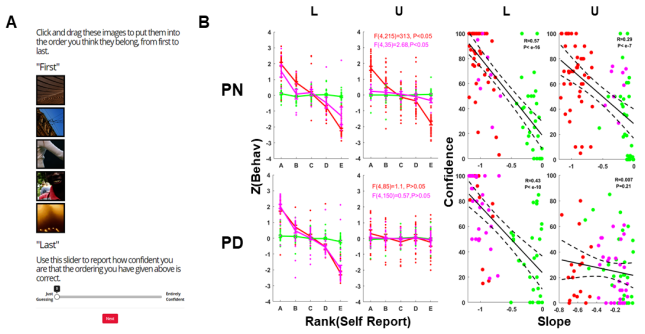

Figure 4: Human's post-task self-report

Overall, our results suggest that under the neutral environment where the relative gains and losses are evenly distributed, humans did not differentiate well between learnable patterns and random relationships. Instead, they tended to engage and treated the random relations as learnable by applying subjective rules on them. It is only when doing so incurs considerable losses that humans started to treat them differently and behave more accordingly to what the ground truth patterns are.

Acknowledgments

This work was supported by grant NIH-R01MH111703 from the National Institutes of Health (VPF).

References

Faraut, M. C., Procyk, E., & Wilson, C. R. (2016). Learning to learn about uncertain feedback. *Learning & memory (Cold Spring Harbor, N.Y.)*, *23*(2), 90–98.

Piray, P., & Daw, N. D. (2021). A model for learning based on the joint estimation of stochasticity and volatility. *Nature communications*, *12*(1), 6587.

Ten, A., Oudeyer, P. Y., & Gottlieb, J. (2021). Humans monitor learning progress in curiosity-driven exploration. *Nature communications*, *12*(1), 5972.

Simoens, J., Verguts, T., & Braem, S. (2024). Learning environment-specific learning rates. *PLoS computational biology*, *20*(3), e1011978.

Jensen, G., Alkan, Y., Ferrera, V. P., & Terrace, H. S. (2019). Reward associations do not explain transitive inference performance in monkeys. *Sci Adv, 5*(7), eaaw2089.

Jensen, G., Kao, T., Michaelcheck, C., Borge, S. S., Ferrera, V. P., & Terrace, H. S. (2021). Category learning in a transitive inference paradigm. *Memory & cognition*, *49*, 1020-1035.

Jin, Y., Jensen, G., Gottlieb, J., & Ferrera, V. (2022). Superstitious learning of abstract order from random reinforcement. *Proceedings of the National Academy of Sciences*, *119*(35), e2202789119.