Modelling Pavlovian Attentional Biases Using Reinforcement Learning

Francisco Garre-Frutos[*]
University of Granada
Granada, Spain
fgfrutos@ugr.es

David Luque[†]
University of Malaga
Málaga, Spain
dluque@uma.es

Pablo Martínez-López[†]
University of Malaga
Málaga, Spain
mlpablocorreo@gmail.com

Juan Lupiáñez[*]
University of Granada
Granada, Spain
jlupiane@ugr.es

Miguel A. Vadillo[‡]
Autonomous University of Madrid
Madrid, Spain
miguel.vadillo@uam.es

Abstract

A central question in modern models of Pavlovian learning is how learning influences attention. Mackintosh's theory proposes that stimuli that reliably predict significant outcomes receive more attention to maximize learning, whereas Pearce-Hall's theory proposes that uncertainty drives attention to promote exploration. Empirical evidence supports both views, showing that stimuli that consistently predict reward or reward variability are more likely to capture attention, even when they act as distractors in conflict with other goal-relevant stimuli. Although these findings are often assumed to reflect the mechanisms of Mackintosh and Pearce-Hall, there have been surprisingly few attempts to explicitly model the learning dynamics underlying such attentional biases. In this work, we developed two hierarchical Bayesian models implementing Mackintosh and Pearce-Hall principles and compared their fit to two datasets in which either value or uncertainty was manipulated. Surprisingly, our results show that the Pearce-Hall model can account for both experimental effects. We argue that this likely reflects a methodological confound in how value is typically manipulated, as high-value cues often also exhibit greater outcome variance, raising the possibility that some experimental effects reflect uncertainty-driven rather than value-driven attention.

Keywords: reinforcement learning, Pavlovian, value, uncertainty, attention

Acknowledgements

This work was supported by an FPU predoctoral grant (ref. FPU20/00826) to FGF.

[*]Mind, Brain and Behavior Research Center (CIMCYC) and Department of Experimental Psychology, University of Granada, Granada, Spain
[†]Department of Basic Psychology and Málaga Institute of Biomedical Research (IBIMA), University of Málaga, Málaga, Spain
[‡]Department of Basic Psychology, Autonomous University of Madrid, Madrid, Spain

1 Introduction

The relationship between learning and attention has been a central question in Pavlovian learning for decades. Since Rescorla & Wagner (1972), numerous accounts have attempted to incorporate attention into associative learning models. For example, Mackintosh (1975) proposed that attention increases for cues that reliably predict significant outcomes, thereby promoting further learning about those cues, whereas Pearce & Hall (1980) argued that attention increases for cues associated with greater uncertainty or variability in outcomes, thereby promoting exploratory learning. Both principle have received empirical support. For instance, in the study by Le Pelley et al. (2015), participants performed a visual search task in which singleton distractors were consistently paired with two different reward magnitudes. In this procedure, although the distractors predicted reward, looking directly at them caused omission of reward, even though attending to the high-reward predictive distractor was counterproductive to the task goals, Le Pelley et al. (2015) found that participants tended to look more at high than the low-value distractor, reflecting a Pavlovian bias that contradicted the participants' goals.

According to Le Pelley et al. (2016), this attentional bias could be explained by a modified Mackintosh model, in which the associative strength (V) of a distractor is described by the following Rescorla-Wagner rule:

$$V_n = V_{n-1} + \eta\, \alpha_n \left(\lambda_n - V_{n-1}\right) \tag{1}$$

where η is a learning rate parameter, λ is the reward in trial n, and α is a weight parameter reflecting attention to the cue. From this formula, V_n is updated as a function of the prediction error, $\left(\lambda_n - V_{n-1}\right)$. Unlike other formulations of the Mackintosh model, Le Pelley et al. (2016) assumed that α is updated as

$$\alpha_n = |V_{n-1}| \tag{2}$$

In other words, attention to the cue is a function of the asymptotic λ. This formulation of the Mackintosh model can explain why value-driven distraction increases over trials and remains stable with training.

In contrast, Pearce et al. (1982) postulated that attention is mainly driven by prediction errors:

$$\alpha_n = \gamma\, |\lambda_n - V_{n-1}| + \left(1 - \gamma\right) \alpha_{n-1} \tag{3}$$

where γ is a decay parameter that weights the relevance of previous values of α. The Pearce-Hall principle predicts that cues associated with more variable outcomes (and thus higher prediction errors) will receive more attention. Consistent with this prediction, and using a procedure similar to Le Pelley et al. (2015), Pearson et al. (2024) showed that when irrelevant distractors were associated with different levels of reward variability, high-variance distractors received more attention than low-variance distractors, even when the overall expected value was higher for low-variance distractors (see also Le Pelley et al., 2019).

The results of both Le Pelley et al. (2015) and Pearson et al. (2024) suggest that the principles of Mackintosh and Pearce-Hall may be applicable in different contexts. Despite explicit formulations of how attention should change as a function of learning, there are few attempts to explicitly model the learning dynamics associated with either theory in Pavlovian attentional biases. In light of this, in this work, we developed two hierarchical Bayesian reinforcement learning (RL) models in which attention is modeled according to formulas (2) and (3). We then fit these two models to two openly aviable datasets where either reward value or uncertainty is manipulated, and compared the peformance of each model to describe the data.

2 Method

2.1 Data

We used the open-source data from Experiment 2 of the Le et al. (2024) study and Experiment 1 of the Chow et al. (2024) study. In both experiments, participants performed the additional singleton task (Theeuwes, 1992). In each trial, participants had to search for unique diamond-shaped stimulus surrounded by circle non-target shapes, one of which was a color-singleton distractor (a circle of a different color) (Figure 1). Critically, during the task, participants could earn rewards (points) depending on the color of the distractor. In the Le et al. (2024) study, if participants managed to look at the target stimuli without looking at the distractor, they would earn either a high or low reward depending on the color of the distractor (a high-reward or a low-reward color). However, looking at the distractor would result in the omission of the reward. In Chow et al. (2024), the color of the distractor signaled outcome variability. One distractor color was always associated with the same outcome (low variance), while the other singleton distractor signaled two possible outcomes (high variance) with a 50% probability. The expected value of both distractors, however, was matched overall.

We employed this two specific datasets because they are directly comparable in design and have an unusually large number of participants compared to much of the literature. In Experiment 2 of Le et al. (2024), there are 84 participants with 384 trials each, whereas in Chow et al. (2024), there are 40 participants with 512 trials each. We provide the data and analysis code for this study in the following repository: `https://github.com/franfrutos/rldm25_vmac`.

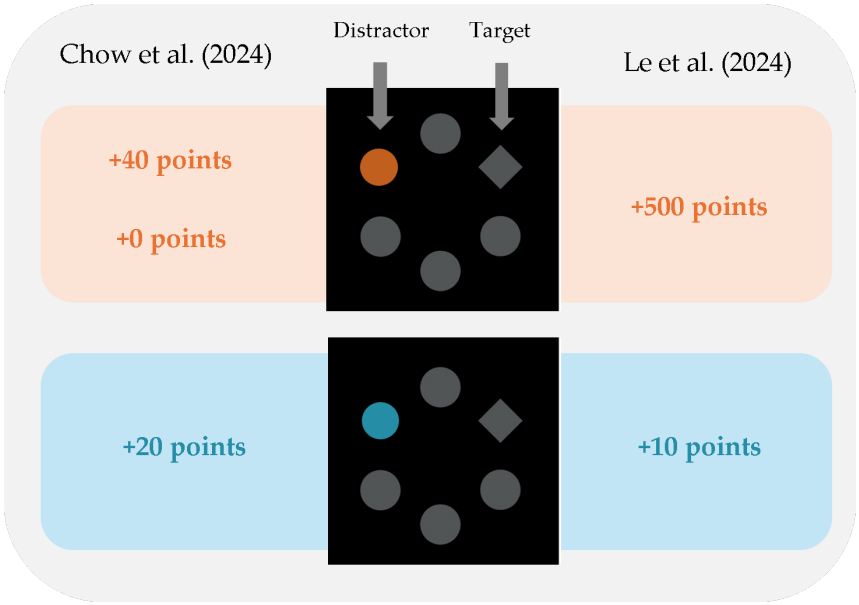

Figure 1: Schematic representation of the task design. In Le et al. (2024), the high-value distractor (top) was associated with more reward than the low-value distractor (bottom), although looking at the distractor before the target produced reward omission. In Chow et al. (2024), the high-variance distractor was associated with two possible outcomes at 50% probability, while the low-variance distractor was always associated with the same outcome.

2.2 Model Specification

For the datasets described above, we model whether a participant looked at the distractor (1) or the target (0) on trial i. This is implemented using a logistic regression (logit^{-1} function), where $\Pr(\text{Look}_i = 1)$ is determined by the latent attention (α) for each distractor (s):

$$\begin{aligned} \Pr(\text{Look}_i = 1) &\sim \text{Bernoulli}(\theta_i), \\ \theta_i &= \frac{1}{1 + e^{-(\beta_{0,j} + \beta_{1,j} \cdot \alpha_{s[i]} + \beta_{2,j} \cdot t_i)}}, \end{aligned} \tag{4}$$

where $\beta_{0,j}$ is the intercept for subject j, $\beta_{1,j}$ is the slope that relates α to the probability of looking at the distractor, and $\beta_{2,j}$ controls for practice effects across trials (t). This simple logistic function can be viewed as a softmax rule with two levels: the target (reference) and the distractor. Thus, β_0 can be interpreted as a bias term (a tendency to choose the target or the distractor), and β_1 as an *inverse temperature* parameter. We parameterized β_1 as $\text{logit}^{-1}(\cdot) \cdot 10$, so it could only take positive values in the range [0, 10]. A higher β_1 increases selection of distractors with higher α.

On trial i, each distractor s updates its V_s according to equation (1), where λ_i equals the observed reward[1] on that trial. Regarding α_s, both models assume that participants start with an initial level of attention (α_0). Then, in the *Mackintosh* model (Le Pelley et al., 2016), α is updated according to equation (2), while in the *Pearce-Hall* model, α is updated following equation (3), where η and γ are subject-specific parameters governing the update of V and α.

All subject-specific parameters (k) are assumed to be drawn from a normal distribution, using non-centered parameterization to improve sampling efficiency:

$$\beta_{k,j} = \mu_k + \sigma_k \cdot z_{k,j}, \quad z_{k,j} \sim \mathcal{N}(0, 1). \tag{5}$$

Subject-specific parameters ($\beta_{k,j}$) are modeled by scaling group-level means (μ_k) with individual deviations (σ_k) and standardized parameters ($z_{k,j}$). Bounded parameters (η, γ, α_0, and β_1) are then transformed using the logit^{-1} function to ensure they remain within the [0, 1] interval. We used weakly informative priors ($\mathcal{N}(0, 1)$) for group-level means μ, and truncated normals for σ.

3 Results

The models described above were programmed in Stan (Stan Development Team, 2024). We fit both models to each dataset using 6000 warm-up iterations and 8000 samples across four chains ($\bar{R} < 1.01$). We assessed the relative fit of the

[1] λ is scaled by $\frac{\lambda_i}{\max(\lambda)}$, which ensures that V takes values in the range [0, 1].

two models to each dataset using the Pareto-smoothed importance sampling leave-one-out cross-validation (PSIS-LOO; Vehtari et al., 2017). Table 1 shows the relative performance of the *Mackintosh* model compared to the *Pearce-Hall* model, where a negative ELPD means worse performance.

Table 1
Relative fit analysis

	Chow et al. (2024)		Le et al. (2024)	
Model	ΔELPD	ΔSE	ΔELPD	ΔSE
Pearce-Hall	0	0	0	0
Mackintosh	-254	24	-79	21

Note. ELPD = Expected Log-Pointwise Predictive Density. SE = Standard Error.

According to Vehtari et al. (2017), if the difference in ELPD between models deviates by more than 2 SEs, the model with the higher ELPD is likely to be the better fit for the data. By this criterion, the *Pearce-Hall* model outperformed the *Mackintosh* model in both datasets.

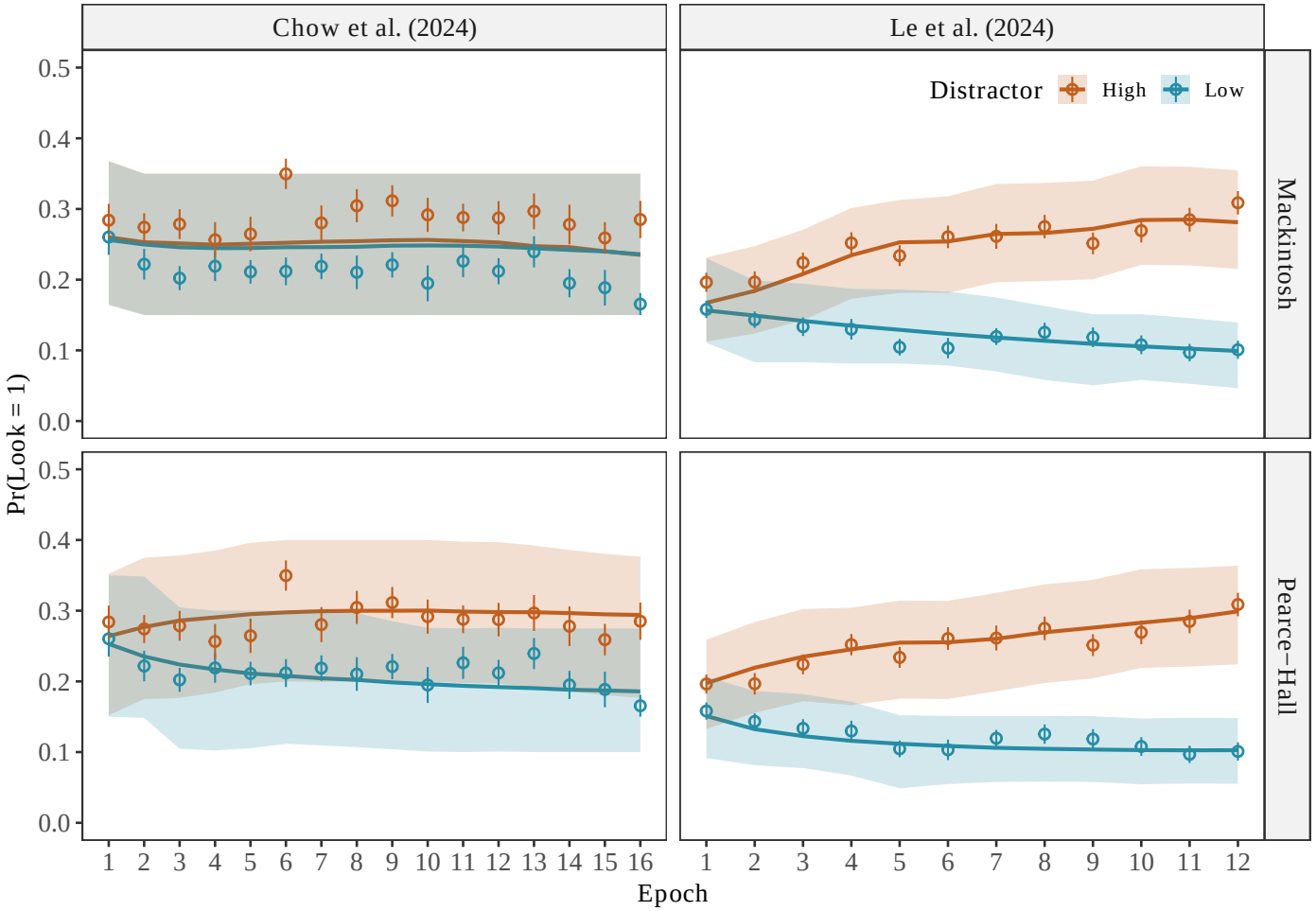

Figure 2: Posterior Predictive Checks (PPC) for the *Mackintosh* and *Pearce-Hall* models, split by dataset. PPCs were generated by simulating new observations based on the fitted parameters and aggregating results at the group level. Solid lines represent the mean, and the shaded area is the 89% High Density Interval (HDI) of the posterior predictions across epochs of 32 trials. The dots represent the observed mean proportion of gazes at the distractor, while the error bars show the SEM.

Although the previous analysis suggests that *Pearce-Hall* has the best relative fit for Le et al. (2024) and Chow et al. (2024) datasets, we should also verify whether both models predict the theoretically relevant experimental effects—

namely, whether participants look more often at the high-value (or high-variance) distractor relative to the low-value (or low-variance) distractor. To examine this, we compared the absolute fit of the *Mackintosh* and *Pearce-Hall* models via Posterior Predictive Checks (PPC). Figure 2 displays the PPCs as a function of distractor type and epochs (32-trial bins). As expected, the *Pearce-Hall* model alone captures the greater attention to the high-variance distractor in the Chow et al. (2024) dataset, whereas the *Mackintosh* model predicts no differences between distractors, because the asymptotic λ was matched. Interestingly, although the *Mackintosh* model provides a reasonable fit to Le et al. (2024) data, the *Pearce-Hall* model's predictions are closer. It also appears to better capture the earlier stages of learning, which may explain why it provides superior predictive accuracy.

4 Conclusions

Our results show that it is possible to use RL models to capture Pavlovian attentional biases. We found that when uncertainty is manipulated (Chow et al., 2024), only the *Pearce-Hall* model explains the observed data. When value is manipulated (Le et al., 2024), somewhat unexpectedly, *Pearce-Hall* also predicts greater attention to the high-value distractor. As suggested by Pearson et al. (2024), outcome variance is a measure of uncertainty because it reflects the amount of prediction error associated with a cue. While outcome variance is explicitly manipulated in Pearson et al. (2024), in the paradigm used by Le Pelley et al. (2015), manipulation of value can be confounded with differences in outcome variance between distractors. As explained above, in Le Pelley et al. (2015), when participants look at the distractors, they cause a reward omission, which can be viewed as a prediction error. This prediction error is larger for the high-value distractor than for the low-value distractor, and the proportion of prediction errors increases as a function of attention because the probability of looking at a distractor increases across trials only for the high-value distractor. Thus, the high-value distractor not only represents the highest value, but is also associated with greater variability. This methodological confound may explain why the *Pearce-Hall* principle is a better fit for Le et al. (2024).

5 References

Chow, J. Y. L., Garner, K. G., Pearson, D., Heber, J., & Le Pelley, M. E. (2024). Effects of instructed and experienced uncertainty on attentional priority. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0001427

Le, J. T., Watson, P., & Le Pelley, M. E. (2024). Effects of outcome revaluation on attentional prioritisation of reward-related stimuli. *Quarterly Journal of Experimental Psychology*, 17470218241236711. https://doi.org/10.1177/17470218241236711

Le Pelley, M. E., Mitchell, C. J., Beesley, T., George, D. N., & Wills, A. J. (2016). Attention and associative learning in humans: An integrative review. *Psychological Bulletin*, *142*(10), 11111140. https://doi.org/10.1037/bul0000064

Le Pelley, M. E., Pearson, D., Griffiths, O., & Beesley, T. (2015). When Goals Conflict With Values: Counterproductive Attentional and Oculomotor Capture by Reward-Related Stimuli. *Journal of Experimental Psychology: General*, *144*, 158–171. https://doi.org/10.1037/xge0000037

Le Pelley, M. E., Pearson, D., Porter, A., Yee, H., & Luque, D. (2019). Oculomotor capture is influenced by expected reward value but (maybe) not predictiveness. *Quarterly Journal of Experimental Psychology*, *72*(2), 168–181. https://doi.org/10.1080/17470218.2017.1313874

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*(4), 276–298. https://doi.org/10.1037/h0076778

Pearce, J. M., & Hall, G. (1980). A model for pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*(6), 532–552. https://doi.org/10.1037/0033-295X.87.6.532

Pearce, J. M., Kaye, H., & Hall, G. (1982). Predictive accuracy and stimulus associability: Development of a model for pavlovian learning. *Quantitative Analyses of Behavior*, *3*, 241–255.

Pearson, D., Chong, A., Chow, J. Y. L., Garner, K. G., Theeuwes, J., & Le Pelley, M. E. (2024). Uncertainty-modulated attentional capture: Outcome variance increases attentional priority. *Journal of Experimental Psychology. General*, *153*(6), 1628–1643. https://doi.org/10.1037/xge0001586

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations on the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). Appleton-Century-Crofts.

Stan Development Team. (2024). *Stan modeling language users guide and reference manual, version 2.36.0*. http://mc-stan.org/

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics*, *51*(6), 599–606. https://doi.org/10.3758/BF03211656

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4