Project 23: Policy compression

1) Implement the policy compression model described in Chapter 11 (Section 3) of *Computational Foundation of Cognitive Neuroscience*. In particular, use Eq. 26 to define an augmented reward function consisting of the immediate reward (r) minus the policy cost, $\lambda \log \frac{\pi(a|s)}{p^*(a)}$, where $p^*(a)$ is an incrementally updated estimate of the recent action probability (e.g., a moving average). This augmented reward function can be plugged into the standard policy gradient learning model described earlier in the chapter. Show that the model reproduces a tendency to repeat recent actions (perseveration).

2) Analyze how the prediction errors change with policy cost, relating this to the findings from Gershman & Lak (2025).

3) What does the model imply about the relationship between action stochasticity and cognitive resource availability (λ)? Simulate the effects of changing λ on action stochasticity and perseveration. Discuss your simulations in connection with the behavioral findings from Wu et al. (2022).

References:

Gershman, S. J., & Lak, A. (2025). Policy complexity suppresses dopamine responses. *Journal of Neuroscience, 45*, e1756242024.

Wu, C. M., Schulz, E., Pleskac, T. J., & Speekenbrink, M. (2022). Time pressure changes how people explore and respond to uncertainty. *Scientific Reports, 12*, 4122.