A Reinforcement Learning Model for Task-modulated Perception

Mohammadreza Bigham (mrez.bigham@gmail.com)
Institute for Cognitive and Brain Sciences, Shahid Beheshti University
Tehran, Iran

Shervin Safavi (research@shervinsafavi.org)
Computational Neuroscience, Department of Child and Adolescent Psychiatry,
Faculty of Medicine, TU Dresden, Dresden, Germany
Department of Computational Neuroscience, Max Planck Institute for Biological Cybernetics,
Tübingen, Germany

Abstract

Perception is a highly active process; therefore, it can and should be approached from a decision-theoretic perspective. Elevating perception from a mere sensory inference to a decision process requires us to consider, for instance, how the value of sensory objects influences what we perceive and how the task at hand affects our perception. Here, we suggest that multistable perception could be a suitable candidate to study task-modulated perception in both humans and animals. Multistable perception is the dynamical alternation that arises when a single sensory input has more than one interpretation or explanation. Multistable perception is one of the most venerable perceptual phenomena that has been formalized as a decision process. We extend the previous model of perceptual multistability (Safavi & Dayan, 2022, 2024) by incorporating a richer state space and action repertoire for a reinforcement learning agent, and we show that this allows us to explain the established task-modulated perception during perceptual multistability. Our model replicates and explains recent findings on the modulation of perception by task observed in previous studies (Dieter et al., 2016). This is achieved by incorporating two key elements– changes in attentional resource allocation and representation of the environment volatility– into a reinforcement learning paradigm. These changes are implemented in the model by systematically adjusting the observation and transition functions in our partially observable Markov decision model of perpetual decisions. Overall, our findings further support the view that perception is an active, goal-directed process, aligned with principles shared by other aspects of cognition.

Keywords: perceptual decision-making; perceptual value; internal reinforcement learning

Perception is an active process

Perception is an active process that shapes sensory processing in response to the demands of the task at hand. Thus, it must be approached through the lens of reinforcement learning (RL) and decision-making. Perceptual multistability – where perception alternates between competing interpretations despite constant sensory input, as seen in phenomena like binocular rivalry and Necker cube (Blake & Logothetis, 2002), offers a unique window into studying perception as a decision process (Safavi & Dayan, 2022).

In this study, we model perceptual multistability based on the RL framework introduced in previous studies (Safavi & Dayan, 2022, 2024). This framework conceptualizes perception as a dynamic value-based decision-making process using a partially observable Markov decision process (POMDP). It offers a novel approach to investigate how cognitive factors (such as attention and reward) affect perception. Here, we extend Safavi and Dayan (2022, 2024)'s framework to explain *how task demands modulate perception*.

A paradigm to study task-modulated perception

In binocular rivalry, each eye is shown a different image (Figure 1A, top), leading to a perceptual alternation between two stimuli (Figure 1A, bottom). The duration of dominant percepts follows a gamma-like distribution (Brascamp et al., 2005), similar to other response distributions.

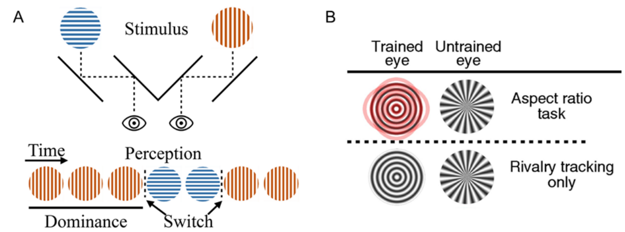

Figure 1: (A) In binocular rivalry (top), participants' two eyes see different images; with a percept that switches spontaneously (bottom). (B) Two experimental settings in Dieter et al. (2016).

Dieter et al. (2016) modified the conventional binocular rivalry experiment and demonstrated that participants' perception is strongly influenced by the task at hand. In their experiment, participants observed pinwheel and bull's-eye images as two competing stimuli, each presented independently to a separate eye (Figure 1B). In one phase of the experiment, called the *rivalry tracking only task*, observers were instructed to report only the dominant image, whether they perceived the bull's-eye or the pinwheel. In the other phase, the *aspect ratio task*, observers viewed the same images, but the bull's-eye image was stretched either horizontally or vertically (see Figure 1B, shaded in red) *and dynamically changed throughout the task*. In this task, observers were required to track and report changes in the bull's-eye aspect ratio by pressing keys. All participants performed both tasks before and after several training sessions. During training, participants completed the aspect ratio task across twelve sessions. Dieter et al. (2016) reported, *prolonged dominance duration of the task-related stimulus (bull's eye) in the post-training session* (see Figure 3A). To explain why perceptual dominance varied across tasks before and after training, we developed an RL model, as described in the following.

A POMDP with task-modulated perception

We extend the previous RL model Safavi and Dayan (2024) to a richer state space and action repertoire. We expand the state space from two states to a three-state configuration, representing distinct perceptual experiences that agents can have (similar to human participants in the experiment): The pinwheel, the widened bull's-eye, and the elongated bull's-eye patterns; and assign perceiving each state to an internal action (3 internal actions), as well as *key presses* as external actions (2 external actions) for behavioral report of widened and elongated bull's eye in *aspect ratio task* (but not in the *rivalry tracking only task* as these actions were not involved). The agent switches between percepts by selecting internal actions based on the dynamic value of possible percepts and belief distribution over all possible states.

We implement a dynamic component to observation and transition functions to incorporate the influence of training on participants' behavior. Before training, the model assumes equal transition probabilities and observation noise for all possible percpets. However, after training, perceptual states relevant to the aspect-ratio tracking task gain heightened volatility, which is incorporated by the increase in state transition probability, specifically between widened and elongated bull's eye (capturing their dynamic changing in the aspect-ratio tracking task), and reduced observation noise for the task-relevant stimulus percepts (capturing the elevated attention).

The reward associated with an agent's perception follows an exponential decay, which represents how sensory value diminishes over time due to cognitive fatigue or boredom (Brielmann & Dayan, 2022) as it was also suggested by Safavi and Dayan (2024). In addition, switching between percepts has a cost for the agent (agent needs to deconstruct one percept and reconstruct a new percept), and a constant (small) reward is given for suppressed percepts (more sophisticated formulation can also be incorporated, see, Safavi & Dayan, 2024).

In particular, our model replicates and explains the change in the dominance **distributions** observed in the aspect ratio tasks before and after the test (observed in Dieter et al., 2016, also see Figure 3). It also indicates shorter dominance durations for the task-relevant stimulus in the pre-test and longer, more sustained dominance in the post-test. Furthermore, the increased predominance of the task-relevant stimulus in both tasks aligns with behavioral data averages (Figure 2).

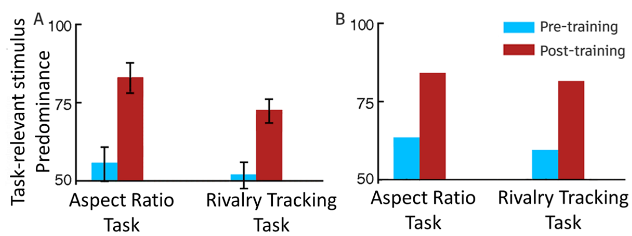

Figure 2: (A) The predominance results reported in Dieter et al. (2016) and (B) with our model.

This task-modulated behavior can be understood as a consequence of how the brain allocates attentional and cognitive resources in accordance with the task's goals/demands and strategies, incorporating a dynamic observation function that is shaped by the training, thus, explaining how attentional allocation reduces uncertainty in perceiving task-relevant stimulus after training sessions. The observer allocates more cognitive and attention resources, which in the model is represented by lower observation noise, to the task-relevant stimulus in the post-tests. Additionally, the update in the transition function indicates the representation of the environment in the agent's representation of the temporal structure of world states, which is dynamically changing in the aspect ratio task.

The change in the transition function captures the learning of the environment's volatility. Alternation between widened and elongated bull's eye (task-relevant stimulus) occurs when the bull's eye is dominant. As the predominance of task-relevant stimulus increases in the training sessions, the probability of transitioning between both percepts enhances. Thus, the transition function is adjusted (during the training) for the post-tests in order to capture the volatility of the environment that the stimulus dynamically changed between widened and elongated bull's-eye stimuli.

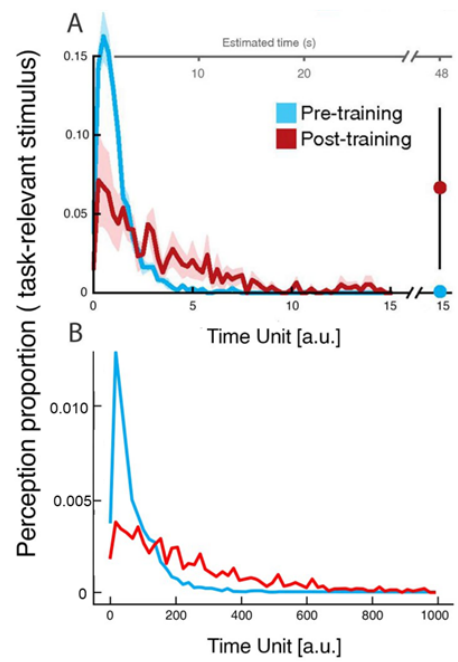

Figure 3: Percept duration histogram for task-relevant stimulus in (A) the experiment of Dieter et al. (2016) and (B) our model.

Overall, our model precisely captures the temporal dynamics of task-modulated perceptual behavior (both averages and distributions). The model captures it through learning the temporal dynamics of the environment incorporated in the task structure and the strategic allocation of cognitive resources (attention), which enhances perceptual predominance for task-relevant stimuli. Crucially, this approach enables us to replicate the shape of perceptual dominance distributions observed in Figure 3 — an aspect that value-free models struggle to explain as the temporal structure was not incorporated in these models (Brascamp et al., 2018), whereas in POMDPs, transition function explicitly takes that into account.

Acknowledgements

This work was supported by the German Research Foundation (DFG) grant 550411021, and the Max Planck Society. SS acknowledges the add-on fellowship from the Joachim Herz Foundation.

References

Blake, R., & Logothetis, N. (2002). Visual Competition. *Nat. Rev. Neurosci.*, *3*(1), 13–21.

Brascamp, J., Sterzer, P., Blake, R., & Knapen, T. (2018). Multistable perception and the role of the frontoparietal cortex in perceptual inference. *Annual review of psychology*, *69*, 77–103.

Brascamp, J., van Ee, R., Pestman, W. R., & van den Berg, A. V. (2005). Distributions of alternation rates in various forms of bistable perception. *Journal of Vision*, *5*(4), 1–1.

Brielmann, A. A., & Dayan, P. (2022). A computational model of aesthetic value. *Psychological Review*, No Pagination Specified–No Pagination Specified.

Dieter, K. C., Melnick, M. D., & Tadin, D. (2016). Perceptual training profoundly alters binocular rivalry through both sensory and attentional enhancements. *Proceedings of the National Academy of Sciences*, *113*(45), 12874–12879.

Safavi, S., & Dayan, P. (2022). Multistability, perceptual value, and internal foraging. *Neuron*.

Safavi, S., & Dayan, P. (2024). A decision-theoretic model of multistability: Perceptual switches as internal actions. *bioRxiv*, 2024–12.