Recent successes in building agents with superhuman performance have led to reinforcement learning (RL), becoming a dominant theoretical framework to understand decision-making through interaction with the world (1). However, recent RL algorithms still have major limitations, such as lack of the ability to develop goal-directed policies or reliance on large amounts of experience to learn (2). These limits impede the ability to rapidly adapt in dynamic environments where tasks or contexts frequently change.

In contrast, humans have a remarkable ability to rapidly adapt to environmental changes with limited experience. Recent findings in decision neuroscience suggest that the brain uses not only multiple control systems for RL but also a flexible metacontrol mechanism to select among control options, each different trait associated with prediction performance, cognitive load, and learning speed (3). Understanding how the brain implements these options could lead to brain-inspired RL algorithms that can work in real control problems for robots (4). Here, we discuss recent findings on human RL that may address several key challenges in robotics: performance-efficiency-speed trade-offs, conflicting demands in multirobot settings, and the exploration-exploitation dilemma.

First, accumulating evidence in decision neuroscience indicates that humans take advantage of two different behavior control strategies: (i) stimulus-driven habitual and (ii) goal-directed cognitive control (5). Habitual control is automatic and fast, despite being fragile in a volatile environment, and is well accounted for by model-free RL, which incrementally learns the values of actions through trial and error without a model of the environment. Conversely, goal-directed control can rapidly adapt to changes in the environment, but it is cognitively demanding. It guides actions by learning a model of the environment and uses this knowledge base to quickly adapt to changes in environmental structure, such as learning latent (hidden) causes within state-action space.

Second, RL algorithms usually require a large amount of experience to adequately learn causal relationships in the presence of different environmental factors (incremental learning). Humans, however, learn fast—often after a single exhibition of an event never experienced before (“one-shot learning”) (5). Recent neuroscience studies (5, 6) found that, when interactions with the environment are limited, humans have a strong tendency to increase their learning rates; they strive for quickly making sense of unknown parts of the environment, even when this compromises safety. These results suggest that the brain directly implements computation to find a trade-off between performance and speed.

Third, accumulating evidence supports the notion that the prefrontal cortex implements metacontrol to flexibly choose between different learning strategies, such as between model-based and model-free RL (7, 8) and between incremental and one-shot learning (5). In a new environment, metacontrol accentuates performance by favoring model-based RL. Because this is computationally expensive, the brain resorts to model-free RL when it finds little benefit from further learning: Either the environment is sufficiently stable to make precise predictions or highly unstable such that predictions from model-based RL become less reliable than those from model-free RL. In other situations, metacontrol prioritizes speed. When the uncertainty in the estimated cause-effect relationships is high, the brain tends to transition to one-shot learning to quickly resolve uncertainty in predicting outcomes. However, when the agent is equally uncertain about all possible causal relationships, it resorts to incremental learning to ensure safe learning. Together, they suggest that brain-like metacontrol can deal with performance-efficiency-speed trade-offs.

Fourth, human RL may account for social phenomena that have been important in human evolution. In human societies where multiple agents are interacting, there are social dilemmas that have partially competitive and partially aligned incentives (9). Approaches using model-based RL successfully achieve cooperation in more complex temporally extended settings [e.g., (10, 11)]. These models often work in two stages: First, there is a planning stage where the agent uses its model of the game’s rules to simulate a large number of games with itself and learns separate cooperation and defection policies by
Causal learning
Incremental learning
One-shot learning

Social learning

Metacognition

Brain-inspired solutions to robot learning. Neuroscientific views on various aspects of learning and cognition converge and create a new idea called prefrontal metacontrol, which can inspire researchers to design learning agents that can address various key challenges in robotics such as performance-efficiency-speed, cooperation-competition, and exploration-exploitation trade-offs.

for rapid adaptation to the context change while maintaining robustness against environmental noise. Such a strategy has potential for augmenting robot decision-making in several ways—for instance, in resolving exploration-exploitation trade-offs by over-seeing how much confidence should drive the desire to learn.

In conclusion, the integration of findings from human decision neuroscience can offer valuable insights into action control systems for robots, leading to safer, more capable, and more efficient learning. Such an interdisciplinary approach should also yield insights for neuroscience, providing a robust test base for developing new theories of human decision computation.

REFERENCES AND NOTES


