Cognition 113 (2009) 259–261

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Editorial

Reinforcement learning and higher level cognition: Introduction to special issue

Reinforcement learning (RL), originally an area of computer science concerned with learning to obtain rewards or avoid punishments by trial and error (Sutton & Barto, 1998), has recently played an influential role on cognitive science and systems neuroscience. Spurred initially by remarkable parallels between computational algorithms for solving such problems in engineered systems such as robots, and the observed firing properties of midbrain dopamine neurons in primates working for reward (Montague, Dayan, & Sejnowski, 1997), such models have been extended to encompass the basal ganglia circuitry primarily targeted by dopamine and, importantly, its putative behavioral and cognitive functions (in motor control, reward, and learning) and dysfunctions (as in Parkinson's disease and drug addiction) (Frank, Seeberger, & O'Reilly, 2004; McClure, Daw, & Montague, 2003; Moustafa, Cohen, Sherman, & Frank, 2008; Niv, Daw, Joel, & Dayan, 2007; Redish, 2004). These models have also recently been fruitfully applied to the analysis and explanation of neuroimaging studies involving choice and reward in humans (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006; O'Doherty, Dayan, Friston, Critchley, & Dolan, 2003; Pessiglione, Seymour, Flandin, Dolan, & Frith, 2006).

The psychological content of these theories, however, tends to resemble formalized versions of Thorndike's (Thorndike, 1911) law of effect. To deliver on the tantalizing promise of a theory spanning biology, psychology, and behavior will therefore require moving beyond these simple behaviorist roots and perhaps belatedly embracing the cognitive revolution. Indeed, an active research frontier is the extension of these theories toward learning and representation in more complex, higher level cognitive domains, including the adaptive regulation of working memory and cognitive control; goal representations and motivation; planning; search; and problem solving (Daw, Niv, & Dayan, 2005). At the neurobiological level, this work extends the models from their subcortical home territory to explore interactions with higher level circuitry such as prefrontal cortex (O'Reilly & Frank, 2006). The papers in this special issue provide insight into how sophisticated cognitive processes might leverage the brain's simpler reinforcement machinery and build on it to support adaptive behavior in a complex world.

A well known shortcoming in simple stimulus-response RL models is a failure to represent or exploit any environmental structure. Botvinick, Niv, and Barto (2009) review frameworks from computational RL that address one part of this problem by exploiting hierarchical structure in the sequences of actions used to obtain goals. In the version they focus on, sequences of primitive actions are strung together into higher level "options" that can then operate as building blocks so as more easily to learn still richer behaviors in complex environments (Sutton, Precup, & Singh, 1999). The authors review classic findings in neuroscience and psychology through the lens of this formalism, and sketch its ramifications for the conventional view of neural RL.

Reynolds and O'Reilly (2009) study a related problem of hierarchy in RL – representing the levels of contingent rules for determining a response – using quite a different methodology. These authors use large-scale neural network models simulating interactions between prefrontal cortex, basal ganglia and dopaminergic systems. Here, among the actions selected by the basal ganglia, which are acquired via RL, are those controlling whether or not to update prefrontal working memory states (O'Reilly and Frank, 2006). The tasks used in this domain are typically designed such that the behavioral relevance of stimuli depends on those that had appeared previously, and thus working memory updating should depend contingently on prior working memory context. Reynolds and O'Reilly show that multiple interacting BG-PFC circuits may be arranged hierarchically such that more anterior PFC regions come to represent more abstract (hierarchical) structure (e.g., Badre, 2008), and that further, the degree to which this segregation occurs facilitates learning.

The working memory contexts envisioned by Reynolds and O'Reilly also address another central problem of simple RL: how these mechanisms determine the "state" relevant to action choice. Gureckis and Love (2009) are among

0010-0277/$ - see front matter © 2009 Elsevier B.V.. All rights reserved.
doi:10.1016/j.cognition.2009.09.005

the first to examine this issue directly by manipulating the state information available to subjects and studying how this affects RL. In particular, they study the behavior of subjects performing a set of choice tasks which are identical in the underlying action-reward contingencies but differ in terms of the cues signaling the current state of the game. Using computational modeling, they demonstrate that the vast changes in subjects' learning between conditions can be understood in terms of an RL model that adopts different internal state representations according to the provided cues.

Two other articles use ideas from RL to develop formal, computational theories of higher level ideas that had previously been treated more abstractly. First, Huys and Dayan (2009) consider how subjects might learn, at a higher level, about the general statistical properties of an environment, and then transfer this knowledge (as a "prior" in their Bayesian formulation) to guide RL in subsequently encountered environments. They propose such a mechanism can explain the phenomenon of learned helplessness – often used as a model of depression – in which an animal is subject to inescapable punishments and then fails to learn normally to avoid punishment or obtain reward in subsequent environments. This work rationalizes the seemingly maladaptive behavior in the animal models, and at the same time formalizes the idea of beliefs about "control" as a key clinical aspect of depression in humans.

Using related Bayesian theoretical models Baker, Tenenbaum, and Saxe (2009) offer a formal, quantitative notion of the notoriously subtle concept of theory of mind, in terms of determining the goals of others by inferring the hidden reasons for their (otherwise ambiguous) actions. Here, RL is not applied to the inference process itself; rather, the perceiver assumes that the other agent is using RL to plan action sequences in an effort to maximize its own reward. This assumption is then inverted to infer the goals behind the other agent's actions. The authors report three experiments in which subjects' judgments about agents' goals are well explained by this formalism.

Finally Chater (2009) challenges the RL enterprise as a means for understanding particular cognitive and neural mechanisms. He argues that apparent evidence for RL should not be taken as suggesting that the brain literally implements a special class of learning mechanism, but instead reflects the operation of a more general rational solution as specialized to the sort of task studied in the laboratory. Chater contrasts this view to a number of two-system accounts that envision behavior as reflecting conflict between multiple neural mechanisms, including both simple RL mechanisms and more general cognitive ones (e.g., conflict between Pavlovian and instrumental systems (Dayan, Niv, Seymour, & Daw, 2006), or habitual and goal-directed systems (Balleine & Dickinson, 1991).

Whether Chater is correct that a more general understanding of higher level cognitive processes will displace the RL view, or (as most of the other articles in the issue envision) extend it, the encounter between higher level cognition and lower-level systems neuroscience shows promise and some success already. It has taken cognitive neuroscience out of the desert of simple associationism to the rich soil offered by theories of higher level cognition. Conversely, it has grounded more abstract ideas about cognition in biologically-constrained and computationally more detailed formalisms of RL, which have proven useful for understanding not only cognition itself, but also how it is altered by disease, genetics, and pharmacological manipulations. Thus, RL is shaping up as a showcase example for the notion that tangible progress can be achieved by the union of the cognitive revolution and the neuroscientific revolution, when the link between them is informed by computational considerations.

References

Badre, D. (2008). Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends in cognitive sciences, 12*(5), 193–200.

Baker, C. L., Tenenbaum, J. B., & Saxe, R. B. (2009). Action understanding as inverse planning. *Cognition, 113*(3), 329–349.

Balleine, B., & Dickinson, A. (1991). Instrumental performance following reinforcer devaluation depends upon incentive learning. *The Quarterly Journal of Experimental Psychology Section B, 43*(3), 279–296.

Botvinick, M., Niv, Y., & Barto, A. C. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition, 113*(3), 262–280.

Chater, N. (2009). Rational and mechanistic perspectives on reinforcement learning. *Cognition, 113*(3), 350–364.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience, 8*(12), 1704–1711.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature, 441*(7095), 876–879.

Dayan, P., Niv, Y., Seymour, B., & Daw, N. D. (2006). The misbehavior of value and the discipline of the will. *Neural networks: The Official Journal of the International Neural Network Society, 19*(8), 1153–1160.

Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By carrot or by stick: Cognitive reinforcement learning in Parkinsonism. *Science, 306*, 1940–1943.

Gureckis, T. M., & Love, B. C. (2009). Short-term gains, long-term pains: How cues about state aid learning in dynamic environments. *Cognition, 113*(3), 293–313.

Huys, Q. J., & Dayan, P. (2009). A Bayesian formulation of behavioral control. *Cognition, 113*(3), 314–328.

McClure, S. M., Daw, N. D., & Montague, P. R. (2003). A computational substrate for incentive salience. *Trends in Neurosciences, 26*, 423–428.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1997). A framework for mesencephalic dopamine systems based on predictive hebbian learning. *The Journal of Neuroscience, 16*, 1936–1947.

Moustafa, A. A., Cohen, M. X., Sherman, S. J., & Frank, M. J. (2008). A role for dopamine in temporal decision making and reward maximization in Parkinsonism. *Journal of Neuroscience, 28*(47), 12294–12304.

Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007). Tonic dopamine: Opportunity costs and the control of response vigor. *Psychopharmacology, 191*(3), 507–520.

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron, 38*, 329–337.

O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation, 18*, 283–328.

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature, 442*(7106), 1042–1045.

Redish, A. D. (2004). Neuroscience. Addiction as a computational process gone awry. *Science, 306*(5703), 1944–1946.

Reynolds, J. R., & O'Reilly, R. C. (2009). Developing pfc representations using reinforcement learning. *Cognition, 113*(3), 281–292.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Sutton, R., Precup, D., & Singh, S. (1999). Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence, 112*(1–2), 181–211.

Thorndike, E. L. (1911). *Animal intelligence: Experimental studies*. MacMillan Press.

Nathaniel D. Daw
Center for Neural Science and Department of Psychology,
New York University, 4 Washington Place, Room 809,
NY 10003, New York, USA
Tel.: +1 212 998 2104
E-mail address: nathaniel.daw@nyu.edu (N.D. Daw)

Michael J. Frank
Department of Cognitive and Linguistic Science and Psychology,
Brown Institute for Brain Science, Brown University,
190 Thayer St. Providence, RI 02912-1978, USA
Tel.: +1 401 863 6872; fax: +1 401 863 2255
E-mail address: michael_frank@brown.edu (M.J. Frank)