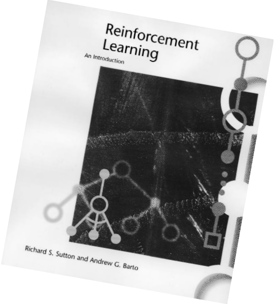
Reinforcement Learning: An Introduction

by Sutton, R.S. and Barto, A.G., MIT Press (1998). £31.95 (xi + 322 pages)
ISBN 0 262 19398 1



Reinforcement is a term with different meanings for different people. In the psychological lexicon, it conjures up the ominous images of Pavlov, Thorndike, Skinner and their intellectual brethren. Although these behaviorists helped to operationalize experimental psychology, their insistence on the non-existence of internal mental states provided a roadblock to modern cognitive science. In fact, the rise of cognitive science since the 1950s could be viewed as a rejection of the stultifying behaviorist views that declared the mind to be a vacuous construct. With such a salvo on behaviorists, what is a review of a book on reinforcement learning doing in *Trends in Cognitive Sciences*? The short answer is that reinforcement, in the context of the new book by Sutton and Barto, is not what it seems. 'Reinforcement learning is learning what to do – how to map situations to actions – so as to maximize a numerical reward signal', according to the introduction of the book.

The primary aim here is to cast learning as a problem involving agents that interact with an environment, sense their state and the state of the environment, and choose actions based on these interactions (which sounds very much like a bug or a rat moving about in some territory in search of food or mates). The twist in reinforcement learning is that the agent comes pre-equipped with goals that it seeks to satisfy. These goals are embodied in the influence of a 'numerical reward signal' on the way that the agent chooses actions, categorizes its sensations and changes its internal model of the environment. Despite the obvious connection of these terms to behavioral psychology, some of the more impressive applications of reinforcement learning have been in computer science and engineering applications. For example, Tesauro's TD-gammon, a reinforcement-learning system, is now one of the best backgammon players in the world[1].

Reinforcement learning typically divides a problem into four parts: (1) a policy; (2) a reward function; (3) a value function; and (4) an internal model of the environment. In this context, a policy is similar to an association in psychological terms; it maps states to actions (behavioral choices). One interesting part of reinforcement-learning problems is the complimentary concepts of *goals* and *evaluation*. A reward function provides a numerical evaluation of a state, and therefore embodies the agent's definition of what is immediately 'good' and what is immediately 'bad'. By contrast, value functions evaluate a state in terms of the total amount of reward an agent can expect from that state into the distant future; that is. they represent long-term evaluations. In this sense, value functions represent something more akin to judgements on the likely payoffs that will follow the current state.

Some of the most exciting work in reinforcement learning has taken place in the past 10 years with the discovery of several mathematical connections between separate methods for solving reinforcement-learning problems. These connections showed that apparently disparate mathematical techniques for solving reinforcement-learning problems were related in fundamental ways. This book provides the best historical details of these mathematical connections found anywhere, and frames clearly the ideas underlying this history.

What is the direct conceptual payoff of reinforcement learning for cognitive science? The descriptions so far show that reinforcement-learning problems could arise in a number of settings. Why should we expect this framework to enrich our understanding of cognition or the connection of the brain to cognition? I think that the direct benefit is twofold. The first benefit is that the lexicon of reinforcement learning is appropriate for describing the problems faced by mobile creatures in a complex, stochastic environment, in which the evaluation of a sequence of decisions might be significantly delayed. Consonant with the appropriateness of the lexicon, a number of modern efforts have successfully used reinforcement learning to describe biological systems related to motor learning in the cerebellum, and reward learning by dopaminergic systems[2,3].

The second benefit is the emphasis that reinforcement learning places on representation. This emphasis emerges from the two serious complaints about reinforcement learning as a framework for artificial intelligence or models of brain function: (1) speed, and (2) the size of the state space[4]. For even modest problems, the state space can be huge (e.g. for backgammon, the state space is ~10^{20} states). If any sizeable fraction of this state space must be explored for a reinforcement-learning system to converge to an answer, then one might have to wait an unacceptably long time for a suitable answer to emerge. These problems were a likely source of discouragement for early work in reinforcement learning. However, more modern work has shown that if careful consideration is given to the representations of states or actions, then reinforcement-learning systems can be a powerful way of learning certain problems.

The present book is an excellent entry point for someone who wants to understand intuitively the ideas of reinforcement learning and the general connection between its parts. It is not, however, a mathematical 'how-to' book, replete with proofs and pointers to unsolved problems in the field (as are, for example, Refs 3,5).

The end of each chapter contains a scholarly set of biographical and historical notes. These sections are particularly pleasing because they provide an easy-to-read review of the history of papers and ideas that contributed to the chapter in question. The authors go above and beyond the call of duty in these sections by providing their own perspective on how and why subfields developed in particular ways. Their effort is useful because this kind of perspective is very difficult to come by, yet it often provides conceptual insights by demonstrating which paths of investigation resulted from historical accident or the prevailing biases of the day. Furthermore, these sections are accessible to the casual peruser as well as the serious student seeking a historical record of publication on the subject.

Anyone interested in the internal representation of goals should read this book. In particular, the success of TD-gammon, and the connection of reinforcement-learning algorithms to the function of identified neural systems, suggests that reinforcement learning might have a lot more yet to say about cognition. That possibility awaits future evaluation.

P. Read Montague
Division of Neuroscience, Baylor College of Medicine, Houston, TX 77030, USA.
tel: +1 713 798 3134
fax: +1 713 798 3130
e-mail: read@bohr.neusc.bcm.tmc.edu

References

1 Tesauro, G.J. (1994) TD-Gammon, a self-teaching backgammon program, achieves master-level play *Neural Comput.* 6, 215–219

2 Schultz, W., Dayan, P. and Montague, P.R. (1997) A neural substrate of prediction and reward *Science* 275, 1593–1599

3 Bertsekas, D.P. and Tsitsiklis, J.N. (1996) *Neuro-Dynamic Programming*, Athena Scientific

4 Kaebling, L.P., Littman, M.L. and Moore, A.W. (1996) Reinforcement learning: a survey *J. Artif. Intell. Res.* 4, 237–285

5 Bellman, R. (1957) *Dynamic Programming*, Princeton University Press