Abstract: Learning from reinforcements is a promising approach for creating intelligent agents. However, reinforcement learning usually requires a large number of training episodes. We present and evaluate a design that addresses this shortcoming by allowing a connectionist Q-learner to accept advice given, at any time and in a natural manner, by an external observer. In our approach, the advice-giver watches the learner and occasionally makes suggestions, expressed as instructions in a simple imperative programming language. Based on techniques from knowledge-based neural networks, we insert these programs directly into the agent’s utility function. Subsequent reinforcement learning further integrates and refines the advice. We present empirical evidence that investigates several aspects of our approach and show that, given good advice, a learner can achieve statistically significant gains in expected reward. A second experiment shows that advice improves the expected reward regardless of the stage of training at which it is given, while another study demonstrates that subsequent advice can result in further gains in reward. Finally, we present experimental results that indicate our method is more powerful than a naive technique for making use of advice.

13.1 PREFACE

The remainder of this chapter is a verbatim reprint of an article that appeared in a special issue of *Machine Learning* on reinforcement learning (Kaelbling, 1996). The theme of this paper is on “advice-giving” by a human teacher to a connectionist re-
inforcement learner, rather than this book’s topic of “learning to learn.” However, there is a relationship between these two tasks. In “learning to learn,” knowledge is transferred from a previous learner to a future learner, while in our work knowledge is transferred from a human teacher to the current learner.

In both approaches to knowledge transfer, the goal is to help a learner to solve the current learning task by providing knowledge from an “outside” source. In the “learning to learn” framework discussed in this book’s introduction, the outside source is a previous instantiation of the learner, while in our method the outside source is a human agent. These two approaches are really complementary rather than competing. For example, if a previously trained learner was available, one could apply a knowledge-extraction method (Craven & Shavlik, 1996), (Towell & Shavlik, 1993) to the learner and produce rules in our advice language. These rules could then be provided to the current learner as if a human teacher had articulated them. In effect, our advice language would be used as a means to allow learners, machine or human, to communicate previously learned knowledge. Thus our work could be combined with a knowledge-extraction method to produce a “learning to learn” algorithm. In fact, this is an area of future research for our work.

The decision of which approach to knowledge transfer to follow is really dependent on what sources of prior knowledge are available – if a human teacher is available our approach can produce benefits and if a previously trained learner is available a different approach might be more effective. In a general learner it would probably be beneficial to allow a learner to learn from both sources of knowledge – both experience on previous tasks and information provided by teachers, since humans clearly learn from both sources of previous knowledge.

In our work, the focus is on a learner solving a single task and the teacher’s advice is aimed at improving the performance of the learner. As will be seen in this chapter, by inserting the advice provided by the human teacher into the machine learner, we actually cause the learner’s underlying model (in this case, a neural network) to be altered (by changing the network topology and weights). So, the learner after receiving the advice is now trying to find a solution to a different problem – optimizing the network weights in the altered network. In other words, our system is using the human-provided advice to alter its learning task, which is another perspective on the phrase “learning [what] to learn.”

13.2 INTRODUCTION

A successful and increasingly popular method for creating intelligent agents is to have them learn from reinforcements (Barto, Sutton, & Watkins, 1990), (Lin, 1992), (Mahadevan & Connell, 1992), (Tesauro, 1992), (Watkins, 1995). However, these approaches suffer from their need for large numbers of training episodes. Several methods for speeding up reinforcement learning have been proposed; one promising ap-