# Schemes for Learning and Behaviour: A New Expectancy Model

**Ph.D. Thesis**


## Christopher Mark Witkowski


**February 1997**


**Department of Computer Science**
**Queen Mary Westfield College**



**University of London**

# Abstract

This thesis presents a novel form of learning by reinforcement. Existing reinforcement learning algorithms rely on the provision of external reward signals to drive the learning algorithm. This new algorithm relies on reinforcing signals generated internally within the algorithm. The algorithm, SRS/E, described here generates expectancies (μ-hypotheses), each of which gives rise to a specific prediction when the conditions relevant to the expectancy are encountered (the μ-experiment). The algorithm subsequently tests these predictions against actual events and so generates reinforcement signals to corroborate or reject individual expectancies. This procedure allows for self-contained, completely unsupervised learning to an extent not possible with previous reinforcement procedures. The SRS/E algorithm is derived from a number of postulates that constitute a new Dynamic Expectancy Model developed in this thesis.

In contrast to the static policy map generated by existing *Q*-learning based reinforcement algorithms, which limit learning to one goal, the SRS/E algorithm generates a Dynamic Policy Map (DPM) from learned expectancies whenever a new goal is selected by the system. This new approach retains the advantages of reactivity to the environment inherent in existing reinforcement algorithms, while substantially increasing the system's flexibility in responding to varying circumstances and requirements. Also in contrast to previous reinforcement systems, goals may be selected arbitrarily and are not limited to those which were associated with reward during the learning steps. This new method allows multiple goals to be pursued either simultaneously or sequentially.

The single SRS/E implementation has been compared directly to the published results from of a family of reinforcement based algorithms, Dyna-PI, Dyna-Q and Dyna-Q+ (Sutton, 1990), themselves extensions to the groundbreaking *Q*-learning algorithm (Watkins, 1989). Under equivalent "ideal learning conditions" the SRS/E algorithm was found to outperform the equivalent Dyna reinforcement program to learn a simple maze task by a factor of some 40:1. The SRS/E learning algorithm was also found to be robust when tested under controlled "noise" conditions. SRS/E was also compared directly to Sutton's Dyna-Q+ algorithm on a range of alternative path and route blocking tasks and was found to offer a similar performance, but SRS/E employs a "biologically plausible" extinction mechanism, mirroring findings from animal behaviour research.

Finally SRS/E was tested with experimental designs for "latent learning" and "place learning", drawn directly from animal learning research. Both are regarded as presenting severe challenges to conventional reinforcement learning theories. SRS/E performs well on both tasks, and in a manner consistent with findings from animal experiments.

# Table of Contents

# List of Figures and Tables