

MODELLING MOTIVATION FOR EXPERIENCE-BASED ATTENTION FOCUS IN REINFORCEMENT LEARNING



A thesis submitted in fulfilment of the requirements for the
degree of Doctor of Philosophy in the School of Information Technologies at
The University of Sydney

Kathryn Elizabeth Merrick
August, 2007

© Copyright by Kathryn Merrick 2007
All Rights Reserved

I hereby declare that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other university or institution.

Kathryn E. Merrick

Sydney
August, 2007

Acknowledgements

A PhD thesis is rarely the work of just one person. This thesis has benefited from the input and advice of members of the Key Centre for Design Computing and Cognition at the University of Sydney. My supervisor, Mary Lou Maher, has been both enthusiastic and supportive throughout. Mary introduced me to John Gero who, in turn, showed me the work of Rob Saunders, which has influenced the models developed in this thesis. Rob has provided support and guidance during the writing process. I would also like to thank Professor Peter Eades from the National ICT Australia for his help and advice over the past three years. This research was supported by an Australian Postgraduate Award and a National ICT Australia PhD scholarship.

Abstract

Computational models of motivation are software reasoning processes designed to direct, activate or organise the behaviour of artificial agents. Models of motivation inspired by psychological motivation theories permit the design of agents with a key reasoning characteristic of natural systems: experience-based attention focus. The ability to focus attention is critical for agent behaviour in complex or dynamic environments where only small amounts of available information is relevant at a particular time. Furthermore, experience-based attention focus enables adaptive behaviour that focuses on different tasks at different times in response to an agent's experiences in its environment. This thesis is concerned with the synthesis of motivation and reinforcement learning in artificial agents. This extends reinforcement learning to adaptive, multi-task learning in complex, dynamic environments.

Reinforcement learning algorithms are computational approaches to learning characterised by the use of reward or punishment to direct learning. The focus of much existing reinforcement learning research has been on the design of the learning component. In contrast, the focus of this thesis is on the design of computational models of motivation as approaches to the reinforcement component that generates reward or punishment. The primary aim of this thesis is to develop computational models of motivation that extend reinforcement learning with three key aspects of attention focus: rhythmic behavioural cycles, adaptive behaviour and multi-task learning in complex, dynamic environments. This is achieved by representing such environments using context-free grammars, modelling maintenance tasks as observations of these environments and modelling achievement tasks as events in these environments. Motivation is modelled by processes for task selection, the computation of experience-based reward signals for different tasks and arbitration between reward signals to produce a motivation signal. Two specific models of motivation based on the experience-oriented psychological concepts of interest and competence are designed within this framework. The first models motivation as a function of environmental experiences while the second models motivation as an introspective process.

This thesis synthesises motivation and reinforcement learning as motivated reinforcement learning agents. Three models of motivated reinforcement learning are presented to explore the combination of motivation with three existing reinforcement learning components. The first model combines motivation with flat reinforcement learning for highly adaptive learning of behaviours for performing multiple tasks. The second model facilitates the recall of learned behaviours by combining motivation with multi-option reinforcement learning. In the third model, motivation is combined with an hierarchical reinforcement learning

component to allow both the recall of learned behaviours and the reuse of these behaviours as abstract actions for future learning.

Because motivated reinforcement learning agents have capabilities beyond those of existing reinforcement learning approaches, new techniques are required to measure their performance. The secondary aim of this thesis is to develop metrics for measuring the performance of different computational models of motivation with respect to the adaptive, multi-task learning they motivate. This is achieved by analysing the behaviour of motivated reinforcement learning agents incorporating different motivation functions with different learning components. Two new metrics are introduced that evaluate the behaviour learned by motivated reinforcement learning agents in terms of the variety of tasks learned and the complexity of those tasks.

Persistent, multi-player computer game worlds are used as the primary example of complex, dynamic environments in this thesis. Motivated reinforcement learning agents are applied to control the non-player characters in games. Simulated game environments are used for evaluating and comparing motivated reinforcement learning agents using different motivation and learning components. The performance and scalability of these agents are analysed in a series of empirical studies in dynamic environments and environments of progressively increasing complexity. Game environments simulating two types of complexity increase are studied: environments with increasing numbers of potential learning tasks and environments with learning tasks that require behavioural cycles comprising more actions.

A number of key conclusions can be drawn from the empirical studies, concerning both different computational models of motivation and their combination with different reinforcement learning components. Experimental results confirm that rhythmic behavioural cycles, adaptive behaviour and multi-task learning can be achieved using computational models of motivation as an experience-based reward signal for reinforcement learning. In dynamic environments, motivated reinforcement learning agents incorporating introspective competence motivation adapt more rapidly to change than agents motivated by interest alone. Agents incorporating competence motivation also scale to environments of greater complexity than agents motivated by interest alone. Motivated reinforcement learning agents combining motivation with flat reinforcement learning are the most adaptive in dynamic environments and exhibit scalable behavioural variety and complexity as the number of potential learning tasks is increased. However, when tasks require behavioural cycles comprising more actions, motivated reinforcement learning agents using a multi-option learning component exhibit greater scalability. Motivated multi-option reinforcement

learning also provides a more scalable approach to recall than motivated hierarchical reinforcement learning.

In summary, this thesis makes contributions in two key areas. Computational models of motivation and motivated reinforcement learning extend reinforcement learning to adaptive, multi-task learning in complex, dynamic environments. Motivated reinforcement learning agents allow the design of non-player characters for computer games that can progressively adapt their behaviour in response to changes in their environment.

Table of Contents

Acknowledgements	iv
Abstract	v
Index of Figures	xii
Index of Tables	xix
List of Abbreviations	xx
1. Introduction	1
1.1. RESEARCH OBJECTIVES	7
1.2. RESEARCH CONTRIBUTIONS AND SIGNIFICANCE	8
1.3. METHODOLOGY	9
1.4. THESIS OVERVIEW	10
1.4.1. Background and Related Work	10
1.4.2. Modelling Motivation for Experience-Based Attention Focus in Reinforcement Learning.....	10
1.4.3. Performance Metrics for Motivated Reinforcement Learning	10
1.4.4. Motivated Reinforcement Learning in a Simulated Game World.....	11
1.4.5. Empirical Analysis of Motivated Reinforcement Learning in Complex, Dynamic Environments	11
1.4.6. Motivated Reinforcement Learning in an Open-Ended Virtual World	11
1.4.7. Conclusion.....	12
2. Background and Related Work	13
2.1. THEORIES OF MOTIVATION FOR NATURAL AND ARTIFICIAL SYSTEMS	14
2.1.1. Biological Theories and Models of Motivation.....	16
2.1.2. Cognitive Theories and Models of Motivation.....	22
2.1.3. Social Theories and Models of Motivation	31
2.1.4. Combined Motivation Theories.....	33
2.1.5. Discussion	35
2.2. REINFORCEMENT LEARNING MODELS	36
2.2.1. Reinforcement Learning	37
2.2.2. Reinforcement Learning in Partially Observable Environments	42
2.2.3. Reinforcement Learning with Function Approximation	43
2.2.4. Hierarchical Reinforcement Learning	44
2.2.5. Motivated Reinforcement Learning.....	47
2.2.6. Discussion	56

2.3.	CONCLUSION	58
3.	Modelling Motivation for Experience-Based Attention Focus in Reinforcement Learning.....	59
3.1.	MOTIVATION AND REINFORCEMENT LEARNING IN COMPLEX, DYNAMIC ENVIRONMENTS	60
3.1.1.	Motivation	62
3.1.2.	Behavioural Cycles.....	63
3.1.3.	Motivated Reinforcement Learning.....	64
3.2.	MODELLING MOTIVATION FOR ATTENTION FOCUS IN REINFORCEMENT LEARNING	65
3.2.1.	Observations	66
3.2.2.	Events	67
3.2.3.	Tasks and Task Selection	69
3.2.4.	Experience-Based Reward Functions using Cognitive Motivation Theories	72
3.2.5.	Arbitration Functions.....	76
3.2.6.	Summary and Discussion	77
3.3.	ALGORITHMS FOR MOTIVATED REINFORCEMENT LEARNING.....	79
3.3.1.	Motivated Flat Reinforcement Learning	80
3.3.2.	Motivated Multi-Option Reinforcement Learning	82
3.3.3.	Motivated Hierarchical Reinforcement Learning	85
3.3.4.	Discussion	88
3.4.	CONCLUSION	88
4.	Performance Metrics for Motivated Reinforcement Learning.....	89
4.1.	EXISTING PERFORMANCE METRICS FOR REINFORCEMENT LEARNING.....	90
4.1.1.	Models of Optimality and Performance Metrics for Reinforcement Learning	90
4.1.2.	Metrics for Characterising the Motivation Function or Behaviour of Motivated Reinforcement Learning Agents.....	94
4.1.3.	Discussion	96
4.2.	NEW PERFORMANCE METRICS FOR ADAPTIVE, MULTI-TASK, MOTIVATED REINFORCEMENT LEARNING	97
4.2.1.	Statistical Model for Identifying Learned Tasks	98
4.2.2.	Behavioural Variety.....	98
4.2.3.	Behavioural Complexity.....	100
4.2.4.	Alternatives to Motivation in Reinforcement Learning.....	101
4.2.5.	Discussion	103
4.3.	CONCLUSION	103

5. Motivated Reinforcement Learning in a Simulated Game World	104
5.1. RELATED WORK.....	105
5.1.1. Character Roles in MMORPGs	106
5.1.2. Existing Artificial Intelligence Techniques for NPCs in MMORPGs.....	107
5.2. MODELS OF MOTIVATION FOR SUPPORT CHARACTERS IN MMORPGS ..	109
5.2.1. Modelling Motivation as Interesting Events	109
5.2.2. Modelling Motivation Using Interest and Competence.....	113
5.3. EXPERIMENT 1: MOTIVATED REINFORCEMENT LEARNING FOR SUPPORT CHARACTERS IN MMORPGS	115
5.3.1. Experimental Setup: A Simulated Game Environment	115
5.3.2. Results and Discussion.....	118
5.4. CONCLUSION	126
6. Scalability of Motivated Reinforcement Learning in Complex and Dynamic Environments	128
6.1. EXPERIMENT 2: INCREASING THE NUMBER OF POTENTIAL LEARNING TASKS	129
6.1.1. Experimental Setup	129
6.1.2. Results and Discussion.....	131
6.2. EXPERIMENT 3: TASKS OF INCREASING COMPLEXITY	138
6.2.1. Experimental Setup	138
6.2.2. Results and Discussion.....	138
6.3. EXPERIMENT 4: ENVIRONMENTS THAT CHANGE WHILE THE AGENT IS LEARNING.....	145
6.3.1. Experimental Setup	145
6.3.2. Results and Discussion.....	147
6.4. CONCLUSION	153
7. Motivated Reinforcement Learning in an Open-Ended Virtual World	155
7.1. RELATED WORK.....	156
7.1.1. Open-Ended Virtual Worlds.....	156
7.1.2. Simulation Games	158
7.2. EXPERIMENT 5: MOTIVATED REINFORCEMENT LEARNING IN OPEN- ENDED SIMULATION GAMES.....	159
7.2.1. Game Design	160
7.2.2. Character Design	160
7.2.3. Analysis of Character Behaviour in Response to Game Play Sequences.....	163
7.2.4. Discussion	169
7.3. CONCLUSION	170

8. Conclusion	172
8.1. SUMMARY OF RESEARCH CONTRIBUTIONS AND CONCLUSIONS	173
8.1.1. Synthesis of Motivation Theory and Reinforcement Learning	173
8.1.2. A Model of Motivation for Experience-Based Attention Focus in Reinforcement Learning	173
8.1.3. Performance Metrics for Motivated Reinforcement Learning	174
8.1.4. An Empirical Analysis of Motivated Reinforcement Learning in Complex, Dynamic Environments	174
8.1.5. Non-Player Characters for Open-Ended Virtual Worlds	175
8.2. LIMITATIONS AND FUTURE WORK	175
8.2.1. Motivation in other Reinforcement Learning Settings	175
8.2.2. Scalability of Motivated Reinforcement Learning	176
8.2.3. Alternative Approaches to Motivation in Reinforcement Learning	177
8.2.4. Motivation in other Machine Learning Settings	178
8.2.5. Additional Metrics for Motivated Reinforcement Learning	179
8.3. CONCLUDING REMARKS	179
Glossary	180
Appendix A – Details of the Experimental Method	183
Appendix B – Additional Results from the Aussie Outback Challenge	186
Appendix C – Publications Related to this Research	191
References	198

Index of Figures

Figure 1.1. The cliff walking task (Sutton and Barto, 2000). Reinforcement learning uses a reward signal (left) to direct learning of a policy (right) for every state. Blue arrows are resets that occur outside the world model.	2
Figure 1.2. The cliff-walking task situated in an environment with multiple tasks that may change unpredictably over time.	3
Figure 1.3. (a) Existing reinforcement learning approaches solve an isolated task for all states. (b) The motivated reinforcement learning approaches in this thesis focuses attention using motivation to support the emergence of behavioural cycles for life-long, adaptive, multi-task learning.	6
Figure 2.1. Motivational state theory describes hunger more accurately than drive theory by using multiple motivational variables (McFarland, 1995).	19
Figure 2.2. Arousal results from the joint action of positive and negative reward for a stimulus. (Berlyne, 1970; Berlyne, 1971).	20
Figure 2.3. Operant theory describes the relationship between reward and behaviour.	25
Figure 2.4. Incentive for behavioural response in individuals with (a) differing motivational strength to approach success and (b) differing motivational strength to avoid failure.	28
Figure 2.5. The Naive Analysis of Action Theory (Heider, 1958) divides motivational forces into two classes: personal forces and environmental forces.	29
Figure 2.6. Maslow’s hierarchy of needs (Maslow, 1954) synthesises different motivation theories in an hierarchy.	34
Figure 2.7. Existence, Relatedness, Growth theory (Alderfer, 1972) synthesises different motivation theories without ordering.	35
Figure 2.8. The reinforcement learning process takes states and rewards as input, updates a policy stored in memory and outputs an action.	38
Figure 2.9. Episodic temporal difference reinforcement learning algorithms process experiences in distinct episodes.	41
Figure 2.10. Continuing task, temporal difference reinforcement learning algorithms process experiences as a single, infinite trajectory.	42
Figure 2.11. The reinforcement learning process for partially observable environments. Sensed states differ from the actual world state.	43
Figure 2.12. The hierarchical reinforcement learning process takes states and rewards as input, creates behavioural options, updates an hierarchical policy stored in memory and outputs an action.	45
Figure 2.13. Category (I) motivated reinforcement learning models: (a) MRL(I) and (b) MHRL(I) extend reinforcement learning and hierarchical reinforcement learning by introducing a motivation signal in addition to the reward signal.	48

Figure 2.14. Category (II) motivated reinforcement learning models extend reinforcement learning by using a motivation signal instead of the reward signal.	49
Figure 2.15. Simsek and Barto (2006) model MHRL(I) agents using a task value function and a behaviour value function.	52
Figure 2.16. (a) Reinforcement learning is extended by Singh et al. (2005) by splitting the environment into internal and external components to create (b) category (I) reinforcement learning agents.....	52
Figure 3.1. An agent in an environment.....	60
Figure 3.2. Behavioural cycles of complexity one, two, three and n.....	64
Figure 3.3. The motivated reinforcement learning algorithm uses a continuing task flow of control. Motivation is computed as an experience-based reward signal.....	65
Figure 3.4. Difference between a world state, a sensed state and an observation.....	66
Figure 3.5. The novelty of a stimulus decreases with repeated exposure to the stimulus (habituation). Novelty increases when the stimulus is removed (recovery).	73
Figure 3.6. The Wundt curve is the difference between positive and negative reward functions. It peaks at a moderate degree of novelty.....	74
Figure 3.7. Learning error and competence have an inverse relationship.....	76
Figure 3.8. Algorithmic description of motivation for attention focus in reinforcement learning.	78
Figure 3.9. Diagrammatic representation of motivation for attention focus.....	79
Figure 3.10. The motivated Q-learning algorithm.	80
Figure 3.11. The motivated SARSA algorithm.....	80
Figure 3.12. Comparison of (a) flat reinforcement learning agents and (b) motivated flat reinforcement learning agents. Flat reinforcement learning agents take a reward signal from the environment, but motivated flat reinforcement learning agents incorporate a motivation process to compute an experience-based reward signal. (Saunders and Gero, 2002)	81
Figure 3.13. The motivated, multi-option Q-learning algorithm.	83
Figure 3.14. Comparison of (a) motivated flat reinforcement learning agents and (b) motivated, multi-option reinforcement learning agents. Motivated, multi-option reinforcement learning agents incorporate a reflex process to create, remove and trigger behavioural options.	85
Figure 3.15. The motivated hierarchical reinforcement learning agent model.	86
Figure 3.16. The motivated hierarchical Q-learning algorithm.	87
Figure 4.1. The performance of different reinforcement learning algorithms can be measured as the reward gained in each learning episode (Sutton and Barto, 2000).	92

Figure 4.2. Reinforcement learning is used as a baseline for measuring two hierarchical reinforcement learning algorithms. Performance is measured as the number of steps until the task is complete in each learning episode (Hengst, 2002).	92
Figure 4.3. (a) Singh et al. (2005) measure learning quality as the average number of actions to salient events. (b) Standard reinforcement learning is used as a baseline.	93
Figure 4.4. Schmidhuber (1997) measures learning quality as the number of occurrences of rewarded 'goal' states. Standard (plain) reinforcement learning is used as a baseline.	93
Figure 4.5. Kaplan and Oudeyer (2003) characterise their motivation function in terms of the evolution of its motivational variables (a) predictability, (b) familiarity, (c) head stability and (d) light stability.	95
Figure 4.6. Kaplan and Oudeyer (2003) characterise behaviour in terms of domain specific physical attributes such a head pan position and perceived light position.	95
Figure 4.7. Huang and Weng (2002) characterise (a) their motivation function in terms of the evolution of q-values in the learning component and (b) behaviour in terms of the frequency with which actions are performed.	96
Figure 4.8. Saunders and Gero (2001a) characterise their motivation function in terms of the evolution of novelty values.	96
Figure 4.9. Behavioural cycles of (a) complexity one for a maintenance task, (b) complexity n for n achievement tasks.	97
Figure 4.10. Multi-task learning can be visualised as instantaneous behavioural variety.	99
Figure 4.11. Adaptive, multi-task learning can be visualised as cumulative behavioural variety.	99
Figure 4.12. The scalability of multi-task learning can be visualised in terms of the instantaneous behavioural variety at a particular time in different environments. .	100
Figure 4.13. Multi-task learning can be visualised in terms of maximum behavioural complexity.	101
Figure 4.14. The scalability of multi-task learning can be visualised in terms of the maximum behavioural complexity attained in different environments.	101
Figure 5.1. Algorithmic description of motivation to achieve interesting events, a motivation function for support characters in MMORPGs.	110
Figure 5.2. Change in novelty with $\alpha = 1.05$ and (a) $\tau_1 = 3.3, \tau_2 = 14.3$ and (b) $\tau_1 = 9, \tau_2 = 27$...	111
Figure 5.3. Change in novelty with (a) $\tau_1 = 3.3, \tau_2 = 14.3, \alpha = 1.5$ and (b) $\tau_1 = 9, \tau_2 = 27, \alpha = 1.5$	111
Figure 5.4. Change in interest with (a) $\rho^+ = \rho^- = 5, F_{\min}^+ = 0.5$ and $F_{\min}^- = 1.5$ and (b) $\rho^+ = \rho^- = 30, F_{\min}^+ = 0.5$ and $F_{\min}^- = 1.5$	112
Figure 5.5. Change in interest with (a) $\rho^+ = \rho^- = 10, F_{\min}^+ = 0.1$ and $F_{\min}^- = 1.9$ and (b) $\rho^+ = \rho^- = 10, F_{\min}^+ = 0.9$ and $F_{\min}^- = 1.1$	112

Figure 5.6. Algorithmic description of motivation using interest and competence, a motivation function for support characters in massively multiplayer, online role-playing games.	113
Figure 5.7. Modelling interest with an aversion to low novelty (highly familiar tasks).	114
Figure 5.8. A game environment in Second Life.	115
Figure 5.9. A context-free grammar for sensed states in a game scenario in which agents control non-player characters. Agents have location sensors, inventory sensors and object sensors.	116
Figure 5.10. A context-free grammar for the action set in a game environment in which agents have location effectors, pick-up object effectors and use object effectors. .	116
Figure 5.11. Cumulative behavioural variety by motivated flat reinforcement learning agents using different motivation functions.	119
Figure 5.12. Maximum behavioural complexity achieved by motivated flat reinforcement learning agents after 50,000 time-steps.	120
Figure 5.13. Focus of attention by two individual motivated, flat reinforcement learning agents motivated to achieve interesting events over 50,000 time-steps. Agents that focus attention differently represent different game characters.	121
Figure 5.14. Cumulative behavioural variety by motivated, multi-option reinforcement learning agents using different motivation functions.	122
Figure 5.15. Average behavioural variety achieved by motivated, flat reinforcement learning (MFRL) and motivated, multi-option reinforcement learning (MMORL) agents after 50,000 time-steps.	122
Figure 5.16. Maximum behavioural complexity achieved by motivated flat reinforcement learning (MFRL) and motivated, multi-option reinforcement learning (MMORL) agents after 50,000 time-steps.	123
Figure 5.17. Cumulative behavioural variety by motivated hierarchical reinforcement learning agents using different motivation functions.	124
Figure 5.18. Cumulative behavioural variety by motivated flat reinforcement learning (MFRL), motivated, multi-option reinforcement learning (MMORL) and motivated hierarchical reinforcement learning (MHRL) agents motivated to achieve interesting events.	125
Figure 5.19. Average behavioural variety achieved by motivated flat reinforcement learning (MFRL), motivated multi-option reinforcement learning (MMORL) and motivated hierarchical reinforcement learning (MHRL) agents after 50,000 time-steps.	125
Figure 5.20. Maximum behavioural complexity achieved by motivated flat reinforcement learning (MFRL), motivated multi-option reinforcement learning (MMORL) and motivated hierarchical reinforcement learning (MHRL) agents after 50,000 time-steps.	126

Figure 6.1. State and action spaces for the additional MDP used to compose Environment 2 for Experiment 2.	130
Figure 6.2. State and action spaces for the additional MDP used to compose Environment 3 for Experiment 2.	130
Figure 6.3. State and action spaces for the additional MDP used to compose Environment 4 in Experiment 2.	131
Figure 6.4. State and action spaces for the additional MDP used to compose Environment 5 in Experiment 2.	131
Figure 6.5. Comparison of state and action set sizes for successive experimental environments in Experiment 2.	132
Figure 6.6. Average behavioural variety achieved by motivated flat reinforcement learning agents using different models of motivation. Each environment contains more tasks than the previous.	133
Figure 6.7. Maximum behavioural complexity achieved by motivated flat reinforcement learning agents using different models of motivation. Each environment contains more tasks than the previous environment.	134
Figure 6.8. Average behavioural variety achieved by motivated, multi-option reinforcement learning agents using different models of motivation. Each environment contains more tasks than the previous environment.	135
Figure 6.9. Maximum behavioural complexity achieved by motivated, multi-option reinforcement learning agents using different models of motivation. Each environment containing more tasks than the previous environment.	135
Figure 6.10. Average behavioural variety achieved by motivated hierarchical reinforcement learning agents using different models of motivation. Each environment contains more tasks than the previous environment.	137
Figure 6.11. Maximum behavioural complexity achieved by motivated hierarchical reinforcement learning agents using different models of motivation. Each environment containing more tasks than the previous environment.	137
Figure 6.12. State and action spaces of Environments 2-5 in Experiment 3.	139
Figure 6.13. Comparison of state and action set sizes for successive experimental environments in Experiment 3.	139
Figure 6.14. Maximum behavioural complexity achieved by motivated flat reinforcement learning agents using different models of motivation. Each environment contains tasks requiring progressively more actions.	141
Figure 6.15. Average behavioural variety achieved by motivated flat reinforcement learning agents using different models of motivation. Each environment contains tasks requiring progressively more actions.	141

Figure 6.16. Maximum behavioural complexity achieved by motivated, multi-option reinforcement learning agents using different models of motivation. Each environment contains tasks requiring progressively more actions.	142
Figure 6.17. Average behavioural variety achieved by motivated, multi-option reinforcement learning agents using different models of motivation. Each environment contains tasks requiring progressively more actions.	143
Figure 6.18. Maximum behavioural complexity achieved by motivated hierarchical reinforcement learning agents using different models of motivation. Each environment contains tasks requiring progressively more actions.	143
Figure 6.19. Average behavioural variety achieved by motivated hierarchical reinforcement learning agents using different models of motivation. Each environment contains tasks requiring progressively more actions.	144
Figure 6.20. Cumulative behavioural variety achieved in Environment 4 by motivated hierarchical reinforcement learning agents motivated to achieve interesting events.	144
Figure 6.21. State and action spaces of the environment in Experiment 4 after $t=50,000$.	146
Figure 6.22. Comparison of state and action space sizes before and after the environment changes in Experiment 4.	146
Figure 6.23. Cumulative behavioural variety by motivated flat reinforcement learning agents using different motivation functions.	148
Figure 6.24. Change in attention focus over time exhibited by a single agent motivated by interest and competence in a dynamic environment.	150
Figure 6.25. Cumulative behavioural variety by motivated, multi-option reinforcement learning agents using (a) motivation to achieve interesting events, (b) motivation by interest and competence and (c) the baseline reward function.	151
Figure 6.26. Cumulative behavioural variety by motivated hierarchical reinforcement learning agents using (a) motivation to achieve interesting events, (b) motivation by interest and competence and (c) the baseline reward function.	152
Figure 7.1. In the <i>Second Life</i> virtual world, complex designs can be created using a combination of primitives and uploaded textures.	157
Figure 7.2. In <i>Active Worlds</i> , complex designs can be created using basic shapes and textures.	157
Figure 7.3. Opening sequences of Aussie Outback Challenge, an open-ended, persistent simulation game set in the Second Life virtual world.	161
Figure 7.4. System architecture for designing agents as characters for Second Life.	162
Figure 7.5. In <i>Aussie Outback Challenge</i> , six uncommonly intelligent sheep are located in a paddock in <i>Second Life</i>	164

Figure 7.6. (a) A sheep follows an avatar away from the eelgrass. (b) Finite state automaton representation of the learned ‘following’ behavioural cycle..... 165

Figure 7.7. (a) A sheep explores the food machine. (b) A learned behavioural cycle for eating one food ball at a time. (c) A learned behavioural cycle for eating two food balls at a time. (d) A learned behavioural cycle for obtaining and eating food without moving. 166

Figure 7.8. Two toys built for the sheep: a colour changing screen and a shape changing wall..... 168

Figure 7.9. A hoop and target challenge designed for the sheep. A player attempts to ride a sheep while the sheep is learning..... 169

Index of Tables

Table 3.1. Observation functions that achieve different levels of attention focus at time t . .	67
Table 3.2. Difference functions that achieve different levels of attention focus at time t	68
Table 3.3. Event functions that achieve different levels of attention focus at time t	69
Table 3.4. Arbitration functions for producing a motivation signal from motivation values, when motivation values are computed by multiple computational models of motivation.	77
Table 3.5. Arbitration functions for producing a motivation signal from motivation values, when motivation values are computed for multiple motivating tasks.	77
Table 3.6. Potential experience trajectories as input for motivation functions in motivated reinforcement learning.	78
Table 3.7. Structures associated with behavioural options in motivated, multi-option reinforcement learning.	83
Table 3.8. Reflex rules used in motivated, multi-option reinforcement learning.	84
Table 3.9. Structures associated with behavioural options in motivated hierarchical reinforcement learning.	86
Table 5.1 Parameters and their values for motivated reinforcement learning agents motivated to achieve interesting events.	112
Table 5.2 Parameters and their values for motivated reinforcement learning agents motivated by interest and competence.	114
Table 5.3. Comparison of reward from interest based motivation, experience-based reward without motivation and a termination function.	123
Table 6.1. Scalability characteristics of motivated reinforcement learning approaches using different motivation and learning components.	154

List of Abbreviations

CFG	Context-Free Grammar
GPI	Generalised Policy Interaction
HRL	Hierarchical Reinforcement Learning
HSOM	Habituated Self-Organising Map
IHDR	Incremental Hierarchical Discriminant Regression
MDP	Markov Decision Process
MEU	Maximum Expected Utility
MFRL	Motivated Flat Reinforcement Learning
MHRL	Motivated Hierarchical Reinforcement Learning
MMORL	Motivated Multi-Option Reinforcement Learning
MMORPG	Massively Multiplayer, Online Role-Playing Game
MORL	Multi-Option Reinforcement Learning
MRL	Motivated Reinforcement Learning
MSL	Motivated Supervised Learning
MUL	Motivated Unsupervised Learning
NPC	Non-Player Character
POMDP	Partially Observable Markov Decision Process
RL	Reinforcement Learning
RPG	Role-Playing Game
SL	Supervised Learning
SMDP	Semi-Markov Decision Process
SOM	Self-Organising Map
TD	Temporal Difference
UL	Unsupervised Learning