

THE REFLEXIVE INSTRUCTOR WITH DELIBERATE APPRENTICE  
ARCHITECTURE

by

Alexander Ferworn

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Systems Design Engineering

Waterloo, Ontario, Canada, 1997

© Alexander Ferworn 1997

I hereby declare that I am the sole author of this thesis.

I authorize the University of Waterloo to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize the University of Waterloo to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

The University of Waterloo requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

## **ABSTRACT**

### **The Reflexive Instructor with Deliberate Apprentice Architecture**

A framework allowing a discourse in autonomy applied to autonomous mobile robots is developed based on human autonomy. This framework is extended to mobile robotics and is used to evaluate the level of autonomy in a novel approach for constructing autonomous controllers called the Reflexive Instructor (RI) with Deliberate Apprentice (DA) architecture. We claim that the RI/DA architecture supports the construction of first-order autonomous learning agents restricted only by their ability to interact with their environments.

The architecture uses simple reinforcement signals provided by the RI component to train the DA. The DA is responsible for providing control signals to the agent's actuators based on received sensor input. Like most reinforcement learning systems it is not likely to do this very well until it has learned to avoid collisions and obstacles in its environment. The RI provides a measure of safety in this respect as it is responsible for taking over control of the agent if the DA makes a mistake as well as providing an appropriate signal to the DA so it might learn from the mistake.

The RI/DA interaction is advantageous because it protects the vehicle from its own ignorance and helps accelerate learning in the DA.

## **ACKNOWLEDGEMENTS**

I would like to acknowledge the patient insistence, loving care and effective support provided to me by my wife Charmaine and the benevolent, no-nonsense direction of my supervisor and friend Deb Stacey. I also acknowledge the inspiration provided by my cat Dave.

## **DEDICATION**

This work is dedicated to my wife Charmaine--without her there would be no work, and to Charlotte, my daughter.

# TABLE OF CONTENTS

<b>ABSTRACT</b> .....	<b>IV</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>V</b>
<b>DEDICATION</b> .....	<b>VI</b>
<b>TABLE OF CONTENTS</b> .....	<b>VII</b>
<b>LIST OF TABLES</b> .....	<b>XI</b>
<b>LIST OF ILLUSTRATIONS</b> .....	<b>XII</b>
<b>1. INTRODUCTION</b> .....	<b>1</b>
1.1 INTRODUCTION.....	1
1.2 REFLEXES.....	1
1.3 MOBILE ROBOTICS.....	3
1.4 REINFORCEMENT LEARNING .....	4
1.5 SYNERGY .....	5
1.6 AUTONOMY .....	5
1.7 PROPOSITION .....	7
1.8 ORGANIZATION OF THIS THESIS .....	8
<b>2. TERMINOLOGY, THEORY AND LITERATURE EXAMINATION</b> .....	<b>9</b>
2.1 INTRODUCTION.....	9
2.2 MOBILE ROBOTICS.....	10
2.2.1 <i>Hierarchical Model Builders</i> .....	10
2.2.2 <i>Distributed Reactive Systems</i> .....	13
2.3 REFLEXES.....	16
2.4 LEARNING.....	17
2.4.1 <i>Reinforcement Learning</i> .....	19
2.5 ARTIFICIAL NEURAL SYSTEMS.....	20
2.5.1 <i>ANSs and Robotics</i> .....	22
2.5.2 <i>Rapid Reinforcement Learning Using Neural Networks</i> .....	22
2.6 AUTONOMY .....	24
2.7 CONCLUSION .....	30
<b>3. REFLEXIVE INSTRUCTOR WITH DELIBERATE APPRENTICE ARCHITECTURE</b> ....	<b>31</b>
3.1 INTRODUCTION.....	31
3.2 RIDA ARCHITECTURE DESCRIPTION .....	31
3.2.1 <i>Richer Reinforcement</i> .....	33
3.2.2 <i>Comparison with simple Reinforcement Learning</i> .....	33
3.2.3 <i>Comparison with Supervised Learning Model</i> .....	34
3.3 RIDA FORMALISM.....	35
3.3.1 <i>General Description and Conventions</i> .....	36
3.3.2 <i>The DA Component</i> .....	37
3.3.3 <i>The RI Components</i> .....	38
3.3.4 <i>An Example RIDA Hierarchy</i> .....	39
3.4 RIDA RATIONALE.....	40

3.4.1 Reliability.....	41
3.4.2 Suitability.....	42
3.5 CONCLUSION .....	42
<b>4. LEARNING, ADAPTATION AND THE DELIBERATE APPRENTICE.....</b>	<b>43</b>
4.1 INTRODUCTION.....	43
4.2 LEARNING AS AN AID TO ADAPTATION.....	43
4.3 MOTIVATION FOR SELECTING A DELIBERATE APPRENTICE.....	44
4.3.1 Single Layer Perceptron.....	45
4.3.2 Multi-Layer Perceptron.....	46
4.3.3 Associative Reward Penalty.....	46
4.3.4 Rapid Reinforcement Network.....	48
4.4 THEORETICAL CONTRIBUTION.....	51
4.5 CONCLUSION .....	52
<b>5. SELECTING A REFLEXIVE INSTRUCTOR.....</b>	<b>53</b>
5.1 INTRODUCTION.....	53
5.2 MOTIVATION FOR SELECTING A REFLEXIVE INSTRUCTOR.....	53
5.3 PROTOTYPE REFLEXIVE INSTRUCTORS. THE SOLENODON TRIALS.....	54
5.3.1 The RI Tasks.....	55
5.3.2 SOLENODON IV.....	55
5.3.3 The Collision Avoidance RI.....	57
5.3.4 Simple RI collision avoidance control strategies.....	59
5.3.5 Testing the Control Strategies.....	61
5.3.6 Observations.....	65
5.3.7 The Light Seeking RI.....	65
5.3.8 Testing the Control Strategy.....	66
5.3.9 Multiple RI Coexistence.....	67
5.3.10 Observations.....	67
5.4 CONCLUSION .....	67
<b>6. EXPERIMENTAL DESIGN.....</b>	<b>69</b>
6.1 INTRODUCTION.....	69
6.2 LINKING AUTONOMY AND ROBOTICS .....	69
6.3 THE AUTONOMY FRAMEWORK.....	71
6.4 PLACING RIDA IN THE AUTONOMY FRAMEWORK.....	73
6.5 TASKS FOR MEASURING RIDA WITHIN THE FRAMEWORK .....	73
6.5.1 Task 0:.....	74
6.5.2 Task 1.....	75
6.5.3 Task 2.....	76
6.5.4 Task 3.....	78
6.5.5 Task 4.....	79
6.6 CONCLUSION .....	79
<b>7. TESTING THE RIDA ARCHITECTURE AGAINST AUTONOMY TASKS.....</b>	<b>81</b>
7.1 INTRODUCTION.....	81
7.2 THE SIMULATOR AND PLAYBACK MODULES.....	81
7.2.1 Experimental Environment.....	81
7.2.2 The Simulator.....	82
7.2.3 The Sensor Array.....	84
7.2.4 The Playback Module.....	85
7.3 TASK 0: REFLEXIVE AVOIDANCE OF OBJECTS.....	86
7.3.1 Description of Task.....	86
7.3.2 Description of Vehicle.....	86



7.3.3 Discussion of Results.....	87
7.4 TASK 1: LEARNED AVOIDANCE OF OBJECTS .....	89
7.4.1 Description of Task.....	89
7.4.2 Description of the Vehicle.....	89
7.4.3 Discussion of Results.....	90
7.5 TASK 2: MIXED GOALS--AVOID COLLISIONS AND FIND THE LIGHT .....	93
7.5.1 Description of Task.....	93
7.5.2 Description of the Vehicle.....	94
7.5.3 Discussion of Results.....	95
7.6 TASK 3: MIXED GOALS--AVOID THE LIGHT AND AVOID COLLISION.....	97
7.6.1 Description of Task.....	97
7.6.2 Description of the Vehicle.....	98
7.6.3 Discussion of Results.....	99
7.7 TASK 4: CASCADING RIDA CONTROL: RELIABLE DA BECOMES RI TO TEACH NEW DA.....	100
7.7.1 Description of Task.....	100
7.7.2 Description of the Vehicle.....	101
7.7.3 Discussion of Results.....	102
7.8 CONCLUSION .....	104
<b>8. RELEVANCE .....</b>	<b>105</b>
8.1 INTRODUCTION.....	105
8.2 REVISITING THE GOALS OF THIS WORK.....	105
8.3 CONTINUING WORK.....	108
8.3.1 N-CART, The Natural Selection Research Group and Autonomous Vehicles.....	108
8.4 THE REAL TIME PROBLEM.....	109
8.5 AFTER WORD.....	109
<b>APPENDIX A : SAMPLE SIMULATOR TELEMETRY INFORMATION .....</b>	<b>112</b>
<b>APPENDIX B : THE EMMA EXPERIMENTS.....</b>	<b>113</b>
INTRODUCTION.....	113
THE ENVIRONMENT.....	113
EMMA I: FORAGING.....	114
<i>Lessons Learned from EMMA I.....</i>	<i>118</i>
EMMA II: FOLLOWING AND WANDERING.....	118
<i>Lessons Learned from EMMA II.....</i>	<i>122</i>
EMMA II.5: SWITCHING BEHAVIOUR.....	122
<i>Lessons Learned.....</i>	<i>126</i>
EMMA III: DEFENCE.....	126
<b>APPENDIX C :RAPID REINFORCEMENT NETWORK ARCHITECTURE AND ALGORITHM.....</b>	<b>128</b>
REINFORCEMENT NETWORK .....	128
<i>Initialization.....</i>	<i>129</i>
TEMPORAL DIFFERENCE LINEAR NETWORK .....	131
<i>Initialization.....</i>	<i>132</i>
<i>Feed Forward Operation.....</i>	<i>132</i>
<i>Feed Backward Operation.....</i>	<i>132</i>
<b>APPENDIX D : IMPLEMENTATION DETAILS.....</b>	<b>133</b>
VEHICLE 1.....	133
VEHICLE 2.....	136
VEHICLE 3.....	139
VEHICLE 4.....	140

**BIBLIOGRAPHY.....144**

## LIST OF TABLES

TABLE 5-1 RESULTS WITH NO CONTROL STRATEGY .....	62
TABLE 5-2 SHUNTED CONTROL WITH SHORT SUSTAIN TIME.....	63
TABLE 5-3 SHUNTED CONTROL WITH LONG SUSTAIN TIME.....	64
TABLE 5-4 CROSS CONNECTED REACTIVE CONTROL WITH SHORT SUSTAIN TIME .....	64
TABLE 5-5 CROSS CONNECTED CONTROL WITH LONG SUSTAIN TIME .....	65
TABLE 7-1 PERFORMANCE OF DIFFERENT VEHICLE CONFIGURATIONS .....	96

# LIST OF ILLUSTRATIONS

FIGURE 2-1 REINFORCEMENT LEARNING .....	21
FIGURE 3-1 THE RIDA MODEL .....	32
FIGURE 3-2 REINFORCEMENT LEARNING .....	34
FIGURE 3-3 SUPERVISED LEARNING.....	34
FIGURE 3-4 THE RIDA ARCHITECTURE .....	35
FIGURE 3-5 AN EXAMPLE RIDA CASCADING HIERARCHY .....	39
FIGURE 3-6 SPECIFIC RIDA DESIGN.....	40
FIGURE 3-7 GRACEFUL DEGRADATION OF SERVICE.....	40
FIGURE 4-1 AN ARP PROCESSING ELEMENT .....	47
FIGURE 5-1 SOLENODON IV (SCHEMATIC VIEW).....	56
FIGURE 5-2 SOLENODON IV (3/4 VIEW).....	56
FIGURE 5-3 THE SOLENODON COLLISION AVOIDANCE RI.....	58
FIGURE 5-4 RI REACTING TO CONTACT WITH ONE OF ITS SENSORS.....	59
FIGURE 5-5 RI STRATEGY ONE .....	60
FIGURE 5-6 RI STRATEGY TWO AND THREE.....	60
FIGURE 5-7 RI STRATEGY FOUR AND FIVE .....	61
FIGURE 5-8 THE TRACK AFTER A RUN.....	61
FIGURE 5-9 HALF OF LIGHT SEEKING RI CIRCUIT.....	66
FIGURE 6-1 REFLEXIVE COLLISION AVOIDANCE.....	75
FIGURE 6-2 LEARNED OBSTACLE AVOIDANCE .....	76
FIGURE 6-3 LEARNED COLLISION AVOIDANCE AND LIGHT SEEKING .....	77
FIGURE 6-4 LEARNED COLLISION AND LIGHT AVOIDANCE.....	78
FIGURE 7-1 SIMULATOR ARCHITECTURE.....	83
FIGURE 7-2 PERMITTED MOVEMENT.....	84
FIGURE 7-3 SCREEN SHOT OF SIMULATOR RUNNING .....	85
FIGURE 7-4 SPECIFIC CONFIGURATION FOR TEST 0.....	86
FIGURE 7-5 TRIAL RUN FOR TEST 0.....	87
FIGURE 7-6 TASK 0 SHOWING COLLISIONS PER 10 TIME STEPS.....	88
FIGURE 7-7 SPECIFIC CONFIGURATION FOR TEST 1.....	90
FIGURE 7-8 TEST 1 TRIAL RUN SHOWING BOTH RESULTS (WITH RI AND WITHOUT).....	91
FIGURE 7-9 TEST 1 WITH RI AND DA ACTIVE.....	92
FIGURE 7-10 TEST 0 VS. TEST 1 WITH AND WITHOUT RI.....	93
FIGURE 7-11 VEHICLE SPECIFIC CONFIGURATION FOR TEST 2.....	95
FIGURE 7-12 LIGHT SEEKING RI'S RELATIONSHIP TO POSITIVE REINFORCEMENT .....	97
FIGURE 7-13 SPECIFIC VEHICLE CONFIGURATION FOR TEST 3.....	98
FIGURE 7-14 TEST 3 VEHICLE PERFORMANCE CHARACTERISTICS WITH RI ACTIVE.....	100
FIGURE 7-15 SPECIFIC VEHICLE CONFIGURATION FOR TEST 4.....	102
FIGURE 7-16 TRAINING OF PERCEPTRON DA BY REINFORCEMENT RI.....	103
FIGURE 7-17 PERFORMANCE OF PERCEPTRON DA AFTER TRAINING.....	104
FIGURE A-1 EXAMPLE TELEMETRY INFORMATION .....	112
FIGURE B-1 THE ARENA.....	114
FIGURE B-2 SOLAR RECHARGING CIRCUIT .....	114
FIGURE B-3 BRAITENBERG'S CROSSED CONNECTIONS.....	115
FIGURE B-4 EMMA I SEARCH CIRCUIT.....	116
FIGURE B-5 EMMA I MOVING, STOPPING AND RECHARGING.....	117
FIGURE B-6 EMMA I.....	117
FIGURE B-7 EMMA I TYPICAL TRAJECTORY IN THE ARENA.....	118
FIGURE B-8 CONTACT SWITCHES USED TO PROMOTE WALL FOLLOWING IN EMMA II.....	119
FIGURE B-9 MODIFIED WHISKERS ON EMMA II.....	120

FIGURE B-10 EMMA II NEGOTIATING A CORNER .....	121
FIGURE B-11 TYPICAL EMMA II TRAJECTORY IN ARENA .....	121
FIGURE B-12 EMMA II.....	122
FIGURE B-13 EMMA II.5 BEHAVIOUR MODIFICATION CIRCUIT .....	124
FIGURE B-14 EMMA II.5.....	125
FIGURE B-15 EMMA III (THE PHOTO-PINE).....	127
FIGURE B-16 EMMA III.....	128
FIGURE C-1 RAPID REINFORCEMENT NETWORK ARCHITECTURE .....	129
FIGURE C-2 TEMPORAL DIFFERENCE NETWORK ARCHITECTURE.....	131
FIGURE D-1 RIDA VEHICLE 1 IMPLEMENTATION DETAIL.....	134
FIGURE D-2 VEHICLE 1 RIDA ARCHITECTURE.....	136
FIGURE D-3 RIDA VEHICLE 2 IMPLEMENTATION DETAILS.....	137
FIGURE D-4 VEHICLE 2 RIDA ARCHITECTURE.....	139
FIGURE D-5 RIDA VEHICLE 4 IMPLEMENTATION DETAILS.....	141
FIGURE D-6 VEHICLE 4 RIDA ARCHITECTURE.....	143