Table of Contents

1. Introduction 1
  1.1. Intractability and Machine Learning 1
    1.1.1. The Ubiquity of Intractable Theories 1
    1.1.2. A Machine Learning Approach to Intractability 1
    1.1.3. Human Learning and Intractable Theories 2
    1.1.4. Heuristics and Intractable Theories 3
    1.1.5. Dimensions of Machine Learning 4
    1.1.6. Questions Addressed by this Research 5
  1.2. The Intractable Theory Problem 5
    1.2.1. Computational Theories 6
    1.2.2. Completeness and Consistency 6
    1.2.3. Approximations, Approximate Theories and Heuristics 6
    1.2.4. A Pragmatic Notion of Intractability 7
    1.2.5. Problem Specification 8
    1.2.6. An Example Intractable Theory Problem 9
    1.2.7. Trading Accuracy for Efficiency 10
    1.2.8. Learning in Context 10
    1.2.9. Dual Views of the Intractable Theory Problem 11
  1.3. The Proposed Solution 11
    1.3.1. Generic Simplifying Assumptions (*GSAs*) 12
    1.3.2. The *POLLYANNA* Architecture 13
      1.3.2.1. Learning as Search in *POLLYANNA* 15
      1.3.2.2. Inductive Bias in *POLLYANNA* 16
  1.4. Research Methodology 17
    1.4.1. Test Domains 17
    1.4.2. Key Implementation Tasks 18
    1.4.3. History and Status of the Implementation 18
  1.5. Related Research in Machine Learning 19
    1.5.1. EBG and the Intractable Theory Problem 19
      1.5.1.1. Preserving Truth v. Trading Accuracy for Efficiency 20
      1.5.1.2. Deterministic and Non-Deterministic Intractability 20
      1.5.1.3. Absolute and Relative Intractability 21
      1.5.1.4. The Scope of EBG 23
    1.5.2. Research on Learning Heuristics 23
    1.5.3. Research on Learning Approximations 24
  1.6. Thesis of the Dissertation 25
    1.6.1. Viability of the *POLLYANNA* Methodology 25
    1.6.2. Subsidiary Claims 26
      1.6.2.1. Properties of Theories Generated from *GSAs* 27
      1.6.2.2. The Importance of Empirical Learning 29
    1.6.3. The Generality of *GSAs* 30
    1.6.4. A Principled Method of Deriving Heuristics 31
    1.6.5. Analytic Reasoning about Properties of Approximations 32
    1.6.6. Interaction of Approximation and Reformulation 32
    1.6.7. Comparison to Alternative Methods 32

1.6.8. The Power of a Paradigm 34
1.7. Recurring Themes of this Research 34
1.8. Reader's Guide to the Dissertation 36
2. Analytic Generation of Heuristics 38
2.1. Introduction 38
2.1.1. A View of Heuristics 38
2.1.2. Overview 38
2.2. *H* = *IT* + *GSA* + *R*: A Framework for Generating Heuristics 39
2.2.1. An Example Heuristic 40
2.2.2. Derivation of the Example Heuristic 42
2.2.3. A Second Example Heuristic 44
2.2.4. The Generality of *GSAs*: Positive and Negative Consequences 45
2.2.4.1. Discovery of New Heuristics 45
2.2.4.2. The Value of Empirical Learning 45
2.2.5. Deep v. Shallow Representations of Heuristics 46
2.2.6. Development of *GSAs* through Reconstruction of Heuristics 47
2.2.7. Goal Heuristics for Hearts 48
2.3. Architecture of the Approximation Generator 48
2.3.1. Representation of Problem States 49
2.3.2. The Operator Application Cycle 50
2.3.3. Local v. Global Theory Operations 51
2.3.4. Classification of Knowledge in *POLLYANNA* 52
2.3.5. Theories, Classes and Generic Functions 53
2.4. Representation of Intractable Theories 54
2.4.1. The *PT-FORMALISM*: An Abstract Class of Domain Theories 54
2.4.2. A Theory of Hearts 56
2.4.2.1. Issues in Formulating a Hearts Theory 56
2.4.2.2. Representation of the Hearts Theory 57
2.4.3. A Job Scheduling Theory 59
2.4.4. Testing and Verification of Intractable Theories 62
2.5. The *PT-GSAs*: A Family of Generic Simplifying Assumptions 63
2.5.1. Function Invariance 64
2.5.2. Equiprobable Random Variables 66
2.5.3. Probabilistic Independence 68
2.5.4. Efficiency Impact of *PT-GSAs* 69
2.5.5. Justification of *PT-GSAs* 69
2.5.6. Generalizations of *PT-GSAs* 71
2.5.7. Subsumption Relations among *PT-GSAs* 72
2.6. Reformulation Knowledge 74
2.6.1. Folding and Unfolding Function Definitions 75
2.6.2. Reformulation of Composite Functions 75
2.6.3. Memoization 76
2.6.4. Soundness of Reformulation Operations 77
2.6.5. Interaction of Assumption and Reformulation 77
2.7. Search Control for Generation of Approximations 77
2.7.1. Goals for Approximation Generation 77
2.7.2. Problems with Blind Search Control 78
2.7.3. A Proposed Automatic Control Strategy 78
2.7.4. Extensions of the Control Strategy 80
2.7.5. Restrictions of the Control Strategy 81
2.7.6. Human v. Automatic Control 83
2.8. Properties of *POLLYANNA*'s Approximate Theories 84
2.8.1. Accuracy v. Efficiency 84
2.8.2. Asymptotic Time Complexity 84
2.9. Summary of Analytic Generation Techniques 85

3. Analytic Learning Results in *POLLYANNA* 87
3.1. Introduction 87
3.2. Overview 88
3.3. Heuristics Generated in *POLLYANNA* 89
3.3.1. Heuristics for Hearts 89
3.3.1.1. Approximation Generator Output in Hearts 89
3.3.1.2. Equivalence of Hearts Heuristics and Approximate Theories 94
3.3.1.3. The Generality of Hearts Heuristics 99
3.3.1.4. Additional Heuristics for Hearts 100
3.3.2. Heuristics for Job Scheduling 100
3.3.2.1. Approximation Generator Output in Scheduling 100
3.3.2.2. Equivalence of Scheduling Heuristics and Approximate Theories 103
3.3.2.3. Additional Heuristics for Job Scheduling 108
3.4. Covered and Discovered Heuristics 108
3.4.1. Coverage of the Heuristic Hearts Rule Set 108
3.4.2. Coverage of Known Scheduling Heuristics 109
3.4.3. The Originality of Generated Heuristics 110
3.4.4. The Quality of Generated Heuristics 113
3.4.5. Coverage Properties of Automatic Control Strategies 114
3.5. Generality of the Approximation Generator 115
3.5.1. Generality within Domains 116
3.5.2. Generality across Domains 118
3.6. The Lessons of Analytic Learning 119
3.6.1. Limitations and Implementation Difficulties 120
3.6.1.1. Inefficient Nested Summations 120
3.6.1.2. Efficiency Cliffs 121
3.6.1.3. Variables with Unknown or Infinite Ranges 122
3.6.1.4. Second Order Mapping Functions 122
3.6.1.5. Normalization of Probabilities 123
3.6.1.6. Learning Special Case Approximations 123
3.6.1.7. Learning Weights for Decision Criteria 124
3.6.2. Strategies for Representation of Domain Theories 124
3.6.2.1. Transparency and Efficiency of Representations 125
3.6.2.2. Incomplete and Redundant Representations 125
3.6.2.3. Intensional and Extensional Representations 126
3.7. Summary of Analytic Generation Results 127
4. Empirical Testing of Heuristics 128
4.1. Introduction 128
4.2. Overview 128
4.3. Learning Goals and Contextual Knowledge 129
4.3.1. Accuracy and Efficiency Goals for *POLLYANNA* 130
4.3.2. Defining "Accuracy" and "Efficiency" 131
4.4. Theory Space Generation (*TSG*) 132
4.4.1. The Role of *TSG* in *POLLYANNA* 132
4.4.2. *TSG* as an Analytic Reasoning Process 132
4.4.3. The Value of Monotonic Theory Spaces 133
4.4.4. Generation of a Monotonic Theory Space 134
4.4.5. Defining Analytic "Efficiency" 136
4.4.6. Efficiency Impact Theorems 137
4.4.6.1. Impact of Function Invariance 138
4.4.6.2. Impact of Argument Abstraction 139
4.4.6.3. Impact of Equiprobable Random Variables 141
4.4.6.4. Impact of Probabilistic Independence 141
4.4.6.5. Implementation of Efficiency Impact Rules 142
4.4.7. The Significance of Potential Non-Monotonicity 142

4.4.8. Local v. Global Comparisons of Efficiency 143
4.4.9. Equivalence and Redundancy among Candidate Theories 144
4.4.10. Completeness and Consistency of Theories 146
4.4.11. The TSG Algorithm 147
4.4.12. Theory Spaces Generated for Hearts 149
4.4.13. The Size of *POLLYANNA*'s Theory Spaces 150
4.4.13.1. Blind Generation 151
4.4.13.2. Restricted Generation 151
4.4.13.3. The Size of Hearts Theory Spaces 153
4.5. Theory Space Search (*TSS*) 154
4.5.1. Search Algorithms 154
4.5.1.1. TSS-C: Best Accuracy First Search 155
4.5.1.2. TSS-E: Efficiency Bounded Depth First Search 155
4.5.1.3. TSS-G: Best Efficiency First Search 156
4.5.1.4. TSS-K: Efficiency Bounded Depth First Search 157
4.5.1.5. Search Algorithms Implemented in *POLLYANNA* 157
4.5.2. Empirical Evaluation of Theories 158
4.5.2.1. Representation of Training Examples in Hearts 159
4.5.2.2. Generation of Training Examples in Hearts 161
4.5.2.3. Invocation of Candidate Theories 163
4.5.2.4. Measuring Accuracy of Candidate Theories 163
4.5.2.5. Measuring Efficiency of Candidate Theories 165
4.6. Summary of Empirical Testing Techniques 167
5. Empirical Learning Results in *POLLYANNA* 168
5.1. Introduction 168
5.2. Overview 169
5.3. Analysis of TSS Experiments 169
5.3.1. Types of TSS Output 170
5.3.2. A Sample TSS Run 170
5.3.3. Measuring the Accuracy / Efficiency Tradeoff 172
5.3.4. Domain Level Interpretation of Results 176
5.3.4.1. Empirical Results for Selected Theories 176
5.3.4.2. Interactions Among Sub-Expressions 179
5.3.5. Redundancy in the Theory Space 181
5.3.6. Decomposability of the Theory Space 185
5.3.7. Univariate Analysis of Accuracy and Efficiency 186
5.3.7.1. Comparison of Accuracy Measures 186
5.3.7.2. Comparison of Efficiency Measures 186
5.3.7.3. Empirical Tests of Theory Space Monotonicity 188
5.4. Measures of Learning in *POLLYANNA* 190
5.4.1. Knowledge Level Learning in *POLLYANNA* 190
5.4.2. Learning Curves for *POLLYANNA* 192
5.4.3. The Contribution of Analytic Learning 194
5.4.4. The Contribution of Empirical Learning 201
5.5. The Feasibility of Empirical Learning 203
5.5.1. Information and Computation Costs 204
5.5.2. The Impact of Redundancy 204
5.5.3. Control Strategies and the Strength of Inductive Bias 205
5.5.4. The Performance Context 205
5.5.5. The Learning Context 205
5.5.6. The Feasibility of Empirical Testing in Hearts 206
5.6. Enhancements to Theory Space Search 207
5.6.1. Reducing the Costs of Empirical Learning 207
5.6.2. Intelligent Search Control 208
5.6.3. Extended Analysis of Training Examples 208

5.6.4. Use of Enriched Training Examples 208
5.6.5. Early Testing of Search Goals 209
5.6.6. Truth Maintenance Techniques 210
5.6.7. Incremental Learning Methods 211
5.6.8. Starting at the Initial Intractable Theory 212
5.6.9. Alternative Theory Space Organization 212
5.7. Summary of Empirical Learning Results 213
6. Conclusion 214
6.1. Introduction 214
6.2. Evaluation Overview 215
6.3. Viability of the *POLLYANNA* Methodology 217
6.3.1. Non-Deterministic Capabilities of Analytic Generation 217
6.3.2. Semantic Properties of Approximate Theories 217
6.3.3. Accuracy, Efficiency and Optimality of Approximate Theories 219
6.3.4. Operational Usability of Approximate Theories 219
6.3.5. Context Sensitivity of Approximate Theories 220
6.3.6. The Feasibility of Complete Automation 220
6.4. Generality and Domain Independence 222
6.4.1. Evidence of Domain Independence 222
6.4.2. Guidelines for Selection of Application Domains 224
6.5. Principled Derivations of Heuristics 225
6.6. Comparisons to Existing Machine Learning Techniques 226
6.6.1. Explanation-Based Generalization (*EBG*) 226
6.6.2. Traditional Inductive Learning 226
6.7. Knowledge Level Learning in *POLLYANNA* 227
6.8. *POLLYANNA* as a Paradigm 227
6.8.1. Components of the Paradigm 228
6.8.2. A Methodology for Reconstruction of Heuristics 228
6.8.3. Potential Instances of the Paradigm 230
6.8.4. Limits of the Paradigm 231
6.8.5. Extensions of the Paradigm 232
6.8.5.1. Abstraction: Trading Scope or Precision for Efficiency 232
6.8.5.2. An Expanded View of the Intractable Theory Problem 233
6.9. Summary of Contributions 234
6.9.1. An Approach to the Intractable Theory Problem 234
6.9.2. A Family of Generic Simplifying Assumptions 234
6.9.3. A New Type of Learning Problem 234
6.9.4. A Style of Learning 235
Appendix A. Derivations of Two Hearts Heuristics 241
Appendix B. Goal Heuristics for Hearts 247
B.1. A Heuristic Rule Set 247
B.2. *POLLYANNA*'s Coverage of the Heuristic Rule Set 248
Appendix C. The Initial Intractable Hearts Theory 249
Appendix D. The Initial Intractable Scheduling Theory 255
Appendix E. Primitive Functions for the *PT-Formalism* 258
Appendix F. GSA Operators 260
Appendix G. Reformulation Operators 263
Appendix H. Outline of Automatic Control 272
Appendix I. Sample Approximation Generator Operation 275
I.1. Sample Approximation Generator Trace 275

I.2. Sample Approximation Generator Final State 281
Appendix J. Approximation Generator Output: Hearts 283
Appendix K. Semantics of Approximate Hearts Theories 295
K.1. Equivalent forms of Approximate Hearts Functions 295
K.1.1. Analysis of *Exp-Win* 295
K.1.2. Analysis of *Exp-Trick-Value* 297
K.1.3. Analysis of *Exp-Next-Win* 298
K.1.4. Analysis of *Exp-Next-Trick-Value* 299
K.1.5. Additional Key Functions 300
K.2. Assumptions used in Approximate Hearts Functions 301
Appendix L. Approximation Generator Output: Scheduling 304
Appendix M. Semantics of Approximate Scheduling Theories 310
M.1. Equivalent forms of Approximate Scheduling Theories 310
M.1.1. Analysis of *Soonest Scheduling Deadline First* 310
M.1.2. Analysis of *Most Critical Job First* 314
M.2. Assumptions used in Approximate Scheduling Theories 318

List of Figures

Figure 1-1: Empirical v. Analytic Learning 5
Figure 1-2: Questions Addressed by this Research 5
Figure 1-3: The Intractable Theory Problem 8
Figure 1-4: Two Views of the Intractable Theory Problem 11
Figure 1-5: Generic Simplifying Assumptions 12
Figure 1-6: The *POLLYANNA* Architecture 14
Figure 1-7: Types of Intractability 21
Figure 1-8: A Selection of the Heuristics Generated by *POLLYANNA* 28
Figure 2-1: Heuristic Generation Framework 40
Figure 2-2: A Hearts Problem and Solution 41
Figure 2-3: Dump High Rank Rule 41
Figure 2-4: Dump High Point Value Rule 44
Figure 2-5: Approximation Generator Architecture 49
Figure 2-6: Approximation Generator Problem State 50
Figure 2-7: Classification of Knowledge in *POLLYANNA* 52
Figure 2-8: The *PT-FORMALISM* for Representing Domain Theories 54
Figure 2-9: Definitions of Some Hearts Functions 58
Figure 2-10: Hearts Random Variable Tree 59
Figure 2-11: Job Scheduling Problem Definition 60
Figure 2-12: Job Scheduling Random Variable Tree 61
Figure 2-13: The *PT-GSA* Family of Generic Simplifying Assumptions 63
Figure 2-14: Instances of *PT-GSAs* in Verbal Form 64
Figure 2-15: Independence Assumptions Enable Factorization 69
Figure 2-16: Generalizations of the *PT-GSAs* 71
Figure 2-17: Subsumption Relations among *GSAs* 74
Figure 2-18: Reformulation Knowledge 74
Figure 2-19: Folding and Unfolding Function Definitions 75
Figure 2-20: Probabilities and Expectations of Composite Functions 76
Figure 2-21: Insertion of Memo Functions 76
Figure 2-22: Interactions of *GSAs* and Reformulations 77
Figure 2-23: Automatic Control Strategy 79
Figure 2-24: Restricted Control Strategies 82
Figure 3-1: Reference Relations among Hearts Functions 90
Figure 3-2: Structure of Approximate Hearts Theories 91
Figure 3-3: Hearts Functions with Multiple Versions 91
Figure 3-4: Versions of Selected Hearts Functions 92
Figure 3-5: Terms used in Hearts Function Definitions 92
Figure 3-6: Hearts Heuristics and Equivalent Approximate Theories 95
Figure 3-7: Additional Cases of Approximate Hearts Theories 97
Figure 3-8: Sophisticated Hearts Theories 98
Figure 3-9: Reference Relations among Scheduling Functions 101
Figure 3-10: Scheduling Static Evaluation Function 102

Figure 3-11: Selected Scheduling Functions **102**
Figure 3-12: *Soonest Scheduling Deadline First* **104**
Figure 3-13: *Most Critical Job First* **105**
Figure 3-14: Special Cases of *Most Critical Job First* **107**
Figure 3-15: Originality of Generated Heuristics **111**
Figure 3-16: Quality of Generated Heuristics **113**
Figure 3-17: Specific Assumption Generality in the Hearts Domain **117**
Figure 3-18: Specific Assumption Generality in the Scheduling Domain **117**
Figure 3-19: *GSA* Operator Usage Comparison Across Domains **118**
Figure 3-20: Reformulation Operator Usage Comparison Across Domains **119**
Figure 4-1: Learning Goals and Theory Space Search Algorithms **130**
Figure 4-2: Types of Monotonicity Properties **133**
Figure 4-3: Specification of the Theory Space **135**
Figure 4-4: Semantics of *Primitive-Refinement* **135**
Figure 4-5: Theory Space Efficiency Monotonicity Condition **135**
Figure 4-6: Uniform Dominance Cost Comparisons **137**
Figure 4-7: Cost Function Definitions **138**
Figure 4-8: Efficiency Impact of Function Invariance **138**
Figure 4-9: Efficiency Impact of Argument Abstraction **140**
Figure 4-10: Auxiliary Assumptions for Functional Invariance **142**
Figure 4-11: Auxiliary Assumptions for Argument Abstraction **143**
Figure 4-12: Types of Equivalence Among Theories **145**
Figure 4-13: Equivalence Test Algorithm **146**
Figure 4-14: Theory Space Generation Algorithm **148**
Figure 4-15: Small Hearts Theory Space **149**
Figure 4-16: Analysis of Theory Space Sizes **152**
Figure 4-17: Search Given Error and Cost Bounds **155**
Figure 4-18: Search to Minimize Error Given Cost Bound **156**
Figure 4-19: Search to Minimize Cost Given Error Bound **156**
Figure 4-20: Search to Pareto Optimize Given Cost Bound **158**
Figure 4-21: Training Example Representation **159**
Figure 4-22: Public and Private Information **160**
Figure 4-23: Generation of a Training Set **162**
Figure 4-24: Definitions of Accuracy and Error Rate **164**
Figure 4-25: The Abstract System Clock **166**
Figure 5-1: Training Set *A* Specifications **171**
Figure 5-2: Accuracy v. Efficiency Scatterplot **172**
Figure 5-3: The Accuracy / Efficiency Tradeoff **173**
Figure 5-4: The Accuracy / Efficiency Tradeoff: Composite Theories **175**
Figure 5-5: Accuracy and Efficiency of Selected Theories **177**
Figure 5-6: Impact of Refining *Exp-Win* **180**
Figure 5-7: Impact of Refining *Exp-Trick-Value* **181**
Figure 5-8: Impact of Refining *Exp-Next-Win* **182**
Figure 5-9: Impact of Refining *Exp-Next-Trick-Value* **183**
Figure 5-10: Accuracy Level Histogram **184**
Figure 5-11: Correlation of Accuracy Measures **187**
Figure 5-12: Correlation of Efficiency Measures **188**
Figure 5-13: Correlation of Efficiency Measures **189**
Figure 5-14: Comparing States of Knowledge in *POLLYANNA* **191**
Figure 5-15: Procedure for Generating Learning Curves **193**
Figure 5-16: Learning Curve for a Fixed Cost Bound **194**
Figure 5-17: Learning Curve for a Fixed Cost Bound **195**

Figure 5-18: Learning Curve for a Fixed Cost Bound 196
Figure 5-19: Learning Curve for a Fixed Cost Bound 197
Figure 5-20: Learning Curve for a Fixed Cost Bound 198
Figure 5-21: Comparing the Hearts Theory Space to Random Theories 200
Figure 5-22: Training Set *B* Specifications 201
Figure 5-23: Training Set *C* Specifications 202
Figure 5-24: Variation of Results with Training Sets 203
Figure 6-1: Criteria for Evaluating the Viability of *POLLYANNA* 215
Figure 6-2: Components of the Paradigm 228
Figure 6-3: Reconstruction of Heuristics 229
Figure 6-4: Solutions to the Intractable Theory Problem 233