

CONTENTS

Preface	xvii
Acknowledgments	xix
1 Introduction	1
2 Background on Genetic Algorithms, LISP, and Genetic Programming	21
3 Hierarchical Problem-Solving	45
4 Introduction to Automatically Defined Functions – The Two-Boxes Problem	57
5 Problems that Straddle the Breakeven Point for Computational Effort	109
6 Boolean Parity Functions	157
7 Determining the Architecture of the Program	201
8 The Lawnmower Problem	225
9 The Bumblebee Problem	275
10 The Increasing Benefits of ADFs as Problems are Scaled Up	301
11 Finding an Impulse Response Function	307
12 Artificial Ant on the San Mateo Trail	349
13 Obstacle-Avoiding Robot	365
14 The Minesweeper Problem	377
15 Automatic Discovery of Detectors for Letter Recognition	389
16 Flushes and Four-of-a-Kinds in a Pinochle Deck	417
17 Introduction to Biochemistry and Molecular Biology	429
18 Prediction of Transmembrane Domains in Proteins	445
19 Prediction of Omega Loops in Proteins	493
20 Lookahead Version of the Transmembrane Problem	505
21 Evolutionary Selection of the Architecture of the Program	525
22 Evolution of Primitives and Sufficiency	575
23 Evolutionary Selection of Terminals	597
24 Evolution of Closure	601
25 Simultaneous Evolution of Architecture, Primitive Functions, Terminals, Sufficiency, and Closure	611
26 The Role of Representation and the Lens Effect	619
27 Conclusion	643

Appendix A: List of Special Symbols	647
Appendix B: List of Special Functions	651
Appendix C: List of Type Fonts	653
Appendix D: Default Parameters	655
Appendix E: Computer Implementation	661
Appendix F: Annotated Bibliography of Genetic Programming	697
Appendix G: Electronic Mailing List and Public Repository	717
Bibliography	719
Index	735

Detailed Table of Contents

1	Introduction	1
1.1	Overview	7
2	Background on Genetic Algorithms, LISP, and Genetic Programming	21
2.1	Background on Genetic Algorithms	21
2.2	Background on LISP	31
2.3	Background on Genetic Programming	35
2.4	Sources of Additional Information	42
3	Hierarchical Problem-Solving	45
3.1	Hierarchical Decomposition	45
3.2	Recursive Application and Identical Reuse	47
3.3	Parameterized Reuse and Generalization	49
3.4	Abstraction	51
3.5	SOAR and Explanation-Based Generalization	53
4	Introduction to Automatically Defined Functions – The Two-Boxes Problem	57
4.1	The Problem	57
4.2	Preparatory Steps without ADFs	60
4.3	Results without ADFs	63
4.4	The Idea of Subroutines	67
4.5	The Idea of Automatically Defined Functions	73
4.6	Preparatory Steps with ADFs	81
4.7	Creation of the Initial Random Population	84
4.8	Structure-Preserving Crossover and Typing	85
4.9	Results with ADFs	87
4.10	Comparison of the Structural Complexity of the Solutions	98
4.11	Comparison of Computational Effort	99
4.12	Summary	106
5	Problems that Straddle the Breakeven Point for Computational Effort	109
5.1	Sextic versus Quintic Polynomial	110
5.1.1	Sextic Polynomial $x^6 - 2x^4 + x^2$	111
5.1.1.1	Preparatory Steps without ADFs	111
5.1.1.2	Results without ADFs	112
5.1.1.3	Preparatory Steps with ADFs	113
5.1.1.4	Results with ADFs	114
5.1.1.5	Comparison with and without ADFs	117
5.1.2	Quintic Polynomial $x^5 - 2x^3 + x$	118
5.1.2.1	Preparatory Steps without ADFs	118

5.1.2.2	Results without ADFs	119
5.1.2.3	Results with ADFs	120
5.1.2.4	Comparison with and without ADFs	120
5.2	The Boolean 6-Symmetry versus 5-Symmetry	122
5.2.1	The Boolean 6-Symmetry Problem	122
5.2.1.1	Preparatory Steps without ADFs	123
5.2.1.2	Results without ADFs	124
5.2.1.3	Preparatory Steps with ADFs	125
5.2.1.4	Results with ADFs	129
5.2.1.5	Comparison with and without ADFs	132
5.2.2	The Boolean 5-Symmetry Problem	132
5.2.2.1	Results without ADFs	132
5.2.2.2	Results with ADFs	132
5.2.2.3	Comparison with and without ADFs	134
5.3	The Four-Sine versus Three-Sine Problems	134
5.3.1	The Four-Sine Problem – $\sin x + \sin 2x + \sin 3x + \sin 4x$	135
5.3.1.1	Preparatory Steps without ADFs	135
5.3.1.2	Results without ADFs	135
5.3.1.3	Preparatory Steps with ADFs	137
5.3.1.4	Results with ADFs	138
5.3.1.5	Comparison with and without ADFs	141
5.3.2	The Three-Sine Problem – $\sin x + \sin 2x + \sin 3x$	142
5.3.2.1	Results without ADFs	142
5.3.2.2	Results with ADFs	144
5.3.2.3	Comparison with and without ADFs	144
5.4	Four Occurrences versus Three Occurrences of a Reusable Constant	144
5.4.1	Three-Term Expression $x / \pi + x^2 / \pi^2 + 2\pi x$	145
5.4.1.1	Preparatory Steps without ADFs	146
5.4.1.2	Results without ADFs	146
5.4.1.3	Preparatory Steps with ADFs	147
5.4.1.4	Results with ADFs	148
5.4.1.5	Comparison with and without ADFs	150
5.4.2	The Two-Term Expression $x / \pi + x^2 / \pi^2$	151
5.4.2.1	Results without ADFs	151
5.4.2.2	Results with ADFs	152
5.4.2.3	Comparison with and without ADFs	153
5.5	Summary	153
6	Boolean Parity Functions	157
6.1	The Even-Parity Problem	157
6.2	Preparatory Steps without ADFs	158
6.3	Even-3-Parity without ADFs	161
6.4	Even-4-Parity without ADFs	162
6.5	Even-5-Parity without ADFs	162
6.6	Even-6-Parity without ADFs	164
6.7	Multiple Function-Defining Branches	166
6.8	Hierarchical Automatically Defined Functions	167
6.9	Preparatory Steps with ADFs	170
6.10	Even-3-Parity with ADFs	175
6.11	Even-4-Parity with ADFs	178
6.12	Even-5-Parity with ADFs	180
6.13	Even-6-Parity Problem with ADFs	188
6.14	Summary for the Even-3-, 4-, 5-, and 6-Parity problems	189
6.15	Scaling for the Even-3-, 4-, 5-, and 6-Parity problems	190
6.16	Higher-Order Even-Parity problems	194

6.16.1	Even-7-Parity Problem	194
6.16.2	Even-8-Parity Problem	195
6.16.3	Even-9-Parity Problem	196
6.16.4	Even-10-Parity Problem	197
6.16.5	Even-11-Parity Problem	197
7	Determining the Architecture of the Program	201
7.1	Method of Prospective Analysis	202
7.2	Method of Providing Seemingly Sufficient Capacity	203
7.3	Method of Using Affordable Capacity	204
7.4	Method of Retrospective Analysis	204
7.4.1	Baseline for the Even-5-Parity Problem without ADFs	205
7.4.2	One Two-Argument ADF	206
7.4.3	One Three-Argument ADF	206
7.4.4	One Four-Argument ADF	206
7.4.5	Two Two-Argument ADFs	208
7.4.6	Two Three-Argument ADFs	208
7.4.7	Two Four-Argument ADFs	208
7.4.8	Three Two-Argument ADFs	208
7.4.9	Three Three-Argument ADFs	210
7.4.10	Three Four-Argument ADFs	210
7.4.11	Four Two-Argument ADFs	210
7.4.12	Four Three-Argument ADFs	212
7.4.13	Four Four-Argument ADFs	212
7.4.14	Five Two-Argument ADFs	212
7.4.15	Five Three-Argument ADFs	212
7.4.16	Five Four-Argument ADFs	215
7.5	Summary of Retrospective Analysis	215
8	The Lawnmower Problem	225
8.1	The Problem	226
8.2	Preparatory Steps without ADFs	227
8.3	Lawn Size of 64 without ADFs	228
8.4	Lawn Size of 32 without ADFs	240
8.5	Lawn Size of 48 without ADFs	242
8.6	Lawn Size of 80 without ADFs	242
8.7	Lawn Size of 96 without ADFs	242
8.8	Preparatory Steps with ADFs	244
8.9	Lawn Size of 64 with ADFs	246
8.10	Lawn Size of 32 with ADFs	257
8.11	Lawn Size of 48 with ADFs	259
8.12	Lawn Size of 80 with ADFs	260
8.13	Lawn Size of 96 with ADFs	260
8.14	Summary for Lawn Sizes of 32, 48, 64, 80, and 96	262
8.15	Scaling for Lawn Sizes of 32, 48, 64, 80, and 96	263
8.16	Wallclock Time for the Lawnmower Problem	268
9	The Bumblebee Problem	275
9.1	The Problem	275
9.2	Preparatory Steps without ADFs	275
9.3	Results with 25 Flowers without ADFs	277
9.4	Preparatory Steps with ADFs	279
9.5	Results with 25 Flowers with ADFs	280
9.6	Results with 20 Flowers without ADFs	283
9.7	Results with 20 Flowers with ADFs	284
9.8	Results with 15 Flowers without ADFs	284

9.9	Results with 15 Flowers with ADFs	286
9.10	Results with 10 Flowers without ADFs	286
9.11	Results with 10 Flowers with ADFs	288
9.12	Summary for 10, 15, 20, and 25 Flowers	288
9.13	Scaling with 10, 15, 20, and 25 Flowers	290
9.14	Wallclock time for the Bumblebee Problem	293
10	The Increasing Benefits of ADFs as Problems are Scaled Up	301
10.1	The Benefits of ADFs as a Function of Problem Size	301
10.2	Wallclock time	304
11	Finding an Impulse Response Function	307
11.1	The Problem	307
11.2	Preparatory Steps without ADFs	309
11.3	Results of One Run without ADFs	311
11.4	Results of Series of Runs without ADFs	320
11.5	Preparatory Steps with ADFs	322
11.6	Results of One Run with ADFs	323
11.7	Genealogical Audit Trail with ADFs	335
11.7.1	Crossover in the Result-producing Branch	335
11.7.2	Crossover in the Function-Defining Branch	338
11.8	Results of Series of Runs with ADFs	344
11.9	Summary	347
12	Artificial Ant on the San Mateo Trail	349
12.1	The Problem	349
12.2	Preparatory Steps without ADFs	350
12.3	Results without ADFs	354
12.4	Preparatory Steps with ADFs	355
12.5	Results with ADFs	356
12.6	Summary	364
13	Obstacle-Avoiding Robot	365
13.1	The Problem	365
13.2	Preparatory Steps without ADFs	366
13.3	Results without ADFs	368
13.4	Preparatory Steps with ADFs	372
13.5	Results with ADFs	372
13.6	Summary	376
14	The Minesweeper Problem	377
14.1	The Problem	377
14.2	Preparatory Steps without ADFs	377
14.3	Results without ADFs	378
14.4	Preparatory Steps with ADFs	380
14.5	Results with ADFs	381
14.6	Summary	387
15	Automatic Discovery of Detectors for Letter Recognition	389
15.1	The Problem	390
15.2	Preparatory Steps without ADFs	390
15.3	Results without ADFs	399
15.4	Preparatory Steps with ADFs	399
15.5	Results with ADFs	401
15.6	Genealogical Audit Trails with ADFs	412

15.7	Detectors of Different Sizes and Shapes	414
15.8	Translation-Invariant Letter Recognition	415
15.9	Summary	416
16	Flushes and Four-of-a-Kinds in a Pinochle Deck	417
16.1	The FLUSH Problem	417
16.2	Preparatory Steps without ADFs	417
16.3	Results without ADFs	419
16.4	Preparatory Steps with ADFs	423
16.5	Results with ADFs	424
16.6	Flushes and Four-of-a-Kinds	425
17	Introduction to Biochemistry and Molecular Biology	429
17.1	Chromosomes and DNA	429
17.2	Role of Proteins	430
17.3	Transcription and Translation	430
17.4	Amino Acids and Protein Structure	432
17.5	Primary Structure of Proteins	434
17.6	Secondary Structure of Proteins	434
17.7	Tertiary Structure of Proteins	436
17.8	Quarternary Structure of Proteins	442
17.9	Genetic Algorithms and Molecular Biology	442
18	Prediction of Transmembrane Domains in Proteins	445
18.1	Background on Transmembrane Domains in Proteins	446
18.2	The Four Versions of the Transmembrane Problem	452
18.3	The Idea of Settable Variables, Memory and State	453
18.4	The Idea of Restricted Iteration	454
18.5	Preparatory Steps without ADFs	456
18.5.1	Terminal Set and Function Set	457
18.5.2	Correlation as the Fitness Measure	459
18.5.3	Fitness Cases	462
18.6	Results without ADFs for the Subset-Creating Version	471
18.7	Preparatory Steps with ADFs for the Subset-Creating Version	472
18.8	Results with ADFs for the Subset-Creating Version	474
18.9	Summary for the Subset-Creating Version	488
18.10	The Arithmetic-Performing Version	488
18.11	Summary for the Arithmetic-Performing Version	492
19	Prediction of Omega Loops in Proteins	493
19.1	Background on Omega Loops	493
19.2	Preparatory Steps with ADFs	495
19.3	Results for the Subset-Creating Version with ADFs	500
19.4	Results for the Arithmetic-Performing Version with ADFs	502
19.5	Summary of the Omega-loop problem	504
20	Lookahead Version of the Transmembrane Problem	505
20.1	The Problem	505
20.2	Partial Parsing	506
20.3	Preparatory Steps	511
20.4	Results	513
21	Evolutionary Selection of the Architecture of the Program	525
21.1	Creation of the Initial Random Population	527
21.2	Point typing for Structure-Preserving Crossover	532
21.3	Results for the Even-5-Parity Problem	540

21.4	Results for the Even-4-Parity Problem	561
21.5	Results for the Even-3-Parity Problem	572
21.6	Summary	572
22	Evolution of Primitives and Sufficiency	575
22.1	Primitive Defining Branches	576
22.2	Results for the Even-5-Parity Problem	580
22.3	Results for the Boolean 6-Multiplexer Problem	592
22.4	Results for a Single Primitive Function	594
22.4.1	Boolean 6-Multiplexer Problem	594
22.4.2	Even-5-Parity Problem	595
23	Evolutionary Selection of Terminals	597
23.1	Preparatory Steps	597
23.2	Results for the Even-5-Parity Problem	598
24	Evolution of Closure	601
24.1	Undefined Values	601
24.2	Preparatory Steps	602
24.3	Results for the Even-4-Parity Problem	603
24.4	Results for the Even-5-Parity Problem	607
25	Simultaneous Evolution of Architecture, Primitive Functions, Terminals, Sufficiency, and Closure	611
25.1	Preparatory Steps	611
25.2	Results for Even-4-Parity Problem	611
25.3	Results for Even-5-Parity Problem	612
25.4	Summary	617
26	The Role of Representation and the Lens Effect	619
26.1	Even-3-, 4-, 5-, and 6-Parity Problems	621
26.1.1	Even-3-Parity Problem	621
26.1.2	Even-4-Parity Problem	624
26.1.3	Even-5-Parity Problem	626
26.1.4	Even-6-Parity Problem	627
26.1.5	Summary for the Parity Problems	627
26.2	The Lawnmower Problem	628
26.2.1	Lawnmower Problem with Lawn Size of 32	628
26.2.2	Lawnmower Problem with Lawn Size of 48	629
26.2.3	Lawnmower Problem with Lawn Size of 64	629
26.2.4	Lawnmower Problem with Lawn Size of 80	632
26.2.5	Lawnmower Problem with Lawn Size of 96	633
26.2.6	Summary for the Lawnmower Problem	634
26.3	The Bumblebee Problem	634
26.3.1	Bumblebee Problem with 10 Flowers	634
26.3.2	Bumblebee Problem with 15 Flowers	636
26.3.3	Bumblebee Problem with 20 Flowers	636
26.3.4	Bumblebee Problem with 25 Flowers	636
26.3.5	Summary for Bumblebee Problem	637
26.4	Obstacle-Avoiding-Robot Problem	637
26.5	Minesweeper Problem	639
26.6	Artificial Ant Problem	639
26.7	DISCUSSION	640
27	Conclusion	643
	Appendix A: List of Special Symbols	647

Appendix B: List of Special Functions	651
Appendix C: List of Fonts	653
Appendix D: Default Parameters	655
Appendix E: Computer Implementation	661
E.1 Problem Specific Code for Boolean Even-5-Parity Problem	663
E.2 Kernel	674
Appendix F: Annotated Bibliography of Genetic Programming	697
F.1 Design	698
F.1.1 Design of Stack Filters and Fitting Chaotic Data	698
F.2 Pattern Recognition and Classification	698
F.2.1 Feature Discovery and Image Discrimination	698
F.2.2 Pattern Recognition using Automatically Defined Features	699
F.2.3 Upgrading Rules for an OCR System	699
F.2.4 Prediction of Secondary Structure of Proteins	699
F.2.5 The Donut Problem	699
F.2.6 Evolution of a Model for a Jetliner	700
F.3 Robotic Control	700
F.3.1 Crawling and Walking of a Six-Legged Creature	700
F.3.2 Evolution of Herding Behavior	700
F.3.3 Obstacle-Avoiding Behavior	701
F.3.4 Corridor-Following and the Lens Effect	701
F.3.5 Control of Autonomous Robots	701
F.3.6 Evolution of Cooperation among Autonomous Robots	701
F.3.7 Incorporating Domain Knowledge into Evolution	701
F.3.8 Monitoring Strategy for Independent Agents	702
F.3.9 Genetic Planner for Robots	702
F.3.10 AI Planning Systems	702
F.4 Neural Networks	702
F.4.2 Synthesis of Sigma-Pi Neural Networks	704
F.4.3 New Learning Rules for Neural Networks	704
F.5 Induction and Regression	704
F.5.1 Induction of Regular Languages	704
F.5.2 Levenberg-Marquardt Regression	704
F.5.3 Multiple Steady States of a Dynamical System	705
F.5.4 Inverting and Co-Evolving Randomizers	705
F.5.5 Adaptive Learning using Structured Genetic Algorithms	705
F.5.6 Minimum Description Length and Group Method of Data Handling	705
F.5.7 Sequence Induction	706
F.6 Financial	706
F.6.1 Horse Race Prediction	706
F.6.2 Double Auction Market Strategies	706
F.6.3 C++ Implementation	706
F.7 Art	706
F.7.1 Interactive Evolution of Equations of Images	706
1.7.1.1 Genetic Art in Virtual Reality	707
F.7.2 Jazz Melodies from Case-Based Reasoning and Genetic Programming	707
F.8 Databases	707
F.8.1 News Story Classification by Dow Jones	707
F.8.2 Building Queries for Information Retrieval	707
F.9 Algorithms	707

F.9.1	Evolution of the Schedule for Simulated Annealing	707
F.9.2	Sorting Programs	708
F.10	Natural Language	709
F.10.1	Word Sense Disambiguation	709
F.10.2	Classification of Swedish Words	709
F.11	Modules	709
F.11.1	Module Acquisition and the Genetic Library Builder	709
F.11.2	Modules and Automatically Defined Functions	710
F.11.3	Learning by Adapting Representations	711
F.12	Programming Methods	711
F.12.1	Directed Acyclic Graphs for Representing Populations of Programs	711
F.12.2	Co-Routine Execution Model	711
F.12.3	Stack-Based Virtual Machine	712
F.13	Variations in Genetic Operations	712
F.13.1	Context-Preserving Crossover	712
F.13.2	Brood Selection and Soft Selection	712
F.13.3	Implementation in C++	712
F.13.4	Effect of Locality	712
F.13.5	Biologically Motivated Representation of Programs	712
F.13.6	Niches	713
F.13.7	Recombination and Selection	713
F.13.8	Strongly Typed Genetic Programming	714
F.14	Memory, State, and Mental Models	715
F.14.1	Evolution of Indexed Memory	715
F.14.2	Map-Making and Map-Using	715
F.15	Theoretical Foundations	716
F.15.1	Evolution of Evolvability	716
F.15.2	Fitness Landscapes and Difficulty	716
F.15.3	Schema in Genetic Programming	716
F.15.4	Turing Completeness	716
	Appendix G: Electronic Mailing List and Public Repository	717
	Bibliography	719
	Index	735