

# Contents

<b>1</b>	<b>Turing, Functionalism, and Emergence</b>	<b>1</b>
1.1	Turing Is Among Us	1
1.2	Functionalism	2
1.3	Emergence	3
1.4	Concluding Remarks	4
	References	4
 <b>Part I The Individual Realm</b>		
<b>2</b>	<b>The Individual Realm of Machine Ethics: A Survey</b>	<b>7</b>
2.1	TRUTH-TELLER and SIROCCO	7
2.2	JEREMY and W.D.	8
2.3	MEDTHEX and ETHEL	9
2.4	A Kantian Machine Proposal	11
2.5	Machine Ethics via Theorem Proving	11
2.6	Particularism versus Generalism	12
2.7	Concluding Remarks	14
	References	16
<b>3</b>	<b>Significant Moral Facets Amenable to Logic Programming</b>	<b>19</b>
3.1	Moral Permissibility	19
3.1.1	The Doctrines of Double Effect and Triple Effect	20
3.1.2	Scanlonian Contractualism	22
3.2	The Dual-Process Model	23
3.3	Counterfactual Thinking in Moral Reasoning	24
3.4	Concluding Remarks	26
	References	27
<b>4</b>	<b>Representing Morality in Logic Programming</b>	<b>29</b>
4.1	Preliminaries	29
4.2	Abduction	35
4.3	Preferences Over Abductive Scenarios	37

4.4	Probabilistic LP . . . . .	38
4.5	LP Updating . . . . .	39
4.6	LP Counterfactuals . . . . .	40
4.7	Tabling . . . . .	41
4.8	Concluding Remarks . . . . .	43
	References . . . . .	43
<b>5</b>	<b>Tabling in Abduction and Updating . . . . .</b>	<b>47</b>
5.1	Tabling Abductive Solutions in Contextual Abduction . . . . .	47
5.1.1	TABDUAL Program Transformation. . . . .	49
5.1.2	Implementation Aspects . . . . .	57
5.1.3	Concluding Remarks. . . . .	65
5.2	Incremental Tabling of Fluents for LP Updating. . . . .	66
5.2.1	The EVOLP/R Language . . . . .	67
5.2.2	Incremental Tabling . . . . .	68
5.2.3	The EVOLP/R Approach . . . . .	70
5.2.4	Concluding Remarks. . . . .	75
	References . . . . .	78
<b>6</b>	<b>Counterfactuals in Logic Programming. . . . .</b>	<b>81</b>
6.1	Causation and Intervention in LP . . . . .	82
6.1.1	Causal Model and LP Abduction . . . . .	83
6.1.2	Intervention and LP Updating . . . . .	84
6.2	Evaluating Counterfactuals via LP Abduction and Updating. . . . .	84
6.3	Concluding Remarks . . . . .	89
	References . . . . .	92
<b>7</b>	<b>Logic Programming Systems Affording Morality Experiments . . . . .</b>	<b>95</b>
7.1	ACORDA. . . . .	95
7.1.1	Active Goals . . . . .	97
7.1.2	Abduction and A Priori Preferences . . . . .	98
7.1.3	A Posteriori Preferences . . . . .	98
7.2	PROBABILISTIC EPA . . . . .	99
7.2.1	Abduction and A Priori Preferences . . . . .	99
7.2.2	A Posteriori Preferences . . . . .	100
7.2.3	Probabilistic Reasoning. . . . .	100
7.3	QUALM . . . . .	102
7.3.1	Joint Tabling of Abduction and Updating . . . . .	102
7.3.2	Evaluating Counterfactuals . . . . .	105
7.4	Concluding Remarks . . . . .	106
	References . . . . .	107

<b>8</b>	<b>Modeling Morality Using Logic Programming</b> . . . . .	109
8.1	Moral Reasoning with ACORDA . . . . .	109
8.1.1	Deontological Judgments via A Priori Integrity Constraints . . . . .	117
8.1.2	Utilitarian Judgments via A Posteriori Preferences . . . . .	118
8.2	Moral Reasoning with PROBABILISTIC EPA . . . . .	121
8.3	Moral Reasoning with QUALM . . . . .	123
8.3.1	Moral Updating . . . . .	123
8.3.2	Counterfactual Moral Reasoning . . . . .	128
8.4	Concluding Remarks . . . . .	136
	References . . . . .	137
 <b>Part II The Collective Realm</b>		
<b>9</b>	<b>Modeling Collective Morality via Evolutionary Game Theory</b> . . . . .	141
9.1	The Collective Realm of Machine Ethics . . . . .	141
9.2	Software Sans Emotions but with Ethical Discernment . . . . .	142
9.2.1	Introduction . . . . .	142
9.2.2	Learning to Recognize Intentions and Committing Resolve Cooperation Dilemmas . . . . .	143
9.2.3	Emergence of Cooperation in Groups: Avoidance Versus Restriction . . . . .	145
9.2.4	Why Is It so Hard to Say Sorry? . . . . .	146
9.2.5	Apology and Forgiveness Evolve to Resolve Failures in Cooperative Agreements . . . . .	148
9.2.6	Guilt for Non-humans . . . . .	150
9.3	Concluding Remarks . . . . .	155
	References . . . . .	155
<b>10</b>	<b>Bridging Two Realms of Machine Ethics</b> . . . . .	159
10.1	Bridging the Realms . . . . .	159
10.2	Evolutionary Teachings . . . . .	161
10.3	Concluding Remarks . . . . .	164
	References . . . . .	164
 <b>Part III Coda</b>		
<b>11</b>	<b>Conclusions and Further Work</b> . . . . .	169
	References . . . . .	171
	<b>Index</b> . . . . .	173

Programming Machine Ethics

Pereira, L.M.; Saptawijaya, A.

2016, XIX, 175 p. 5 illus., Hardcover

ISBN: 978-3-319-29353-0