Contents

1 Turing, Functionalism, and Emergence
  1.1 Turing Is Among Us ................................. 1
  1.2 Functionalism ...................................... 2
  1.3 Emergence ......................................... 3
  1.4 Concluding Remarks .............................. 4
References ........................................... 4

Part I The Individual Realm

2 The Individual Realm of Machine Ethics: A Survey
  2.1 TRUTH-TELLER and SIROCCO ....................... 7
  2.2 JEREMY and W.D. .................................. 8
  2.3 MEDETHEX 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

3 Significant Moral Facets Amenable to Logic Programming
  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

4 Representing Morality in Logic Programming .................... 29
  4.1 Preliminaries .................................... 29
  4.2 Abduction ....................................... 35
  4.3 Preferences Over Abductive Scenarios ............... 37
<table>
<thead>
<tr>
<th>4.4 Probabilistic LP</th>
<th>38</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5 LP Updating</td>
<td>39</td>
</tr>
<tr>
<td>4.6 LP Counterfactuals</td>
<td>40</td>
</tr>
<tr>
<td>4.7 Tabling</td>
<td>41</td>
</tr>
<tr>
<td>4.8 Concluding Remarks</td>
<td>43</td>
</tr>
<tr>
<td>References</td>
<td>43</td>
</tr>
</tbody>
</table>

5 Tabling in Abduction and Updating
5.1 Tabling Abductive Solutions in Contextual Abduction
5.1.1 TABDUAL Program Transformation
5.1.2 Implementation Aspects
5.1.3 Concluding Remarks
5.2 Incremental Tabling of Fluents for LP Updating
5.2.1 The EVOLP/R Language
5.2.2 Incremental Tabling
5.2.3 The EVOLP/R Approach
5.2.4 Concluding Remarks
References

6 Counterfactuals in Logic Programming
6.1 Causation and Intervention in LP
6.1.1 Causal Model and LP Abduction
6.1.2 Intervention and LP Updating
6.2 Evaluating Counterfactuals via LP Abduction and Updating
6.3 Concluding Remarks
References

7 Logic Programming Systems Affording Morality Experiments
7.1 ACORDA
7.1.1 Active Goals
7.1.2 Abduction and A Priori Preferences
7.1.3 A Posteriori Preferences
7.2 PROBABILISTIC EPA
7.2.1 Abduction and A Priori Preferences
7.2.2 A Posteriori Preferences
7.2.3 Probabilistic Reasoning
7.3 QUALM
7.3.1 Joint Tabling of Abduction and Updating
7.3.2 Evaluating Counterfactuals
7.4 Concluding Remarks
References
8 Modeling Morality Using Logic Programming

8.1 Moral Reasoning with ACORDA

8.1.1 Deontological Judgments via A Priori Integrity Constraints

8.1.2 Utilitarian Judgments via A Posteriori Preferences

8.2 Moral Reasoning with PROBABILISTIC EPA

8.3 Moral Reasoning with QUALM

8.3.1 Moral Updating

8.3.2 Counterfactual Moral Reasoning

8.4 Concluding Remarks

References

Part II The Collective Realm

9 Modeling Collective Morality via Evolutionary Game Theory

9.1 The Collective Realm of Machine Ethics

9.2 Software Sans Emotions but with Ethical Discernment

9.2.1 Introduction

9.2.2 Learning to Recognize Intentions and Committing Resolve Cooperation Dilemmas

9.2.3 Emergence of Cooperation in Groups: Avoidance Versus Restriction

9.2.4 Why Is It so Hard to Say Sorry?

9.2.5 Apology and Forgiveness Evolve to Resolve Failures in Cooperative Agreements

9.2.6 Guilt for Non-humans

9.3 Concluding Remarks

References

10 Bridging Two Realms of Machine Ethics

10.1 Bridging the Realms

10.2 Evolutionary Teachings

10.3 Concluding Remarks

References

Part III Coda

11 Conclusions and Further Work

References

Index
Programming Machine Ethics
Pereira, L.M.; Saptawijaya, A.
2016, XIX, 175 p. 5 illus., Hardcover
ISBN: 978-3-319-29353-0