

## Contents

<b>Acknowledgements</b> . . . . .	ix
<b>Note on the Terminology</b> . . . . .	xi
<b>Chapter 1. Presenting the Actors</b> . . . . .	1
1.1. The book. . . . .	1
1.2. Human and artificial beings. . . . .	4
1.3. The computer. . . . .	7
1.4. The author. . . . .	9
1.5. CAIA, an artificial AI scientist. . . . .	11
1.6. The research domains of CAIA . . . . .	15
1.7. Further reading . . . . .	19
<b>Chapter 2. Consciousness and Conscience</b> . . . . .	21
2.1. Several meanings of “consciousness” . . . . .	22
2.2. Extending the meaning of “conscience” for artificial beings . . . . .	25
2.3. Why is it useful to build conscious artificial beings with a conscience? . . . . .	29
2.4. Towards an artificial cognition. . . . .	31
2.4.1. A new kind of consciousness . . . . .	32
2.4.2. A new kind of conscience . . . . .	33
<b>Chapter 3. What Does “Itself” Mean for an Artificial Being?</b> . . . . .	35
3.1. Various versions of an individual . . . . .	36
3.1.1. The concept of an individual for human beings . . . . .	36
3.1.2. The boundaries of an artificial being. . . . .	39
3.1.3. Passive and active versions of an individual . . . . .	41
3.1.4. Reflexivity . . . . .	47
3.2. Variants of an individual. . . . .	49

3.2.1. An individual changes with time . . . . .	50
3.2.2. Learning by comparing two variants. . . . .	50
3.2.3. Genetic algorithms . . . . .	52
3.2.4. The bootstrap . . . . .	54
3.3. Cloning artificial beings . . . . .	57
3.3.1. Cloning an artificial being is easy . . . . .	57
3.3.2. Cloning artificial beings is useful. . . . .	58
3.4. Dr. Jekyll and Mr. Hyde . . . . .	61
3.5. The Society of Mind . . . . .	63
3.6. More on the subject. . . . .	65
<b>Chapter 4. Some Aspects of Consciousness . . . . .</b>	<b>67</b>
4.1. Six aspects of consciousness . . . . .	68
4.1.1. One is in an active state . . . . .	68
4.1.2. One knows what one is doing . . . . .	72
4.1.3. One examines his/its internal state . . . . .	80
4.1.4. One knows what one knows . . . . .	84
4.1.5. One has a model of oneself. . . . .	87
4.1.6. One knows that one is different from the other individuals. . . . .	90
4.2. Some limits of consciousness. . . . .	92
4.2.1. Some limits of consciousness for man. . . . .	93
4.2.2. Some limits of consciousness for artificial beings . . . . .	100
<b>Chapter 5. Why is Auto-observation Useful? . . . . .</b>	<b>105</b>
5.1. Auto-observation while carrying out a task . . . . .	105
5.1.1. To guide toward the solution. . . . .	106
5.1.2. To avoid dangerous situations . . . . .	111
5.1.3. To detect mistakes . . . . .	121
5.1.4. To find where one has been clumsy . . . . .	125
5.1.5. To generate a trace . . . . .	126
5.2. Auto-observation after the completion of a task . . . . .	129
5.2.1. Creation of an explanation . . . . .	130
5.2.2. Using an explanation . . . . .	133
5.2.3. Finding anomalies . . . . .	138
<b>Chapter 6. How to Observe Oneself . . . . .</b>	<b>143</b>
6.1. Interpreting . . . . .	146
6.2. Adding supplementary orders . . . . .	150
6.3. Using timed interruptions . . . . .	154
6.4. Using the interruptions made by the operating system . . . . .	158
6.5. Knowing its own state . . . . .	159

6.6. Examining its own knowledge . . . . .	160
6.7. The agents of the Society of Mind. . . . .	165
6.8. The attention . . . . .	166
6.9. What is “I” . . . . .	169
<b>Chapter 7. The Conscience . . . . .</b>	<b>173</b>
7.1. The conscience of human beings . . . . .	174
7.2. The conscience of an artificial being . . . . .	179
7.3. Laws for artificial beings . . . . .	183
7.3.1. Asimov’s laws of robotics . . . . .	183
7.3.1. How can moral laws be implemented? . . . . .	184
7.3.3. The present situation. . . . .	191
<b>Chapter 8. Implementing a Conscience . . . . .</b>	<b>195</b>
8.1. Why is a conscience helpful? . . . . .	197
8.1.1. The conscience helps to solve problems . . . . .	197
8.1.2. The conscience helps to manage its life . . . . .	198
8.1.3. Two ways to define moral knowledge . . . . .	199
8.1.4. Who benefits from the conscience of an artificial being? . . . . .	200
8.2. The conscience of CAIA. . . . .	201
8.3. Implicit principles . . . . .	202
8.4. Explicit principles . . . . .	206
8.5. The consciences in a society of individuals . . . . .	215
8.5.1. The Society of Mind . . . . .	216
8.5.2. Genetic algorithms . . . . .	217
<b>Chapter 9. Around the Conscience . . . . .</b>	<b>219</b>
9.1. Emotions. . . . .	220
9.2. Changing its conscience . . . . .	223
9.3. A new human conscience for our relationships with artificial beings . . . . .	228
<b>Chapter 10. What is the Future for CAIA? . . . . .</b>	<b>237</b>
<b>Appendices . . . . .</b>	<b>239</b>
1. Constraint Satisfaction Problems. . . . .	239
2. How to implement some aspects of consciousness. . . . .	253
<b>Bibliography . . . . .</b>	<b>263</b>
<b>Index . . . . .</b>	<b>269</b>