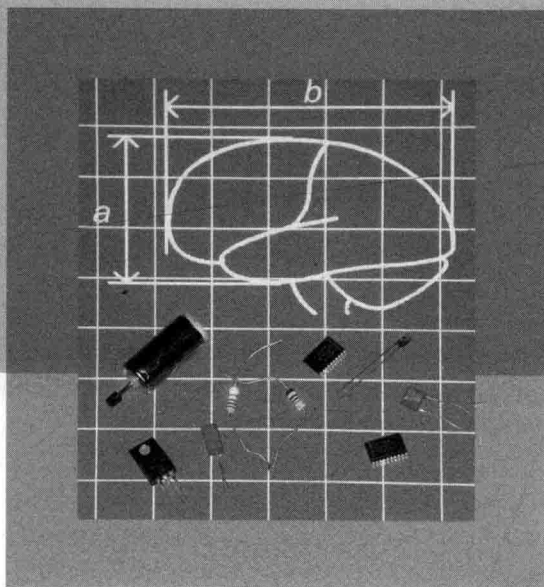


Pentti O Haikonen

CONSCIOUSNESS AND ROBOT SENTIENCE

Series on Machine Consciousness – Vol. 2



CONSCIOUSNESS AND ROBOT SENTIENCE

Pentti O Haikonen

University of Illinois at Springfield, USA

 World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

Series on Machine Consciousness — Vol. 2

CONSCIOUSNESS AND ROBOT SENTIENCE

Copyright © 2012 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN 978-981-4407-15-1

Printed in Singapore by World Scientific Printers.

CONSCIOUSNESS AND ROBOT SENTIENCE

Series on Machine Consciousness

ISSN: 2010-3158

Series Editor: Antonio Chella (*University of Palermo, Italy*)

Published

- Vol. 1 Aristotle's Laptop: The Discovery of our Informational Mind
 by Igor Aleksander and Helen Morton
- Vol. 2 Consciousness and Robot Sentience
 by Pentti Olavi Antero Haikonen

Dedication

This book is humbly dedicated to you, my respected reader and to the international community of machine consciousness researchers.

Preface

Many excellent papers on machine consciousness were presented at the AAAI Fall Symposium on AI and Consciousness at Arlington in 2007. Therefore, at the concluding moments of one presentation the harsh comments of one philosopher were a disturbing shock. The comments were presented in a frank way, but the most disturbing thing was that the philosopher was right. Stevan Harnad lamented that none of the presented papers really addressed the one and only real issue of consciousness and therefore did not deliver the promised goods. This statement was not fully justified, because there were papers that treated the issue, even though not necessarily in a sufficient or even a correct way. Harnad noted that there are many proposed ways to build a machine or a write a computer program that is claimed to be conscious. But alas, none of these approaches really solve the real problem of consciousness and therefore any claims of consciousness in these machines and programs are totally unfounded. The author has no other option than to agree with Harnad.

The real problem of consciousness is recognized by many philosophers in one form or another and many designers of potentially sentient robots are at least vaguely aware of it. But, the real problem has appeared so hard that it has been easier just to ignore it and hope that it will be automatically solved, as soon as a sufficient system complexity is achieved. It will not.

This book is different from most of the contemporary books on machine consciousness. The real problem of consciousness is taken as the starting point and it is explained, not explained away. The rest is developed from the conclusions from this investigation. This book is the

third one in the author's trilogy of machine consciousness books, the previous ones being "The Cognitive Approach to Conscious Machines", Imprint Academic UK 2003 and "Robot Brains; Circuits and Systems for Conscious Machines", Wiley UK 2007. A reader will notice that similar themes are presented in these books. These books augment each other; the first book presents background contemplations from philosophy and cognitive sciences and the second book presents material for engineers. This book amplifies, clarifies and explains the real and fundamental issues and practical aspects of machine consciousness and includes a presentation of the author's experimental cognitive robot XCR-1.*

I would like to thank Prof. Peter Boltuc, Prof. Antonio Chella and Mr. Dylan Drummond for their expert comments and valuable suggestions about the text.

Finally, I want to thank my wonderful, ever young wife Sinikka for her support and encouragement; I can move mountains, if you only hold my hand. Special thanks go also to my media artist son Pete, whose metaphysically captivating and uplifting techno-instrumental composition *Untitle 'Em* with its strong spectrum of amodal qualia gave me inspiration and strength to continue during the longest hours of this effort.

Pentti O A Haikonen

* Demo videos of the robot XCR-1 can be seen at
<http://www.youtube.com/user/PenHaiko/videos>

Contents

Dedication.....	v
Preface	vii
Chapter 1.....	1
Introduction.....	1
1.1. Towards Conscious Robots.....	1
1.2. The Structure of This Book.....	3
Chapter 2.....	5
The Problem of Consciousness	5
2.1. Mind and Consciousness	5
2.2. The Apparent Immateriality of the Mind.....	6
2.3. Cartesian Dualism	7
2.4. Property Dualism.....	10
2.5. The Identity Theory.....	11
2.6. The Real Problem of Consciousness	12
Chapter 3.....	15
Consciousness and Subjective Experience.....	15
3.1. Theories of Consciousness	15
3.2. The Subjective Experience.....	20
3.3. The Internal Appearance of Neural Activity	21
Chapter 4.....	25
Perception and Qualia.....	25
4.1. Perception and Recognition.....	25
4.1.1. What is a Percept?.....	25
4.1.2. Is Perception the Same as Recognition?.....	27
4.2. Qualia.....	30
4.2.1. What Are Qualia?.....	30
4.2.2. The Privacy of Qualia.....	33
4.2.3. No Qualia, No Percepts	35
4.2.4. Different Qualities of Qualia.....	35
4.2.5. Amodal Qualia	37
4.2.6. Externalization, the Apparent Location of Percepts	41

Chapter 5.....	47
From Perception to Consciousness	47
5.1. No Percepts – No Consciousness	47
5.2. Attention and Consciousness.....	48
5.3. The Difference Between Conscious and Non-Conscious Perception	50
5.4. Information Integration and Consciousness	52
5.5. What is Consciousness?	53
Chapter 6.....	55
Emotions and Consciousness	55
6.1. Emotions and Feelings	55
6.2. The Qualia of Emotions	56
6.3. The System Reactions Theory of Emotions (SRTE)	58
6.4. Emotions and Motivation	59
6.5. Free Will.....	60
6.6. Decision Making	61
Chapter 7.....	63
Inner Speech and Consciousness.....	63
7.1. Natural Language	63
7.2. Consciousness and Inner Speech.....	66
7.3. Conscious Perception of Inner Speech	67
Chapter 8.....	73
Qualia and Machine Consciousness	73
8.1. Human Consciousness vs. Machine Consciousness	73
8.2. Preconditions for Machine Qualia.....	74
Chapter 9.....	79
Testing Consciousness	79
9.1. Requirements for Consciousness Tests	79
9.2. Tests for Consciousness	82
9.2.1. The Turing Test.....	82
9.2.2. Picture Understanding Test	83
9.2.3. The Cross-Examination Test	84
9.3. Tests for Self-Consciousness.....	85
9.3.1. Self-Consciousness.....	85
9.3.2. The Mirror Test.....	86
9.3.3. The Name Test	87
9.3.4. The Ownership Test	87
9.3.5. The Cross-Examination Test	88
9.4. Requirements and Tests for Machine Consciousness in Literature	88
9.4.1. Aleksander's Axioms	88
9.4.2. The ConsScale.....	89

Chapter 10.....	91
Artificial Conscious Cognition	91
10.1. Which Model for Artificial Cognition?.....	91
10.2. Sub-symbolic vs. Symbolic Information Processing.....	93
10.3. What Is a Cognitive Architecture?	96
Chapter 11.....	99
Associative Information Processing.....	99
11.1. What Is Associative Information Processing?.....	99
11.2. Basic Associative Processes.....	100
11.2.1. Pavlovian Conditioning	100
11.2.2. Hebbian Learning	102
11.2.3. Autoassociation and Heteroassociation.....	103
11.3. The Representation of Information	104
11.4. Distributed Signal Representations	105
Chapter 12.....	107
Neural Realization of Associative Processing.....	107
12.1. Spiking Neurons or Block Signal Neurons?.....	107
12.2. Associative Neurons and Synapses	109
12.3. Correlative Learning	114
12.4. The Associative Neuron as a Logic Element.....	117
12.5. Associative Neuron Groups	119
12.5.1. The Operation of an Associative Neuron Group.....	119
12.5.2. The Association of Vectors with Vectors	121
12.5.3. Autoassociative Memory	122
12.5.4. Temporal Sequences	122
Chapter 13.....	125
Designing a Cognitive Perception System	125
13.1. Requirements for Cognitive Perception	125
13.2. The Perception/Response Feedback Loop.....	128
13.2.1. The Basic Principle.....	128
13.2.2. The Realization of Priming	129
13.2.3. Match, Mismatch and Novelty Detection	130
13.2.4. Sensory Attention in the Perception/Response Feedback Loop	132
13.2.5. The Realization of Visual Searching.....	133
13.2.6. Predictive Perception	134
13.2.7. Conscious Introspection of Imaginations.....	135
Chapter 14.....	139
Examples of Perception/Response Feedback Loops	139
14.1. The Auditory Perception/Response Feedback Loop	139
14.1.1. The Purpose	139
14.1.2. Auditory Pre-Processes.....	139
14.1.3. The Outline.....	142

14.2. The Visual Perception/Response Feedback Loop	144
14.2.1. The Purpose	144
14.2.2. Visual Pre-Processes	145
14.2.3. The Outline	146
14.3. The Touch Perception/Response Feedback Loop.....	148
Chapter 15.....	151
The Transition to Symbolic Processing.....	151
15.1. From Distributed Signals to Symbols.....	151
15.2. Requirements for a Natural Language.....	151
15.3. Association of Meaning	153
Chapter 16.....	155
Information Integration with Multiple Modules.....	155
16.1. Cooperation and Interaction of Multiple Modules	155
16.2. Sensorimotor Integration.....	157
16.3. Feedback Control Loops	158
16.4. Hierarchical Control Loops.....	160
Chapter 17.....	163
Emotional Significance of Percepts	163
17.1. The Significance of Percepts.....	163
17.2. Emotional Evaluation of Percepts	164
Chapter 18.....	167
The Outline of the Haikonen Cognitive Architecture (HCA).....	167
18.1. General Overview	167
18.2. The Block Diagram of HCA	169
18.3. Control, Motivation and Drivers	172
18.4. Information Integration, Coalitions and Consciousness.....	172
Chapter 19.....	175
Mind Reading Applications	175
19.1. Mind Reading Possible?.....	175
19.2. The principle of the Imaging of Inner Imagery	176
19.3. The Perception/Response Feedback Loop in the Imaging of Inner Imagery	179
19.4. Inner Speech and Unvoiced Speech	182
19.5. Silent Speech Detection with the Perception/Response Loop	182
Chapter 20.....	185
The Comparison of Some Cognitive Architectures.....	185
20.1. Introduction.....	185
20.2. Baars Global Workspace Architecture	186
20.2.1. Baars Concept of Consciousness	186
20.2.2. Baars Model.....	187
20.2.3. What Is Explained by the Baars Model.....	190
20.2.4. What Is Not Explained by the Baars Model.....	190

20.3.	Shanahan Global Workspace Architecture	191
20.3.1.	Shanahan Concept of Consciousness.....	191
20.3.2.	Shanahan Model.....	192
20.3.3.	What Is Explained by the Shanahan Model.....	194
20.3.4.	What Is Not Explained by the Shanahan Model.....	195
20.4.	Haikonen Cognitive Architecture	196
20.4.1.	Haikonen Concept of Consciousness.....	196
20.4.2.	Haikonen Model.....	197
20.4.3.	What Is Explained by the Haikonen Model.....	198
20.5.	Baars, Shanahan and Haikonen Architectures Compared.....	200
Chapter 21	203
Example: An Experimental Robot with the HCA	203
21.1.	Purpose and Design Principles	203
21.2.	Architecture	204
21.3.	The Auditory Module	206
21.3.1.	The Auditory Perception/Response Feedback Loop.....	206
21.3.2.	Speech Recognition.....	207
21.3.3.	Speech Production.....	207
21.4.	The Visual Module	208
21.4.1.	Visual Perception/Response Feedback Loops	208
21.4.2.	Object Recognition.....	210
21.4.3.	Direction Detection	210
21.4.4.	The Selection and Searching of a Target	212
21.5.	The Emotional Module.....	212
21.5.1.	Pain and Pleasure	212
21.5.2.	Non-Events.....	213
21.5.3.	Emotional Significance	213
21.5.4.	Self-Concept.....	214
21.6.	The Gripper Module	215
21.7.	The Wheel Drive Module	217
21.8.	Self-Talk	218
21.9.	Cross-Associative Information Integration	220
21.9.1.	The Verbal-Corporal Teaching Experiment	220
21.10.	Consciousness in the XCR-1	221
Chapter 22	225
Concluding Notes	225
22.1.	Consciousness Explained.....	225
22.2.	The Conclusion.....	228
Bibliography	229
Index	235

Chapter 1

Introduction

1.1. Towards Conscious Robots

Why is it so hard to make computers to understand anything and why is it equally hard to design sentient robots with the ability to behave sensibly in everyday situations and environments? The traditional approach of Artificial Intelligence (AI) has not been able to solve these problems. The modest computing power of the early computers in the fifties was seen as the main limiting factor for Artificial Intelligence, but today, when hundreds of gigabytes of memory can be packed into miniscule microchips, this excuse is no longer valid.

Machines still do not understand anything, because they do not operate with meanings. Understanding everyday meanings calls for embodiment. The machine must be able to interact with its environment and learn how things are. We humans do this effortlessly, we see and understand our environment directly and can readily interact with it according to the varying requirements of each moment. We can do this, because we are conscious. Our existing robots cannot do this.

Traditional Artificial Intelligence has not been able to create conscious robots (except in science fiction movies) and it is very likely that it never will (for the reasons explained later on in this book), therefore novel approaches must be found. We know, how conscious subjects behave and we could make machines imitate conscious behavior using the traditional AI methods. However, the mere imitation of conscious behavior is not enough, imitation has its limits and may fail any moment. The only universal solution would be the creation of truly conscious machines, if we only knew what it takes to be conscious.

The philosophy of mind has tried to solve the mystery of consciousness, but with limited success. Contradicting opinions among philosophers abound, as was demonstrated by Aaron Sloman's target paper [Sloman 2010] and the responses to it in the *International Journal of Machine Consciousness* Vol. 2, No 1.

Psychology has not done much better. Neurosciences may be able to associate some neural activities with conscious states, but how these neural activities could give rise to the experience of being conscious has remained unexplained.

Could engineering do better? Engineers are familiar with complex systems. They know, how components can be combined into systems that execute complicated functions. The creation of a cognitive machine should thus be just another engineering project; feasible and executable as soon as the desired functions are defined. These definitions could be delivered by cognitive sciences. No doubt, machines that behave as if they were more or less conscious can be eventually created. But, were these machines really conscious? Would they have the experience of being conscious? If not, then no real consciousness has been created and the machines would only be mere replicas of the original thing, looking real, but not delivering the bullet.

The bullet must be delivered and conscious robots must be aware of their own thoughts and their own existence and know what they are doing. Unfortunately, there has been no engineering definition for the experience of being conscious. Philosophers have been pondering this question for couple of thousand years and have come to the conclusion that the phenomenon of consciousness seems to involve a problem that is hard or even impossible to solve in terms of physical sciences. This issue is known as the mind-body problem. The mind involves the experience of being conscious; how could this be explained with the application of the laws of physics? The natural laws of physics are able to explain the workings of energy and matter, electromagnetic fields, atoms and electrons and eventually the whole universe, for that matter. Yet the natural laws of physics have not been able to explain consciousness. Therefore, is consciousness something beyond energy and matter?

The central hypothesis beyond this book proposes that consciousness is neither energy nor matter and therefore it cannot be explained by the

physical laws about energy and matter. However, consciousness is achievable by the application of energy and matter, because we already have an example of this kind of a conscious system; the human mind. Thus, the experience of being conscious seems to be a property of certain kinds of perceptual systems. Therefore, the explanation for consciousness would be found at the system level and the creation of robots with conscious minds should be possible. However, before the construction of conscious machines the real problem of consciousness must be identified, isolated and analyzed.

1.2. The Structure of This Book

This book begins with the review of the fundamental philosophical issues of consciousness. This treatment leads to the identification of the one and only real problem of consciousness. It is shown that this problem is related to perception and qualia, which are then discussed in detail. The existence of amodal qualia is proposed based on the known concept of amodal features and it is proposed that amodal qualia can give some insights into the phenomenal nature of qualia.

The relation of emotions and inner speech to consciousness is discussed next.

Can machines have qualia? It is argued that qualia are a mandatory prerequisite for human-like machine consciousness. Systems without qualia are not truly conscious. Next, some preconditions for machine qualia are proposed.

How do we know that a person or a machine is conscious? Some proposed tests exist and are presented and discussed here.

The identified preconditions for conscious cognition lead to the requirement of a perceptive system that combines sub-symbolic and symbolic information processing. Associative information processing with associative neural networks and distributed signal representations is introduced as a method for sub-symbolic processing that inherently facilitates the natural transition from sub-symbolic to symbolic processing.

Conscious robot cognition calls for information integration and sensorimotor integration and these lead to the requirement of an architecture, the assembly of cross-connected perception/response and motor modules. The Haikonen Cognitive Architecture (HCA) is presented as an example of a system that would satisfy the identified requirements.

Modern brain imaging technology seems to allow at least limited mind reading. It is proposed that the HCA might be used to augment mind reading technology. One already implemented example is cited.

Many cognitive architectures have been proposed lately and the comparison of their different approaches with the HCA would be interesting and in this way the approach of this book could be put in a wider perspective. However, a complete comparison is not feasible here, therefore a smaller review is attempted and the compared cognitive architectures are the Baars Global Workspace architecture and the Shanahan Architecture, as these are well-known and share many representative features with several other architectures.

Finally, as an example of a practical implementation of the Haikonen Cognitive Architecture, the author's experimental cognitive robot XCR-1 is presented.

At the end of each chapter, where feasible, a chapter summary is provided for easy assimilation of the text. The concluding chapter of this book summarises the explanation of consciousness, as proposed by the author. This summary should be useful and should reveal the main points quickly.