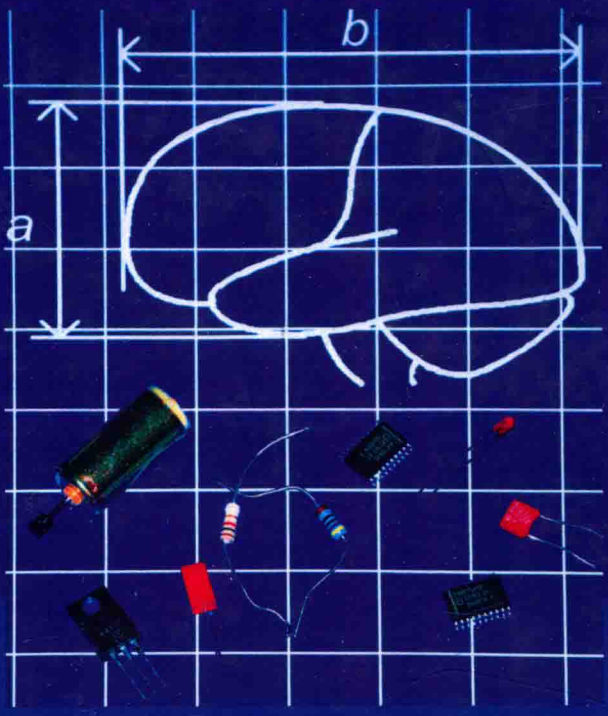
Series on Machine Consciousness – Vol. 2

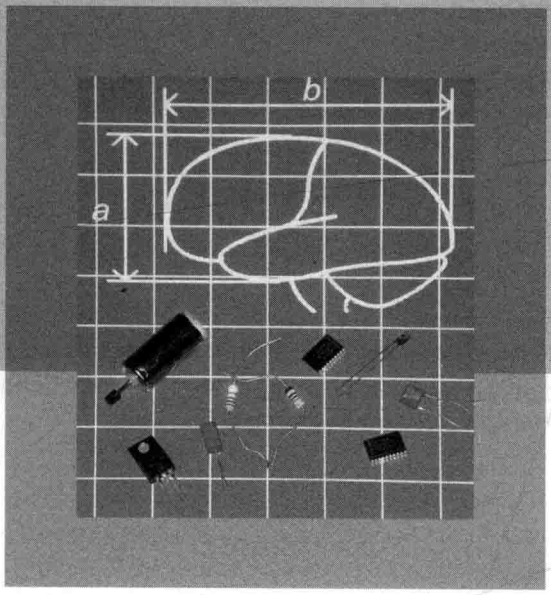


Pentti O Haikonen

CONSCIOUSNESS AND ROBOT SENTIENCE

World Scientific

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

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

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.