

# **Studies in Applied Philosophy, Epistemology and Rational Ethics**

Volume 44

## **Series editor**

Lorenzo Magnani, University of Pavia, Pavia, Italy  
e-mail: [lmagnani@unipv.it](mailto:lmagnani@unipv.it)

## **Editorial Board**

Atocha Aliseda  
Universidad Nacional Autónoma de México (UNAM), Coyoacan, Mexico

Giuseppe Longo  
Centre Cavaillès, CNRS—Ecole Normale Supérieure, Paris, France

Chris Sinha  
School of Foreign Languages, Hunan University, Changsha, P.R. China

Paul Thagard  
Waterloo University, Waterloo, ON, Canada

John Woods  
University of British Columbia, Vancouver, BC, Canada

**Studies in Applied Philosophy, Epistemology and Rational Ethics (SAPERE)** publishes new developments and advances in all the fields of philosophy, epistemology, and ethics, bringing them together with a cluster of scientific disciplines and technological outcomes: ranging from computer science to life sciences, from economics, law, and education to engineering, logic, and mathematics, from medicine to physics, human sciences, and politics. The series aims at covering all the challenging philosophical and ethical themes of contemporary society, making them appropriately applicable to contemporary theoretical and practical problems, impasses, controversies, and conflicts. Our scientific and technological era has offered “new” topics to all areas of philosophy and ethics – for instance concerning scientific rationality, creativity, human and artificial intelligence, social and folk epistemology, ordinary reasoning, cognitive niches and cultural evolution, ecological crisis, ecologically situated rationality, consciousness, freedom and responsibility, human identity and uniqueness, cooperation, altruism, intersubjectivity and empathy, spirituality, violence. The impact of such topics has been mainly undermined by contemporary cultural settings, whereas they should increase the demand of interdisciplinary applied knowledge and fresh and original understanding. In turn, traditional philosophical and ethical themes have been profoundly affected and transformed as well: they should be further examined as embedded and applied within their scientific and technological environments so to update their received and often old-fashioned disciplinary treatment and appeal. Applying philosophy individuates therefore a new research commitment for the 21st century, focused on the main problems of recent methodological, logical, epistemological, and cognitive aspects of modeling activities employed both in intellectual and scientific discovery, and in technological innovation, including the computational tools intertwined with such practices, to understand them in a wide and integrated perspective.

## Advisory Board

|                                            |                                          |
|--------------------------------------------|------------------------------------------|
| A. Abe, Chiba, Japan                       | W. Park, Daejeon, South Korea            |
| H. Andersen, Copenhagen, Denmark           | A. Pereira, São Paulo, Brazil            |
| O. Bueno, Coral Gables, USA                | L. M. Pereira, Caparica, Portugal        |
| S. Chandrasekharan, Mumbai, India          | A.-V. Pietarinen, Helsinki, Finland      |
| M. Dascal, Tel Aviv, Israel                | D. Portides, Nicosia, Cyprus             |
| G. D. Crnkovic, Göteborg, Sweden           | D. Provijn, Ghent, Belgium               |
| M. Ghins, Lovain-la-Neuve, Belgium         | J. Queiroz, Juiz de Fora, Brazil         |
| M. Guarini, Windsor, Canada                | A. Raftopoulos, Nicosia, Cyprus          |
| R. Gudwin, Campinas, Brazil                | C. Sakama, Wakayama, Japan               |
| A. Heeffer, Ghent, Belgium                 | C. Schmidt, Le Mans, France              |
| M. Hildebrandt, Rotterdam, The Netherlands | G. Schurz, Dusseldorf, Germany           |
| K. E. Himma, Seattle, USA                  | N. Schwartz, Buenos Aires, Argentina     |
| M. Hoffmann, Atlanta, USA                  | C. Shelley, Waterloo, Canada             |
| P. Li, Guangzhou, P.R. China               | F. Stjernfelt, Aarhus, Denmark           |
| G. Minnameier, Frankfurt, Germany          | M. Suarez, Madrid, Spain                 |
| M. Morrison, Toronto, Canada               | J. van den Hoven, Delft, The Netherlands |
| Y. Ohsawa, Tokyo, Japan                    | P.-P. Verbeek, Enschede, The Netherlands |
| S. Paavola, Helsinki, Finland              | R. Viale, Milan, Italy                   |
|                                            | M. Vorms, Paris, France                  |

More information about this series at <http://www.springer.com/series/10087>

Vincent C. Müller  
Editor

# Philosophy and Theory of Artificial Intelligence 2017

*Editor*  
Vincent C. Müller  
Interdisciplinary Ethics Applied (IDEA)  
Centre  
University of Leeds  
Leeds, West Yorkshire  
UK

ISSN 2192-6255 ISSN 2192-6263 (electronic)  
Studies in Applied Philosophy, Epistemology and Rational Ethics  
ISBN 978-3-319-96447-8 ISBN 978-3-319-96448-5 (eBook)  
<https://doi.org/10.1007/978-3-319-96448-5>

Library of Congress Control Number: 2018948694

© Springer Nature Switzerland AG 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Editorial Note

The papers in this volume result from the 3rd conference on the “Philosophy and Theory of Artificial Intelligence” (PT-AI) 4–5 November 2017 which I organised in Leeds where I am a university fellow—for details on the conference, see <http://www.pt-ai.org/>.

For this conference, we had 77 extended abstract submissions by the deadline, which were reviewed double-blind by two to four referees. A total of 28 submissions, i.e. 36%, were accepted for presentation. We also accepted 18 posters to be presented. The invited speakers were Thomas Metzinger, Mark Sprevak, José Hernández-Orallo, Yi Zeng, Susan Schneider, David C. Hogg and Peter Millican. All papers and posters were submitted in January at full length (two pages for posters) and reviewed another time by at least two referees among the authors. In the end, we have 32 papers here that represent the current state of the art in the philosophy of AI. We grouped the papers broadly into three categories: “Cognition–Reasoning–Consciousness”, “Computation–Intelligence–Machine Learning” and “Ethics–Law”. This year, we see a significant increase in ethics, more work on machine learning, perhaps less on embodiment or computation—and a stronger “feel” that our area of work has entered the mainstream. This is also evident from more papers in “standard” journals, more book publications with mainstream philosophy presses and more philosophers from neighbouring fields joining us, such as philosophy of mind or philosophy of science. These are encouraging developments, and we are looking forward to PT-AI 2019!

We gratefully acknowledge support from the journal *Artificial Intelligence*, and the *IDEA Centre* at the University of Leeds.

May 2018

Vincent C. Müller

# Contents

## **Cognition - Reasoning - Consciousness**

|                                                                                                           |           |
|-----------------------------------------------------------------------------------------------------------|-----------|
| <b>Artificial Consciousness: From Impossibility to Multiplicity . . . . .</b>                             | <b>3</b>  |
| Chuanfei Chin                                                                                             |           |
| <b>Cognition as Embodied Morphological Computation . . . . .</b>                                          | <b>19</b> |
| Gordana Dodig-Crnkovic                                                                                    |           |
| <b>“The Action of the Brain”: Machine Models and Adaptive<br/>Functions in Turing and Ashby . . . . .</b> | <b>24</b> |
| Hajo Greif                                                                                                |           |
| <b>An Epistemological Approach to the Symbol Grounding Problem . . . . .</b>                              | <b>36</b> |
| Jodi Guazzini                                                                                             |           |
| <b>An Enactive Theory of Need Satisfaction . . . . .</b>                                                  | <b>40</b> |
| Soheil Human, Golnaz Bidabadi, Markus F. Peschl, and Vadim Savenkov                                       |           |
| <b>Agency, Qualia and Life: Connecting Mind and Body Biologically . . . . .</b>                           | <b>43</b> |
| David Longinotti                                                                                          |           |
| <b><i>Dynamic Concept Spaces</i> in Computational Creativity for Music . . . . .</b>                      | <b>57</b> |
| René Mogensen                                                                                             |           |
| <b>Creative AI: Music Composition Programs as an Extension<br/>of the Composer’s Mind . . . . .</b>       | <b>69</b> |
| Caterina Moruzzi                                                                                          |           |
| <b>How Are Robots’ Reasons for Action Grounded? . . . . .</b>                                             | <b>73</b> |
| Bryony Pierce                                                                                             |           |
| <b>Artificial Brains and Hybrid Minds . . . . .</b>                                                       | <b>81</b> |
| Paul Schweizer                                                                                            |           |

|                                                                                                                                                                                                                                                                                      |     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Huge, but Unnoticed, Gaps Between Current AI and Natural Intelligence</b> . . . . .                                                                                                                                                                                               | 92  |
| Aaron Sloman                                                                                                                                                                                                                                                                         |     |
| <b>Social Cognition and Artificial Agents</b> . . . . .                                                                                                                                                                                                                              | 106 |
| Anna Strasser                                                                                                                                                                                                                                                                        |     |
| <b>Computation - Intelligence - Machine Learning</b>                                                                                                                                                                                                                                 |     |
| <b>Mapping Intelligence: Requirements and Possibilities</b> . . . . .                                                                                                                                                                                                                | 117 |
| Sankalp Bhatnagar, Anna Alexandrova, Shahar Avin, Stephen Cave,<br>Lucy Cheke, Matthew Crosby, Jan Feyereisl, Marta Halina,<br>Bao Sheng Loe, Seán Ó hÉigeartaigh, Fernando Martínez-Plumed,<br>Huw Price, Henry Shevlin, Adrian Weller, Alan Winfield,<br>and José Hernández-Orallo |     |
| <b>Do Machine-Learning Machines Learn?</b> . . . . .                                                                                                                                                                                                                                 | 136 |
| Selmer Bringsjord, Naveen Sundar Govindarajulu, Shreya Banerjee,<br>and John Hummel                                                                                                                                                                                                  |     |
| <b>Where Intelligence Lies: Externalist and Sociolinguistic Perspectives on the Turing Test and AI</b> . . . . .                                                                                                                                                                     | 158 |
| Shlomo Danziger                                                                                                                                                                                                                                                                      |     |
| <b>Modelling Machine Learning Models</b> . . . . .                                                                                                                                                                                                                                   | 175 |
| Raül Fabra-Boluda, Cèsar Ferri, José Hernández-Orallo,<br>Fernando Martínez-Plumed, and M. José Ramírez-Quintana                                                                                                                                                                     |     |
| <b>Is Programming Done by Projection and Introspection?</b> . . . . .                                                                                                                                                                                                                | 187 |
| Sam Freed                                                                                                                                                                                                                                                                            |     |
| <b>Supporting Pluralism by Artificial Intelligence: Conceptualizing Epistemic Disagreements as Digital Artifacts</b> . . . . .                                                                                                                                                       | 190 |
| Soheil Human, Golnaz Bidabadi, and Vadim Savenkov                                                                                                                                                                                                                                    |     |
| <b>The Frame Problem, Gödelian Incompleteness, and the Lucas-Penrose Argument: A Structural Analysis of Arguments About Limits of AI, and Its Physical and Metaphysical Consequences</b> . . . . .                                                                                   | 194 |
| Yoshihiro Maruyama                                                                                                                                                                                                                                                                   |     |
| <b>Quantum Pancomputationalism and Statistical Data Science: From Symbolic to Statistical AI, and to Quantum AI</b> . . . . .                                                                                                                                                        | 207 |
| Yoshihiro Maruyama                                                                                                                                                                                                                                                                   |     |
| <b>Getting Clarity by Defining Artificial Intelligence—A Survey</b> . . . . .                                                                                                                                                                                                        | 212 |
| Dagmar Monett and Colin W. P. Lewis                                                                                                                                                                                                                                                  |     |
| <b>Epistemic Computation and Artificial Intelligence</b> . . . . .                                                                                                                                                                                                                   | 215 |
| Jiří Wiedermann and Jan van Leeuwen                                                                                                                                                                                                                                                  |     |

|                                                                                           |            |
|-------------------------------------------------------------------------------------------|------------|
| <b>Will Machine Learning Yield Machine Intelligence? . . . . .</b>                        | <b>225</b> |
| Carlos Zednik                                                                             |            |
| <b>Ethics - Law</b>                                                                       |            |
| <b>In Critique of RoboLaw: The Model of SmartLaw . . . . .</b>                            | <b>231</b> |
| Paulius Astromskis                                                                        |            |
| <b>AAAI: An Argument Against Artificial Intelligence . . . . .</b>                        | <b>235</b> |
| Sander Beckers                                                                            |            |
| <b>Institutional Facts and AMAs in Society . . . . .</b>                                  | <b>248</b> |
| Arzu Gokmen                                                                               |            |
| <b>A Systematic Account of Machine Moral Agency . . . . .</b>                             | <b>252</b> |
| Mahi Hardalupas                                                                           |            |
| <b>A Framework for Exploring Intelligent Artificial Personhood . . . . .</b>              | <b>255</b> |
| Thomas B. Kane                                                                            |            |
| <b>Against Leben's Rawlsian Collision Algorithm<br/>for Autonomous Vehicles . . . . .</b> | <b>259</b> |
| Geoff Keeling                                                                             |            |
| <b>Moral Status of Digital Agents: Acting Under Uncertainty . . . . .</b>                 | <b>273</b> |
| Abhishek Mishra                                                                           |            |
| <b>Friendly Superintelligent AI: All You Need Is Love . . . . .</b>                       | <b>288</b> |
| Michael Prinzing                                                                          |            |
| <b>Autonomous Weapon Systems - An Alleged Responsibility Gap . . . . .</b>                | <b>302</b> |
| Torben Swoboda                                                                            |            |
| <b>Author Index . . . . .</b>                                                             | <b>315</b> |