Philosophical Studies Series

Volume 143

Philosophical Studies Series aims to provide a forum for the best current research in contemporary philosophy broadly conceived, its methodologies, and applications. Since Wilfrid Sellars and Keith Lehrer founded the series in 1974, the book series has welcomed a wide variety of different approaches, and every effort is made to maintain this pluralism, not for its own sake, but in order to represent the many fruitful and illuminating ways of addressing philosophical questions and investigating related applications and disciplines.

The book series is interested in classical topics of all branches of philosophy including, but not limited to:

- Ethics
- Epistemology
- Logic
- Philosophy of language
- Philosophy of logic
- Philosophy of mind
- Philosophy of religion
- Philosophy of science

Special attention is paid to studies that focus on:

- the interplay of empirical and philosophical viewpoints
- the implications and consequences of conceptual phenomena for research as well as for society
- philosophies of specific sciences, such as philosophy of biology, philosophy of chemistry, philosophy of computer science, philosophy of information, philosophy of neuroscience, philosophy of physics, or philosophy of technology; and
- contributions to the formal (logical, set-theoretical, mathematical, information-theoretical, decision-theoretical, etc.) methodology of sciences.

Likewise, the applications of conceptual and methodological investigations to applied sciences as well as social and technological phenomena are strongly encouraged.

Philosophical Studies Series welcomes historically informed research, but privileges philosophical theories and the discussion of contemporary issues rather than purely scholarly investigations into the history of ideas or authors. Besides monographs, *Philosophical Studies Series* publishes thematically unified anthologies, selected papers from relevant conferences, and edited volumes with a well-defined topical focus inside the aim and scope of the book series. The contributions in the volumes are expected to be focused and structurally organized in accordance with the central theme(s), and are tied together by an editorial introduction. Volumes are completed by extensive bibliographies.

The series discourages the submission of manuscripts that contain reprints of previous published material and/or manuscripts that are below 160 pages/88,000 words.

For inquiries and submission of proposals authors can contact the editor-in-chief Mariarosaria Taddeo via: mariarosaria.taddeo@oii.ox.ac.uk

More information about this series at https://link.springer.com/bookseries/6459

Björn Lundgren • Nancy Abigail Nuñez Hernández
Editors

Philosophy of Computing

Themes from IACAP 2019

Springer

Editors
Björn Lundgren
The Department of Philosophy and Religious Studies
Utrecht University
Utrecht, The Netherlands

Institute for Futures Studies
Stockholm, Sweden

Department of Philosophy
Stockholm University
Stockholm, Sweden

Nancy Abigail Nuñez Hernández
Institute of Philosophy
Czech Academy of Sciences
Prague, Czech Republic

ISSN 0921-8599 ISSN 2542-8349 (electronic)
Philosophical Studies Series
ISBN 978-3-030-75266-8 ISBN 978-3-030-75267-5 (eBook)
https://doi.org/10.1007/978-3-030-75267-5

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

This book is dedicated to the memory of Gunilla Lundgren (June 20, 1954–October 6, 2020).

Foreword

In the late 1970s, when I was just starting to think about the ethical issues related to computers, I was struck by how little was being written on the topic and how shallow much of it was. I remember thinking that the discourse could benefit greatly from philosophical concepts and deeper philosophical analysis. At the same time, a small number of philosophers turned their attention to the work of Herbert Simon and Norbert Wiener, seeing in the work of these two big thinkers a way to understand the significance of modern computers. These philosophers thought that computation had enormous implications for, and might even resolve, many fundamental philosophical issues. I was especially taken with Simon's statement that: "Perhaps the most important question of all about the computer is what it has done and will do to man's view of himself and his place in the universe" (Simon, H.A. 1977. "What computers mean for man and society." *Science* 195(4283): 1186-1191, p. 1190). Also, during this time, a small number of philosophers began to see in computers the opportunity to teach logic in a better way than it had ever been taught before; they began developing software for this purpose.

While modern computers attracted the interest of this small group of philosophers, in those early days and for the next several decades, very few scholars in the humanities and social sciences recognized the significance of modern computers and the prescience of Simon's statement. That small group of philosophers was the exception. In the mid-1980s, with an abundance of foresight, they hatched the idea of annual workshops and conferences focused on the connections between philosophy and computing. These meetings brought together computer scientists with philosophers from a variety of fields within philosophy. Over the years, the conferences grew and took various forms; they began as small local events and grew to international events. Eventually, the activities and the organization that made them happen morphed into what is now known as the International Association for Computing and Philosophy (IACAP).

The papers gathered together in this volume are from the 2019 annual conference of IACAP. They testify to the rich and robust territory created by bringing philosophy and computers together. The papers illustrate how traditional philosophical questions have been expanded and extended as a result of the combination. Although it is tempting to seek a broad yet pithy account of the connection between philosophy and computers, a simple or singular account is not possible. Philosophy and computing have a myriad of connections.

The *mere idea* of computers, as suggested in Simon's statement, has had an influence. It has, for example, presented novel ontological questions. Philosophers are now asking whether it would ever be possible (or even obligatory) to recognize robots as persons with rights comparable to those of humans. The question arises because the future trajectory of robot technology seems to point in the direction of robots that, although silicon-based or a form of artificial life, will think and act in ways that could be considered equivalent to human thinking and behavior. There is even research underway to produce behavior in robots that would manifest as something like human emotions. These possibilities have led philosophers to look at traditional philosophical notions differently, that is, to rethink such notions as personhood, equivalence, and ontological status. Grappling with the ontological and ultimately the moral status of robots is evident in this volume in a chapter that asks, should robots be allowed to punish us?

Computers are not, of course, just an idea, they are machines that operate in the physical world. Indeed, since the invention of modern computers, their use has come to permeate many, if not most, domains of human activity. Because of the use of computers in science, the fundamental question of epistemology – how do we know? – has been taken in new directions. Several of the chapters in this volume take up issues relating to computationally produced forms of knowledge in addressing computer simulation, computer-assisted proofs, and simplicial complexes.

Because computers operate in the physical world, and especially because they change how we relate to one another and how social institutions are constituted, an endless array of ethical questions and concerns have been raised. Although ethics has to do with how people behave, computers affect when, where, and how we behave and interact with one another. The use of computers affects individual behavior as well as the nature of human organizations and institutions. Hence, the issues, especially the situations and cases, taken up by moral philosophy must – to be relevant to the experience of those alive today – address a computerized world. The chapter on predictive fairness in this volume is a good example of this. There is a long and deep tradition of philosophical thinking about fairness and the capacity of computers to combine and process huge quantities of data and to use the results to make fine-grained predictions about human behavior has required the extension of notions of fairness to the use of these new and powerful predictive computations.

The chapters in this volume extend the long tradition of philosophical thinking by bringing computers and computing into the conversation. The volume carries forward conversations that began centuries ago and are now infused with the idea of powerful computers and the experience of living in a world in which they operate.

University of Virginia, Charlottesville, VA, USA Deborah G. Johnson
October 2020

Preface

Alan Turing's "On computable numbers, with an application to the Entscheidungsproblem" (1936) and "Computing Machinery and Intelligence" (1950) are among the foundational works on computer science and artificial intelligence. Any reader of these papers can see that Turing confronted deep philosophical worries. Since then, computer science and artificial intelligence have developed at a blistering pace with a tendency to highly technical specialization, growing them apart from philosophy. However, the legacy of Turing should remind us of the significance of bringing together philosophy, computer science, and artificial intelligence. The chapters contained in this volume represent efforts in that fertile direction that seeks to foster collaborations between philosophers, computer scientists, and experts on artificial intelligence.

This book is the official proceedings of the 2019 meeting of the *International Association for Computing and Philosophy* (IACAP). Since 1986, IACAP has organized either international or regional conferences yearly. So far, 2020 is the only exception.[1] The 2019 conference was arranged in Mexico City at the *National Autonomous University of Mexico* by Nancy Abigail Nuñez Hernández, Sergio Rajsbaum, and Luis Estrada. The conference included presentations on a large variety of topics and a special track on Distributed Computing and Epistemic Logic arranged by Sergio Rajsbaum and Alexandru Baltag.

In preparing this volume, we invited all authors presenting at the conference to submit a manuscript for consideration. Among those that opted to submit, seven papers were selected after peer review. Beyond these selected papers, the volume also includes one invited contribution from each of the prize winners for IACAP's two annual awards: *The Covey Award* and *The Herbert A. Simon Award*.

Since 2009, the Covey Award "recognizes senior scholars with a substantial record of innovative research in the field of computing and philosophy broadly

[1] See: https://www.iacap.org/conferences/. At the time of writing this Introduction, the next IACAP conference, in 2021, is currently planned to be arranged *online* by the *University of Hamburg*. It is a collaboration with *CEPE* (*Computer Ethics—Philosophical Enquiry*).

conceived."[2] In 2019, the Covey Award went to John Weckert (*Charles Sturt University*) for his "broad and deep influence on the ethics of information technology" and "the significant impact John has had fostering by both example and personal influence the careers of many scholars who have continued following the fruitful paths of research his work brought to light" (Berkich 2019a).[3]

Since 2010, the Herbert A. Simon Award for Outstanding Research in Computing and Philosophy "recognizes scholars at an early stage of their academic career who are likely to reshape debates at the nexus of Computing and Philosophy by their original research."[4] In 2019, the Simon Award went to Juan M. Durán (Delft University of Technology) "for his emphasis on the epistemic and moral normative dimensions of computational methods in the sciences, representing transformational and seminal scholarship in the philosophy of science." (Berkich 2019b).[5]

The themes of the nine chapters of this book are diverse. The themes include epistemology, epistemic logic, computational logic, ethics, applied ethics, algorithmic fairness, philosophy of technology, human-robot interaction, philosophy of science, and pedagogy, which illustrate the breadth of the topics that are commonly dealt with at an IACAP conference. Below we briefly introduce each of the chapters and raise some questions.

Current technological developments show that it is more pressing than ever to unify philosophers and computer scientists' efforts to understand knowledge in its full complexity. In Chap. 1, "Knowledge and Simplicial Complexes," Hans van Ditmarsch, Eric Goubault, Jérémy Ledent, and Sergio Rajsbaum offer an interesting example of how bringing together philosophical developments on dynamic epistemic logic with topological insights from simplicial complexes and distributed computing sheds new light on core questions about knowledge and belief, like the notion of consensus.

van Ditmarsch et al. offer alternative ways to model epistemic notions such as knowledge and belief, as well as group notions such as mutual, common,

---

[2] http://www.iacap.org/awards/

[3] Previous winners are (chronologically ordered, most recent winner first): Deborah G. Johnson (*University of Virginia*), Raymond Turner (*University of Essex*), Jack Copeland (*University of Canterbury*), William J. Rapaport (*University at Buffalo*, *The State University of New York*), Selmer Bringsjord (*Rensselaer Polytechnic Institute*), Margaret Boden (*University of Sussex*), Luciano Floridi (*University of Hertfordshire*), Terrell Bynum (*Southern Connecticut State University*), John R. Searle (*University of California*, *Berkeley*), and Edward N. Zalta (*Stanford University*). See: http://www.iacap.org/awards/

[4] http://www.iacap.org/awards/

[5] Previous winners are (chronologically ordered, most recent winner first): Thomas C. King (*Oxford Internet Institute*), Andrea Scarantino (*Georgia State University*), Marcin Milkowski (*The Institute of Philosophy and Sociology of the Polish Academy of Sciences*), Michael Rescorla (*University of California-Santa Barbara*), Gualterio Piccinini (U*niversity of Missouri–St. Louis*), Judith Simon (*University of Vienna*), Patrick Allo (*Vrije Universiteit Brussels*), John Sullins (*Sonoma State University*), and Mariarosaria Taddeo (*University of Hertfordshire*; *University of Oxford*). See: http://www.iacap.org/awards/

and distributed knowledge on simplicial complexes.[6] The appeal to algebraic topological tools like simplicial complexes has shown fertile explanatory advantages in diverse research areas, like graph theory (Lovász 1978), neuroscience (Reimann et al. 2017), and semantic networks (Salnikov et al. 2018; Christianson et al. 2020), among many others. For philosophers unfamiliar with algebraic topology, simplicial complexes' potential applications to epistemology may go unnoticed until the duality between simplicial complexes and Kripke models is shown, which is precisely the departure point of this chapter. Roughly speaking, if a Kripke model is viewed as a simplicial complex, we can use algebraic topology tools to study that structure's higher dimensional properties. For instance, the authors state that topological invariants are preserved when the system evolves after communication, and these topological invariants determine the problems that agents can solve. To sum up, the chapter is a contribution that provides novel insights for computer scientists interested in distributed computing and philosophers interested in dynamic epistemic logic, group knowledge, and the development of more accurate formal tools to model epistemic phenomena.

The development of computational technology has allowed us to process large quantities of data, to the point that big data, machine learning, and other data-driven methods have become essential to many scientific enterprises. This achievement comes, perhaps, at a cost: it has been claimed that data per se does not contribute to developing science, and that scientific theories should guide experimental design to efficiently collect data and produce reliable predictive models and conceptual knowledge. In Chap. 2, "Meta-Abduction: Inference to the Probabilistically Best Prediction," Christian J. Feldbacher-Escamilla suggests one way to help bridge the gap between vast troves of algorithmically massaged data and the predictive and explanatory goals of scientific theory by offering a characterization and justification for abduction as an inference to the probabilistically best prediction.

Feldbacher-Escamilla presents a taxonomy of abductive inferences to focus on a species of selective abduction, which aims at inferring the probabilistically best hypothesis, explanation, or theory based on data. Moreover, Feldbacher-Escamilla develops an account of meta-abduction as inference to the best probabilistic explanation, applying Schurz's theory of meta-induction and Feldbacher-Escamilla's framework of prediction games to show how such species of abduction can be justified. This theoretical contribution is especially relevant given the current "need to turn data into true predictions, that is, predictions of events in novel circumstances, or predictions of events before they occur (not *post hoc* explanations)." (Coveney et al. 2016)

[6] Simplicial complexes are made of simplices. Roughly speaking, a simplex can be understood as a generalized triangle (a triangle in any dimension): a 0 simplex is a single vertex, a 1 simplex is two vertices connected by an edge, a 2 simplex is three vertices connected pairwise by edges with a single face (a triangle), and a 3 simplex is four vertices connected pairwise by edges joint by four faces which are filled in to form a solid (a tetrahedron). Hence the duality between simplicial complexes and Kripke models which tend to be graphically represented through vertices connected by edges.

Big data methods in science raise further concerns. There are research ethical worries (such as about P-hacking), which often also raises epistemic worries (can we trust the results?).[7] In Chap. 3, "Is There Anything Special About the Ignorance Involved in Big Data Practices?", María del Rosario Martínez-Ordaz addresses epistemic challenges of big data. Martínez-Ordaz starts off by discussing some preliminary methodological worries relating to how big data is often used to find correlations and the worry about determining whether we can "rationally trust correlations as legitimate instances of scientific knowledge" (p. 117). She argues that since it is impossible for a human agent to process a big data set, we become "necessarily constrained by specific technological resources" (p. 117). Moreover, she argues that technological dependence is a source of epistemic opacity. Next, she connects these methodological worries to related worries about knowledge, before she turns to the main issue of the chapter: ignorance.

Martínez-Ordaz supplies a taxonomy of different types of or sources of ignorance (factual ignorance, objectual ignorance, procedural ignorance, and ignorance of theoretical structures; pp. 123–124) and argues that the main problem with big data has to do with ignorance of theoretical structure, but that there are two sources of reliable knowledge: objectual knowledge (i.e., knowledge of particular objects) and modal knowledge (which we take here to be knowledge of the possibility space) of "how these objects (could) behave and relate to one another" (p. 137). Based on this analysis, Martínez-Ordaz ends her chapter with a case-study on cosmology and big data.

Some of her negative arguments may meet resistance. For example, technological dependence is arguably not something that is special for big data. Most, if not all, empirical studies depend on instruments, from the mundane (a magnifying glass, say) to the technically more complex (an electron microscope, for example), and it would be challenging to find an empirical study in which there is no related dependence. Here it is worth to point out that Jumbly Grindrod (2019) has recently argued that algorithms function like "instrumental knowledge" (cf. Sosa 2006). Moreover, it is possible to construct big data sets with proven validity, so that while the scientist cannot process the data set themself, they can still know that the data set is valid.

As a final point, we believe that a fruitful way to develop her main points would be to introduce a measurement of ignorance and opacity, since the problem arguably is not whether there is or is not some ignorance and opacity, but how pervasive ignorance and opacity are.

Going back at least to the Four Color Theorem, mathematicians have wrestled with whether computer assisted proofs, which may have no traditional "paper-and-pencil" equivalent, can be counted as proofs. Favio E. Miranda-Perea and Lourdes

[7] P-hacking is the process of looking for correlations in a data set to find data points that would manifest as statistically significant. Since a large data set will—due to random variations—include correlations, we cannot apply a significance test to see if it is likely that those correlations are due to random variation or a real connection.

del Carmen González Huesca discuss this topic in Chap. 4, "On the Conciliation of Traditional and Computer-Assisted Proofs." In this work, Miranda-Perea and González Huesca elaborate on the concept of transitional proofs, which intends to bridge computer-assisted proofs and traditional paper-and-pencil proofs. This proposal is instantiated through the use of Coq, a powerful and popular proof assistant. Those unfamiliar with computer-assisted proofs can benefit from the detailed explanations in this chapter; however, various mathematicians, computer scientists, and software engineers specialized in automated reasoning may find some claims controversial. Particularly, novelty claims on this work's contributions might be contested, for example, the formalization of backward proofs, induction on assumption contexts, or the discovery of superior correctness proofs for specific algorithms.

Implicit bias is a well-known problem for algorithmic decision-making. In various ethical guidelines for AI (see, e.g., AI HLEG 2019; Brey et al. 2019) fairness is a key requirement for ethical AI. In Chap. 5, "Predictive Fairness," Anders Herlitz discusses the issue of algorithmic fairness. He starts by presenting a discussion about the fairness of the algorithm COMPAS (the Correctional Offender Management Profiling for Alternative Sanctions)—a tool used by US courts to determine an offender's recidivism rate. Herlitz notes that there has been disagreement about whether COMPAS is biased against blacks or not and that this disagreement is rooted in different conceptions of fairness, what Herlitz calls "Equality in accuracy" and "Equality in inaccuracy."

Kleinberg et al. (2017) have shown that these two conceptions of fairness—in Kleinberg et al. one of these conceptions corresponds to two separate conditions—cannot both be satisfied if the recidivism rate differs between two salient social groups. As Herlitz sums it up: "Either (i) the decision method correctly identifies the relevant property (e.g. recidivism) more often in one subgroup (e.g. black defendants) than another (e.g. white defendants); or (ii) the decision method systematically ascribes higher probabilities to individuals who have and/or individuals who lack the property in one group compared to the probabilities ascribed to individuals who have and/or individuals who lack the property in another group" (p. 141). In the second section, Herlitz turns to discuss three possible responses: dismiss one of the conditions or conceptions of fairness, address the unfairness at another stage of the process, or reject algorithmic decision-tools altogether. Herlitz finds all three options problematic. Thus, in the third section, he suggests three principles that ought to "guide policy that rely on systematic decision methods that generate probability predictions: *Dominance, Transparency* and *Priority to the worse off*" (p. 156). Simply put, Dominance is satisfied if one method is better than another relative to all relevant (fairness) factors and worse relative to none; Transparency demands awareness and transparency from the user of the unfair decision-method; and Priority to the worse off says that unfairness should benefit the group that is worst off.

We find these suggestions sound and helpful, but in practice they may require further operationalization. For example, as Herlitz indicates, Dominance implies a goal to minimize unfairness. However, we would have to specify further how to choose between methods that violate some fairness principles very little and rarely, and those that violate only one principle but do so more extremely and very often.

Robotic applications are breaking new grounds. Techniques are developing for robots in such diverse areas as health care, child care, elderly care, and the transport industry, well beyond their traditional domain in industrial manufacturing. This raises various issues of the role of robots in the society and especially when robots are asked not only to perform a limited task with objects, but to engage in a social environment. In Chap. 6, "Castigation by Robot: Should Robots Be Allowed to Punish Us?", Alan R. Wagner and Himavath Jois offer an introductory overview to the topic of robot punishment. Wagner and Jois apply a broad definition of punishment, according to which it is a behavior "selected for the purpose of adding cost to one's action" (p. 164). Under this definition, the idea that robots should be allowed and need to administer punishment is not surprising, as they note even a warning sound may be a punishment. Wagner and Jois argue that if robots perform substantial roles and tasks in society, the ability to punish (even if only minor) may be central to maintain the relevant authority needed to perform those roles and tasks. However, Wagner and Jois also argue that it is important that robots "support the existing system of social norms," even bad norms—because it is important to maintain consistency, by which we suppose they mean social order (p. 168). We find this problematic. First, we question the empirical claim that maintaining social norms maintains social order or consistency. Indeed, because social norms depend on second-order beliefs (i.e., beliefs about others' normative preferences), it follows that there are social norms that are upheld against a super majority's will (see, e.g., Bicchieri 2017 for empirical examples). Thus, some social norms are upheld because many or even most people hold false second-order beliefs. Hence, it is not clear that maintaining such norms would maintain consistency. Second, and more importantly, we worry that norm adherence can be abused to maintain oppressive norms (e.g., gender mutilation) or that claims about the necessity of social norms can be used to retain undemocratic powers and deny people their human rights. Perhaps Wagner and Jois can temper their claim about norms by reformulating their principle of norm adherence in terms of a *prima facie* or *pro tanto* principle focusing on good or generally accepted norms.

In Chap. 7, "Implementing Algorithmic and Computational Design in Philosophical Pedagogy," Rocco Gangle argues that algorithmic reasoning—or, what he calls "Diagrammatic Methods"—can be used to teach introductory philosophy courses. The pedagogical tools he envisions using might best be described as diagrams illustrating the different relations between arguments and premises. The idea here is simply to use graphical diagrammatic structures to show and analyze these relations. It is an interesting idea that may help students struggling with more formal expressions. However, we wonder to what extent the diagrams can be applied to all arguments even in an introductory course. While that may be a learning experience, because of the—as Gangle notes— "the self-reflexive logical aspect of philosophy" (p. 181), the value of such a learning experience arguably diminishes over time. Nevertheless, the manuscript contains several interesting examples (with detailed illustrations), which many teachers likely will find interesting to try out themselves.

Gangle also suggests that the diagrammatic representation may also be used for more sophisticated philosophical reasoning and research. Ultimately, the proof is in the pudding (or, for the pedagogical claims, the proof is in a well-designed experiment).

According to Moore's law the number of transistors in a microchip will double every second year. This forecast has held steady since the 1970s, and even if some expect the number of transistors to flatten (Shalf 2020; Waldrop 2016), it is fair to say that this development has had tremendous implications. With more and more devices becoming "smart" (phones, TVs, home appliances, etcetera), technology is developing in new ways. But is this development sound?

In Chap. 8, "Our Technology Fetish," John Weckert argues that we (humans) have a technological fetish, which he argues is bad, and we ought to slow down and reflect on technological development, rather than push to the next frontier. Weckert discusses many aspects and problems relating to technology, including surveillance, control, and risk. One might ask, however, is the supposed fetish really a fetish about the technological device as a technological artifact or an accessory? It seems that many desire the latest devices not necessarily for what they offer, as technological artifacts, but because of what they signify—in the same way as a new pair of sneakers or a handbag of the right brand signifies status in the relevant reference frame. If that is the case, then is that really a *technological* fetish?

The nature and understanding of computer simulations have been the subject of great interest in a wide range of fields: software engineering, computer science, mathematics, and philosophy of science—including philosophy of computer science and philosophy of mathematics. In Chap. 9, "Models, Explanation, Representation, and the Philosophy of Computer Simulations," Juan M. Durán offers a survey of the central positions within the interdisciplinary debate on computer simulations and argues that these simulations are units of analysis on their own and cannot be reduced to mathematical models. Thus, Durán's proposal approaches simulations as a branch of computer science and software engineering.

To underpin his defense of computer simulations as distinctive units of analysis, Durán relies on examples that aim to show how the representation of a target system through a computer simulation differs from its representation through a set of equations; the syntax of simulation models uncovers these differences. However, these claims are controversial because, so far, it is unclear that computer simulations actually exceed the expressive power of mathematical languages and models. The controversial nature of Durán's central claims serves as an invitation to discuss computer simulations' role and value within scientific theories and practices.

Lastly, this book would not have been possible without the help of a large number of people. We want to thank Deborah G. Johnson for writing the foreword. We also want to thank all reviewers for their help. While the reviewing was double-blinded, we publish the names of the reviewers consenting to be named below (some reviewers preferred to retain their anonymity also at this stage): Adam Chlipala, Alan Bundy, Aybüke Özgün, Barbro Fröding, Chantal Keller, Charles Rathkopf, Claus Beisbart, Cosmo Grant, Daniel Lim, Don Berkich, Dorna Behdadi, Eckhart Arnold, Gordana Dodig-Crnkovic, Göran Collste, John Symons, Jorge Manero,

Juan M. Durán, Kye Palider, Kyrill Winkler, Lenin Vazquez-Toledo, Marc Jiménez-Rolland, Martin Frické, Matthieu Fontaine, Michael Weisberg, Nicola Angius, Orlin Vakarelov, Otávio Bueno, Pak-Hang Wong, Patricia Mirabile, Paul Humphreys, Rasmus K. Rendsvig, Ryan Jenkins, Soraj Hongladarom, Stefan Buijsman, Stefano Canali, Stefano Vincini, Steve McKinlay, and Yoav Meyrav.

Mexico City, Mexico — Nancy Abigail Nuñez Hernández
Stockholm, Sweden — Björn Lundgren

References

AI HLEG (High-Level Expert Group on Artificial Intelligence). 2019. *Ethics guidelines for trustworthy AI*. https://ec.europa.eu/futurium/en/ai-alliance-consultation.

Berkich, D. 2019a. 2019 Covey Award: Emeritus Professor John Weckert. *IACAP announcement on PHILOS-L email list.* https://listserv.liv.ac.uk/cgi-bin/wa?A2=PHILOS-L;f0e8d75e.1906.

Berkich, D. 2019b. 2019 Simon Award: Dr. Juan M. Durán. *IACAP announcement on PHILOS-L email list.* https://listserv.liv.ac.uk/cgi-bin/wa?A2=PHILOS-L;5222a460.1906.

Bicchieri, C. 2017. Norms in the wild: how to diagnose, measure, and change social norms. New York: Oxford University Press.

Brey, P., B. Lundgren, K. Macnish, and M. Ryan. 2019. Guidelines for the development and the use of SIS. *Deliverable D3.2 of the SHERPA project*. https://doi.org/10.21253/DMU.11316833.v2.

Christianson, N.H., A. Sizemore Blevins, and D.S. Bassett. 2020. Architecture and evolution of semantic networks in mathematics texts. *Proceedings of the Royal Society A* 476(2239): 20190741. https://doi.org/10.1098/rspa.2019.0741.

Coveney, P.V., E.R. Dougherty, & R.R. Highfield. 2016. Big data need big theory too. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374(2080): 20160153. https://doi.org/10.1098/rsta.2016.0153.

Grindrod, J. 2019. Computational beliefs. *Inquiry*. https://doi.org/10.1080/0020174X.2019.1688178.

Kleinberg, Joel, Sendhil Mullainathan, and Manish Raghavan. 2017. *Inherent Trade-Offs in the Fair Determination of Risk Scores*. arXiv:1609.05807v2.

Lovász, L. 1978. Kneser's conjecture, chromatic number, and homotopy. *Journal of Combinatorial Theory*, Series A, 25(3): 319–324. https://doi.org/10.1016/0097-3165(78)90022-5.

Reimann, M.W., M. Nolte, M. Scolamiero, K. Turner, R. Perin, G. Chindemi, P. Dłotko, R. Levi, K. Hess, and H. Markram. 2017. Cliques of neurons bound into cavities provide a missing link between structure and function. *Frontiers in computational neuroscience*, 11, 48. https://doi.org/10.3389/fncom.2017.00048.

Salnikov, V., D. Cassese, R. Lambiotte, and N.S. Jones. 2018. Co-occurrence simplicial complexes in mathematics: Identifying the holes of knowledge. *Applied Network Science* 3, 37. https://doi.org/10.1007/s41109-018-0074-3.

Shalf, J. 2020. The future of computing beyond Moore's Law. *Philosophical Transactions of Royal Society A* 37820190061. https://doi.org/10.1098/rsta.2019.0061.

Sosa, E. 2006. Knowledge: Instrumental and testimonial. In *The epistemology of testimony,* ed. J. Lackey and E. Sosa, 116–123. Oxford: Oxford University Press.

Turing, A.M. 1936. On computable numbers, with an application to the Entscheidungsproblem. In *Proceedings of the London Mathematical Society* S2–42: 230–265. [Published 1937]. https://doi.org/10.1112/plms/s2-42.1.230.

Turing, A.M. 1950. Computing machinery and intelligence. *Mind* LIX(236): 433–460. https://doi.org/10.1093/mind/LIX.236.433.

Waldrop, M.M. 2016. The chips are down for Moore's law. *Nature* News feature 530: 144–147. https://doi.org/10.1038/530144a.

Contents