Computer Modeling in Philosophy

Editorial

Patrick Grim*

Editorial introduction to the Topical Issue "Computer Modeling in Philosophy"

https://doi.org/10.1515/opphil-2019-0049

The role played by logic in 20th century philosophy, it can be argued, will be played by computational modeling in the 21st. This special issue is devoted to discussion, analysis, but primarily to *examples* of computer-aided or computer-instantiated modeling. Over the past several decades, social epistemology and philosophy of science have been important areas in the development of computational philosophy.[1] Here we focus on current work in a wider spread of sub-disciplines: ethics, social philosophy, philosophy of perception, philosophy of mind, metaphysics and philosophy of religion.

The first two pieces in the collection concentrate on computational techniques and philosophical methodology quite generally.

Istvan Berkeley's "The Curious Case of Connectionism" opens the collection, with an examination and analysis of three stages in the history of a major theoretical approach that continues in contemporary computational philosophy. He characterizes a first stage of connectionism as ending abruptly with the critique by Minsky and Papert.[2] A second stage had an important impact on philosophy, but Berkeley documents its waning influence with the declining appearance of the terms 'connectionism' and 'connectionist' in the *Philosopher's Index*. He proposes deep learning as a third stage of connectionism, with new computational technologies promising the possibility of important philosophical application.

The search for formal methods of inquiry and discovery, as opposed to mere justification, can be seen historically as a project in Aristotle, Bacon, Leibniz, and Mill. But in the 20th century, at the hands of Popper, Reichenbach, Rawls, and others, that search was largely abandoned. In "The Evaluation of Discovery: Models, Simulation and Search through 'Big Data'," Joseph Ramsey, Kun Zhang, and Clark Glymour argue that the contemporary development of algorithms for search through big data offers a rebirth for formal methods of discovery. The authors point out that search algorithms also pose a major problem of validation, however. What we want is output with both 'precision,' a high probability that the hypotheses it returns are true, and 'recall,' the probability that the true hypotheses are returned. How are we to assess precision and recall for causal relations if, as in many cases, our data base is huge, the potential correlations are many, but the empirical base available for direct assessment is vanishingly small? Here recourse is often made to modeling or simulation, assessing a search method using not actual data but data simulated from known patterns, sub-structures, or 'motifs' within a domain. Ramsey, Zhang, and Glymour illustrate the approach with two cases, one from neuroscience and another from astrophysics. They emphasize the inherent risk in the simulated data strategy, particularly in cases in which sample selection is not automated and not guaranteed to be representative. In specific contexts, with appropriate safeguards, careful application of search algorithms

1 See for example Zollman, "The Communication Structure of Epistemic Communities," Weisberg and Muldoon, "Epistemic Landscapes and the Division of Cognitive Labor," Hegselmann and Krause, "Truth and Cognitive Division of Labour: First Steps in a Computer Aided Social Epistemology," Grim and Singer et al., "Scientific Networks on Data Landscapes: Question Difficulty, Epistemic Success, and Convergence," and O'Connor and Bruner, "Dynamics and Diversity in Epistemic Communities."
2 Minsky and Papert, *Perceptrons*.

***Corresponding author: Patrick Grim,** Stony Brook University, New York / University of Michigan, Ann Arbor, United States of America; E-mail: patrick.grim@stonybrook.edu

does offer new promise for formalized discovery, they conclude, though "the ambition to find a single method for all problems of empirical inquiry has not been realized, and we do not expect it to be."

Other articles in this issue are collected in terms of topic area: social philosophy and ethics, philosophy of mind and perception, metaphysics and philosophy of religion.

Computational techniques of agent-based modeling are represented in two pieces in social philosophy and ethics.

Spite harms others, though at a cost to the agent. The question then is how spite can evolve—which it demonstrably does, in forms evident from bacteria through birds to humans. Fluctuating populations that may hover on the brink of extinction are a key element in the agent-based model that Rory Smead and Patrick Forber build to explore the question in "Signals and Spite in Fluctuating Populations." Their results show that conditional spite, signaled by pre-play signals, can easily evolve in the context of stochastic population fluctuations. Although carefully qualified, they also suggest that such a dynamic may form a primitive basis for the origin of punishment. Smead and Forber close with some important reminders regarding the limitations of agent-based modeling: though simulations allow the possibility of revealing new explanations, they also allow the possibility of modeling artifacts and over-complicated models no easier to understand than the target phenomenon itself.

In "Towards Computer Simulations of Virtue Ethics," Jeremiah Lasquety-Reyes outlines two approaches. In a simple approach, one might represent an agent's virtue by a single probability of behaving in a certain way and having that probability strengthened by use or 'habit.' Lasquety-Reyes develops a far more sophisticated version in which agents are embodied with interacting physical, cognitive, emotional, and social modules, in line with Schmidt and Urban's PECS model and Joshua Epstein's approach in *Agent_Zero*.[3] He builds simulations for a virtue of temperance—consumption in response to physical demands but in accord with both health and social norms—in an environment of food seeking and consumption that extends Axtell and Epstein's *Sugarscape*.[4] In Lasquety-Reyes' model, agent action reflects 'decision' on the basis of potential tension between internal physical, cognitive, emotion, and social elements. Within some parameters for sparsity and social contact, he is able to track the development of a virtue of temperance in an agent. With other parameters, he tracks a trajectory toward violation of social norms. The results raise questions both regarding the social dynamics of virtue and for concepts of virtue itself.

Three pieces in this special issue use computational techniques in the philosophy of perception and philosophy of mind.

Why do we categorize color as we do? Color universalists point to patterns of color categorization that appear to hold across different languages. Color relativists point to linguistic differences in color categorization. Debate continues regarding both proper methodology in analyzing the data and potential explanations for what the data shows. In "Towards More Realistic Modeling of Linguistic Color Categorization," José Correia and Radek Ocelák use computational techniques to construct a model of color categorization that reflects three major influences: fundamental physical characteristics of color perception, the dynamics of game-theoretic signaling between agents, and the probability of different colors in the experiential environment. Calibrating results against data from the World Color Survey on color terms in different languages,[5] with scrupulous attention to how the data is employed, the authors argue that the model they propose has predictive scores that match 'the state of the art,' while being both motivationally and explanatorily superior. They outline their work not as a finished product but as a baseline, an example of the form categorization theory should take: a concrete model that combines satisfactory explanation with data-accurate prediction.

Robert Pretner and Chris Fields attempt to join formal techniques from research in artificial intelligence with interpretations relevant to philosophy of perception and consciousness, exemplifying an interface intended to benefit both sides. Against a background outline of recurrent neural networks, cellular

---

3 Schmidt, *The Modeling of Human Behavior*; Urban, "PECS: A Reference Model for the Simulation of Multi-Agent Systems;" Epstein, *Agent_Zero: Toward Neurocognitive Foundations for Generative Social Science*.

4 Epstein and Axtell, *Growing Artificial Societies: Social Science from the Bottom Up*.

5 Kay, Berlin, Maffi, Merrifield and Cook, *The World Color Survey*.

automata, random Boolean networks, and agent-based modeling, they construct 'perceptual networks,' incorporating a number of features from other approaches with a specific psychological interpretation in mind. When can 'sensation' in the form of input to the nodes of a network be taken as something more: as the inferential construction of a representation that constitutes 'perception'? Technical questions in Pretner and Fields' "Using AI Methods to Evaluate a Minimal Model for Perception" include the developmental stability of their networks and genetic algorithm optimization of rules for specific agents within a handful of fixed networks. But their main targets remain philosophical throughout. When can the dynamics of such a network be interpreted as perceptual inference? When is it appropriate to think of such a network as a single agent? An overriding methodological goal is to illustrate with a specific example the potential, for both sides, of cross-fertilization between philosophy and research in artificial intelligence.

A number of advocates of the 'hard problem' have argued that work in artificial intelligence will inevitably prove irrelevant to the study of consciousness. Many AI researchers hold that the concept of consciousness remains too ill-defined to be of interest for AI. In "Modeling Working Memory to Identify Computational Correlates of Consciousness," James Reggia, Garrett Katz and Gregory Davis take the contrary view, attempting explicitly to use computational models in order to creep up, however gradually, on functional aspects of consciousness, carefully circumscribed. Their goal is to determine whether there are useful correlates of consciousness, with not phenomenal but functional access consciousness as the explicit target. On the assumption of working memory as a prime example, the authors examine two models of working memory: a neurocomputational model trained to solve simple card-matching problems, and a more traditional symbol-processing framework for robotics. The strategy is to identify high-level computational aspects in each of these that are functionally necessary. If these are necessary in order to model working memory, the argument goes, they will be necessary for working memory itself, and can thus hypothetically serve as correlates of the consciousness it represents. In the end, Reggia, Katz and Davis propose three such potential correlates: 'itinerant' attractor sequences, top-down gating as executive control, and very fast weight changes. Using the historical analogy of vitalism and the DNA demystification of a 'biophysical explanatory gap,' the authors speculate that "bridging the computational explanatory gap via the identification of computational correlates of consciousness may even eventually help with demystifying the hard problem in philosophy."

Computational metaphysics and philosophy of religion constitute the third topic area in this special issue.

In "Computer Science and Metaphysics: A Cross-Fertilization," Daniel Kirchner, Christoph Benzmüller and Edward Zalta survey past work and outline new results in the extension of theorem-proving techniques to issues in metaphysics and philosophy of religion. Prover9 has been used to prove theorems regarding situations and possible worlds, Plato's theory of Forms, and reconstructions of Anselm's ontological argument. An important recent extension has been to higher-order logics using 'shallow semantic embeddings,' with applications that include analysis of an inconsistency in the ontological argument that Gödel left behind (reproduced here) and consistent variations by Anthony Anderson and Melvin Fitting. Toward the end of the piece the authors outline ambitious and ongoing approaches to metaphysical theory in terms of computationally instantiated object theory. Although most of their piece concentrates on results from the application of computer science techniques to philosophy, the authors suggest that the specific techniques developed for philosophical purposes may have broader applications within computer science as well.

The application of theorem-proving techniques in philosophy of religion is further illustrated with Jack Horner's "A Computationally Assisted Reconstruction of an Ontological Argument in Spinoza's *The Ethics*." Using Prover9 and Mace4, Horner demonstrates that Spinoza's conclusion that God exists cannot be derived from his definitions and axioms. Horner then uses the same techniques to argue that a valid (and novel) argument *can* be constructed with the addition of further assumptions that Spinoza would accept.

An entirely different computational approach to philosophy of religion appears in F. LeRon Shults' "Computer Modeling in Philosophy of Religion." Shults' methodology is not that of theorem-provers but of agent-based modeling, with the social dynamics of religious belief as a target. He offers a model of mutually escalating religious violence between two groups as a first model. Progressive decline in religious belief in contemporary societies is the subject of a second. In both cases Shults is explicit in seeking validation of basic modeling mechanisms in empirical data and social psychological theory. The philosophy of religion that is represented in these examples, Shults argues, is importantly different from the apologetics of 'philosophy *for* religion.' Computational modeling of this form, he suggests, can lead us to a philosophical examination of religion more closely aligned with the psychological and social sciences.

We ordinarily think of the world in accord with Sellars' 'manifest image,' populated with ordinary middle-sized dry goods: chairs, tables, mountains, and stars. But contemporary physics seems to tell us that these things don't exist at all, as exemplified in the title of James Ladyman and Don Ross's *Every Thing Must Go: Metaphysics Naturalized*. In "What Simulations Teach Us About Ordinary Objects," the last piece in this special issue, Arthur Schwaninger offers an outline of the conceptual conflict and offers a resolution in terms of simulations from computational neuroscience. Those simulations seem to show our categorization of ordinary objects to be a predictable and evolutionarily advantageous result of the cognitive processing necessary in order to navigate the world of our sensory experience. Objects, one might say, though not metaphysically necessary, are necessary for cognitive beings like us. Computational simulations, Schwaninger proposes, offer us new tools—importantly beyond linguistic analysis—for the further exploration of the character of our manifest image.

One thing that the articles of this special issue illustrate is the wide variety of computational modeling in philosophy, both in application and methodology. One thing that all the articles collectively argue for is the philosophical promise of a wide range of computational techniques.[6]

References

Epstein, Joshua. *Agent_Zero: Toward Neurocognitive Foundations for Generative Social Science*. Princeton: Princeton Univ. Press, 2013.

Epstein, Joshua & Robert Axtell. *Growing Artificial Societies: Social Science from the Bottom Up*. Washington D.C.: Brookings Institution Press, 1996.

Grim, Patrick, Daniel J. Singer, Steven Fisher, Aaron Bramson, William Berger, Christopher Reade, Carissa Flocken & Adam Sales. "Scientific Networks on Data Landscapes: Question Difficulty, Epistemic Success, and Convergence." *Episteme*, 10, 441-464, 2014.

Hegselmann, Rainer & Ulrich Krause. "Truth and Cognitive Division of Labour: First Steps in a Computer Aided Social Epistemology." *Journal of Artificial Societies and Social Simulation*, 9 (3), 2006.

Kay, Paul, Brent Berlin, Luisa Maffi, William R. Merrifield & Richard Cook. *The World Color Survey*. Stanford: Center for the Study of Language and Information, 2009.

Ladyman, James & Don Ross. *Every Thing Must Go: Metaphysics Naturalized*. New York: Oxford Univ. Press, 2007.

Minsky, Marvin & Seymour Papert. *Perceptrons*. Cambridge, MA: MIT Press, 1969.

O'Connor, Cailin & Justin Bruner. "Dynamics and Diversity in Epistemic Communities." *Erkenntnis*, 84, 101-119, 2019.

Schmidt, Bernard. *The Modeling of Human Behavior*. Ghent, Belgium: SCS-Europe BVBA, 2000.

Urban, Christoph. "PECS: A Reference Model for the Simulation of Multi-Agent Systems." In *Tools and Techniques for Social Science Simulation*, edited by Ramzi Suleiman, Klaus Troitzsch and Nigel Gilbert, 83-114. Heidelberg: Physica-Verlag, 2000.

Weisberg, Michael & Ryan Muldoon. "Epistemic Landscapes and the Division of Cognitive Labor." *Philosophy of Science*, 76, 225-252, 2009.

Zollman, Kevin. "The Communication Structure of Epistemic Communities." *Philosophy of Science*, 74, 574-587, 2007.

6 For further work in computer modeling in philosophy, see a special issue of *Synthese* and a *Stanford Encyclopedia of Philosophy* entry on 'Computational Philosophy,' both forthcoming.