Published in M. Sprevak & M. Colombo (Eds.) *Routledge Handbook of the Computational Mind* (2018), London: Routledge, pp. 1–5
doi:10.4324/9781315643670

mark.sprevak@ed.ac.uk
m.colombo@uvt.nl

Introduction to Handbook

Mark Sprevak
University of Edinburgh

Matteo Colombo
University of Tilburg

22 October 2018

Computational approaches to explain how the mind works have bloomed in the last three decades. The idea that computing can explain thinking emerged in the early modern period, but its impact on the philosophy and sciences of the mind and brain owes much to the groundbreaking work of Alan Turing on the foundations of both mathematical computation theory and artificial intelligence (AI) (e.g. Turing, 1936; Turing, 1950). Turing's work set the stage for *the computational theory of mind* (CTM), which, classically understood, claims that thinking is a computational process defined over linguistically structured representations.

Championed by Hilary Putnam (1967), Jerry Fodor (1975), Allen Newell and Herbert Simon (1976), and Zenon Pylyshyn (1984) among others, CTM played a major role in cognitive science from the 1960s to the 1990s. In the 1980s and 1990s, connectionism (Rumelhart, McClelland and the PDP Research Group, 1986) and dynamical systems theory (Thelen and Smith, 1994) began putting pressure on the classical formulation of CTM. Since then, a growing number of cognitive scientists and philosophers have appealed to these alternative paradigms to challenge the idea that the computations relevant to cognition are defined over linguistically structured representations.

Meantime, fueled by increasingly sophisticated machine learning techniques and growing computer power, computers and computational modeling have become ever more important to cognitive science. Since the turn of this century, engineering successes in machine learning and computer science have inspired novel approaches to the mind, like deep learning, reinforcement learning, Bayesian modeling, and other probabilistic frameworks, which straddle dichotomies that previously defined the debate about CTM (e.g. representationalism vs anti-representationalism, logicism vs probability, and nativism vs empiricism). Recently, some researchers have

argued that all these theories can be unified by thinking of the mind as an embodied, culturally situated, computational engine for prediction (Clark, 2015).

The Routledge Handbook of the Computational Mind reflects these historical dynamics, engages with recent developments, and highlights future vistas. It provides readers with a comprehensive, state-of-the-art treatment of the history, foundations, challenges, applications, and prospects of computational ideas for understanding mind, brain, and behavior.

The thirty-five chapters of the Handbook are organized into four sections: 'History and future directions', 'Types of computing', 'Foundations and challenges', and 'Applications'. Although each chapter stands alone and provides readers with understanding of a specific aspect of the computational mind, there are several common threads that contribute to the narrative coherence of the volume. Some of these threads indicate a departure from past directions; others maintain aspects of the heritage of classical ideas about computation and the mind. We survey these briefly below.

An important thread that is continuous with the origin of CTM is that theorists engage with the details of actual scientific practice. In the Preface to *The Language of Thought*, Jerry Fodor explains that he had two reasons for addressing the question of how the mind works: first, 'the question of how the mind works is profoundly interesting, and the best psychology we have is *ipso facto* the best answer that is currently available. Second, the best psychology we have is still research in progress, and I am interested in the advancement of that research.' (1975, p. viii). These two considerations also animate the contributors to this Handbook. Authors rely on the best theories and evidence from the computational sciences to address questions about how the mind works. They also aim to advance research in these sciences by clarifying foundational concepts, illuminating links between apparently different ideas, and suggesting novel experiments. Authors sometimes disagree about which scientific theories and evidence count as 'the best', but their supporting discussion clarifies these disagreements and provides readers with an understanding of differences concerning computational approaches within scientific practice.

Another point of continuity with previous approaches is that many important foundational questions about the computational mind remain largely unresolved. Researchers with different backgrounds and interests continue to wrestle with 'classical' questions. Several contributors to the Handbook engage with the problem of computational implementation: What does it mean for a physical system to implement a computation? Other contributors engage with explanatory questions about the relationship between different levels of analysis, such as, for example, the relationship between David Marr's computational, algorithmic, and implementational levels (Marr and Poggio, 1976). An important question is whether one level

of analysis is somehow epistemically privileged when it comes to explain how the mind works. A further set of issues centers on the notion of representation: What kinds of representation occur in the mind, and how do they fit with computational models? Several contributors to the Handbook explore the relationship between computation, representation, thought, and action, and how we should understand representation in the context of an embodied and acting agent. Others take up questions about the role of representation in computational explanation, including the format used by representations in the computational sciences.

A final point of continuity with previous treatments concerns the challenge of scalability: How can one go from explaining a few aspects of the mind under limited circumstances to explaining the full range of mental capacities across demanding, ecologically realistic settings? One aspect of this challenge is associated with the so-called 'frame problem'. The frame problem was originally formulated as the problem of specifying in a logical language what does and does not change in a situation when an event occurs (McCarthy and Hayes, 1969). This relatively specialized problem has been taken as suggestive of a more general difficulty: accounting for the ability of computing systems to make timely decisions on the basis of what is relevant within an ongoing situation. Concerns about computational complexity and tractability compound the frame problem, creating a scalability challenge. At least since Herbert Simon's (1957) work on bounded rationality, a major question faced by computational approaches to the mind has been: How can computing systems with limited time, memory, attention, and computing power solve complex, ambiguous, and pressing problems in the real world? Taking the lead from Simon and other pioneers of AI, researchers in the computational sciences, including authors in this Handbook, develop strategies to cut through the complexity of computational problems and allow limited computing systems to solve complex, real-world problems.

Despite these points of continuity, the contributions in the Handbook also present salient points of departure from previous work on the computational mind. One such point of departure is the plurality of approaches we currently observe in the computational sciences. Instead of there being 'only one game in town' (Fodor, 1975), there are now many computational approaches to explain how the mind works, each of which illuminates a different aspect of mental phenomena.

The diversity of approaches within the computational sciences has helped to motivate several epistemological and methodological views that go under the general banner of 'pluralism'. According to these views, the plurality of computational approaches we observe in the sciences is an essential feature of scientific inquiry into the mind. The explanatory and practical aims of studying the mind are best pursued with the aid of many theories, models, concepts, methods, and sources of evidence

from different fields, including philosophy, computer science, AI, psychology, and neuroscience. As several of the contributors to this Handbook suggest, these fields are converging on a pluralistic conception of the computational foundations of the mind that promotes fruitful exchanges on questions, methods, and results.

Pluralist views about the computational mind are reflected in the progressive erosion of dichotomies that, as noted above, have traditionally defined the field. The contributions in this Handbook show that the historical roots of CTM are broad: even at its origins, CTM did not necessitate or reflect a monistic approach to the mind. Today, an increasing number of researchers realize that they do not need to pick between Turing machines, logic, neural networks, probability calculus, and differential equations as 'the' approach to the mind. Nor do they need to claim that any one of these is 'the' computational approach. Instead, they are able to choose between diverse questions, problems, and approaches that fall within the ambit of the computational sciences. They have the power to illuminate different aspects of mental or neural phenomena. None of these approaches has the monopoly on computation. This has led some researchers to reconceive of apparently competing approaches – such as connectionism or dynamical systems theory – as different aspects of the computational framework rather than as non-computational alternatives.

Work in this area reflects broader trends in the philosophy of science. Many contributors use ideas developed in the context of other sciences to illuminate the practice of the computational sciences. Examples include appealing to work on explanation and the relationship between models and mechanisms; the role of idealization in modeling and perspectivalism about models in general; and the influence of values and social structures on scientific practice. With respect to explanation, philosophers of science have articulated various accounts emphasizing different constraints on what constitutes an explanation. One recent trend salient in this Handbook is to think of scientific explanation in terms of mechanisms and models rather than in terms of laws, general principles, and encompassing theories. A turn to mechanisms and models has informed computational modeling, and raises questions about the conditions under which a computational model has explanatory value. Work on idealization and perspectivalism in the philosophy of science emphasizes that the growth of scientific knowledge is always a situated process carried out by interest-driven human beings interacting in social institutions and seeking to find their way in a complex world. This helps us to understand why there might not be a unique, universally true, computational account of the mind: different inquirers may need different models to answer different questions.

Early computational treatments of mind were closely tied to traditional metaphysical questions such as the mind–body problem (What is the relationship between mental states and physical states?) and semantic externalism (Does the semantic content of

our mental states supervene on our brains and bodies alone?). In this Handbook, these metaphysical debates often take a back seat to questions about the explanatory role of computational models in scientific practice.

One last point of departure from previous treatments arises from the increase in power of computing machinery over recent years. Technological change has contributed to dramatic advances in machine learning and brain simulation. The success of machine learning models is felt in the chapters. Machine learning techniques have inspired models of the mind based around predictive processing, statistical inference, deep learning, reinforcement learning, and related probabilistic notions. These algorithms extract statistical information from large datasets, use that information to recognize patterns, make inferences, and learn new tasks like how to play a video game, a board game, or drive a car. A question that occupies many contributors in this Handbook is whether, and to what extent, these techniques also describe the workings of the human mind. While current AI excels at narrowly defined tasks, the problem of how to re-create human general intelligence remains largely unsolved. General intelligence describes the ability to solve many diverse tasks and change goals flexibly and rationally in response to contextual cues. We do not know how humans do this. Reconstructing the process that underlies general intelligence poses a challenge to both current machine learning and computational models of the mind.

As editors, we see *The Routledge Handbook of the Computational Mind* as fulfilling three main goals. First, we see it as a 'time capsule' of current trends, marking points of departure and continuity with respect to classical computational treatments. Since the Handbook crystallizes many of the important ideas we can identify today, it will be a helpful resource for those researchers who will look back at the historical trajectory of the field in a couple of decades or so. Second, we see the Handbook as informing present-day scholars and practitioners of the accomplishments and challenges of computational approaches to the mind. Third, we see the Handbook as a pedagogical resource, appropriate for graduate and advanced undergraduate courses in disciplines ranging from the philosophy of mind and cognitive science, to computational cognitive neuroscience, AI, and computer science.

Acknowledgements

We would like to thank a number of individuals for helping to make this volume possible: Fahad Al-Dahimi and Jonathan Hoare for copy-editing and helping to prepare the volume for final submission; Adam Johnson for guiding us through the process and providing much needed support and encouragement; and, most important of all, the authors for providing thoughtful, bold, and valuable contri-

butions, for their constructive responses to our comments, and for their patience throughout the production process.

Matteo gratefully acknowledges financial support from the Deutsche Forschungsgemeinschaft (DFG) within the priority program 'New Frameworks of Rationality' ([SPP 1516]), and from the Alexander von Humboldt Foundation.

Bibliography

Clark, A. (2015). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford: Oxford University Press.

Fodor, J. A. (1975). *The Language of Thought*. Cambridge, MA: Harvard University Press.

Marr, D. and T. Poggio (1976). *From understanding computation to understanding neural circuitry*. Artificial Intelligence Laboratory. A.I. Memo. Massachusetts Institute of Technology.

McCarthy, J. and P. J. Hayes (1969). "Some philosophical problems from the standpoint of artificial intelligence". In: *Machine Intelligence 4*. Ed. by B. Meltzer and D. Michie. Edinburgh: Edinburgh University Press, pp. 463–502.

Newell, A. and H. A. Simon (1976). "Computer Science as Empirical Enquiry: Symbols and Search". In: *Communications of the ACM* 19, pp. 113–126.

Putnam, H. (1967). "Psychological predicates". In: *Art, Mind, and Religion*. Ed. by W. H. Capitan and D. D. Merrill. Pittsburgh, PA: University of Pittsburgh Press, pp. 37–48.

Pylyshyn, Z. W. (1984). *Computation and Cognition*. Cambridge, MA: MIT Press.

Rumelhart, D. E., J. McClelland and the PDP Research Group (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, MA: MIT Press.

Simon, H. A. (1957). *Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting*. New York, NY: Wiley & Sons.

Thelen, E. and L. B. Smith (1994). *A Dynamical Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press.

Turing, A. M. (1936). "On computable numbers, with an application to the *Entscheidungsproblem*". In: *Proceeding of the London Mathematical Society, series 2* 42, pp. 230–265.

Turing, A. M. (1950). “Computing machinery and intelligence”. In: *Mind* 49, pp. 433–460.