*Draft only. Do not cite without permission.*

Forthcoming in M. Colombo & M. Sprevak (Eds.) *Routledge Handbook of the Computational Mind*, London: Routledge

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

# Introduction to Handbook

Matteo Colombo
*University of Tilburg*

Mark Sprevak
*University of Edinburgh*

24 March 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 mind and brain owes much to the groundbreaking work of Alan Turing on both the foundations of mathematical computation theory and artificial intelligence (e.g., Turing, 1936–1937; 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 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 dynamicism (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 over linguistically structured representations.

Meantime, fueled by increasingly sophisticated machine learning techniques and growing computer power, computers and computational modelling have become ever more important to cognitive science. At the turn of the century, engineering successes in machine learning and computer science inspired novel approaches, like deep learning, reinforcement learning, Bayesian modelling, and other probabilistic frameworks, which straddle dichotomies in the philosophy of mind 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 these apparently opposing positions can be synthesized inside a picture of the mind as an embodied, culturally situated, active computational engine for prediction (Clark, 2015).

*The Routledge Handbook of the Computational Mind* reflects these historical dynamics, engaging with recent developments and highlighting 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 35 chapters of the Handbook are organized into four sections: *History and Future Directions*, *Types of Computing*, *Foundations and Challenges*, and *Applications*. Although each of the 35 chapters in the volume stands alone and provides readers with close understanding of a specific aspect of the computational mind, there are several common threads that contribute to the narrative coherence of the Handbook. 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 should actively engage with the details of actual scientific practice. In the Preface of *The Language of Thought*, Jerry Fodor explains that he had two reasons to attempt to say 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 animate the contributions of this Handbook. Authors rely on the best theories and evidence from the computational sciences to address questions about how minds work. 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 among computational approaches in the light of actual 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 contributions in the Handbook engage with the problem of computational implementation: What does it mean for a physical system to implement an abstract computational model? Other contributions engage with explanatory questions about the relationships between different levels of analysis, such the relationships between David Marr's computational, algorithmic and imple-

mentational levels (Marr and Poggio, 1976). An important question here is whether one level of analysis is somehow epistemically privileged when it comes to explain how minds work and why they work the way they do. A further set of issues center on the notion of representation: What kinds of representation occur in the mind and how do they fit with computational models? Several contributions in the Handbook explore the relationship between computation, representation, thought and action, and how we are to understand representation in the context of an embodied and acting agent. Others take up questions about the role of representation in computational explanations, and the adequate format of representations in the computational sciences.

Yet another point of continuity with previous treatments concerns the challenge of scalability: How can one scale up from explaining a few aspects of the mind under limited circumstances to explaining the full range of behavior across many demanding, ecologically valid 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 changes and what does not change in a situation when an event occurs (McCarthy and Hayes, 1969). But the frame problem has been taken to be suggestive of a more general problem: accounting for the ability of computational systems to make timely decisions on the basis of what is relevant in an ongoing situation. Concerns about complexity and tractability compound the frame problem, contributing to the overall challenge of scalability. At least since Herbert Simon's (1957) work on bounded rationality, a major question faced by computational approaches has been: How can computational systems with bounded resources like time, memory, attention, and computational power solve complex, ambiguous, and uncertain problems in the real world? Taking the lead from Simon and other pioneers in AI, many researchers in the computational sciences, as well as in this Handbook, have continued to develop strategies and approximations to cut through the complexity of computational systems aimed at solving large, 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 point of departure is the remarkable plurality of approaches we currently observe in the computational sciences. Instead of 'only one game in town' (Fodor, 1975), there are various ways of understanding the nature of computational systems, and a great many computational approaches to explain how minds work, each one of which illuminates different aspects of some mental phenomenon.

This plurality of approaches has helped to motivate several epistemological and methodological views going under the general banner of 'pluralism'. According to these views, the plurality of computational approaches we currently observe in

science is an essential feature of scientific inquiries 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 contributions of this Handbook suggest, these fields are converging on a pluralistic view of the computational foundations of 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 have traditionally defined the field. Partly because of a broader sense of the historical roots of CTM, and appreciation of how even at its earliest origins CTM did not necessitate or reflect a monistic approach to the mind, an increasing number of researchers have realized that they do not have to pick between Turing machines, logic, neural networks, probability calculus, and differential equations as the approach to the mind or brain. Or that any of these approaches has right to be called 'the' computational theory of the mind. Most chapters in this Handbook offers readers understanding of the kinds of questions and problems different approaches that fall within a computational framework are apt to formalize and address. Their accompanying concepts have the power to illuminate certain aspects of mental or neural phenomena. None has the monopoly on computation. Many researchers have thus reconceived apparently competing approaches—like connectionism or dynamicism—as part of the computational framework rather than as non-computational alternatives.

A related point is that work in this area now reflects broader trends in the philosophy of science. Many contributions in the Handbook appeal to work developed in other areas of the philosophy of science to illuminate the practice of the computational sciences of mind and brain. Examples include work on explanation and on the relationship between models and mechanisms, work on perspectivalism about models and the role of idealization and abstraction in modelling, and work on 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 has been to think of scientific explanation in terms of mechanisms and models, and not only in terms of laws, general principles, and encompassing theories. This turn to mechanisms and models has informed understanding of computational modelling, and raised questions about the conditions under which a computational model has explanatory value. Work on perspectivalism and on idealization in the philosophy of science emphasizes that the growth of scientific knowledge is always a situated process carried out by limited human beings interacting in structured social institutions to find their way in a complex world. This work has highlighted that there cannot be a unique, universally true computational account of the mind.

It has also helped to distance current computational approaches from issues about the metaphysics of mind.

Early computational treatments of mind were closely tied to metaphysical issues like 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 brains and bodies or also on the environment?). In this Handbook, these metaphysical debates now appear to take a backseat 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 and sophistication of computing machinery over recent years. Technological change has contributed to dramatic advances in machine learning and brain simulation. We can create electronic computers and robots that are 'smarter' and realistic computer simulations of organic brains. The power and success of machine learning models is felt in the chapters. These techniques have inspired models of the mind based around predictive processing, statistical inference, deep learning, reinforcement learning, and related probabilistic notions. Machine learning extracts statistical information by searching large datasets, combines information to recognize patterns, make inferences, and learn new tasks like playing video games, board games, or driving 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 brain. While current AI excels at relatively narrow tasks, the problem of how to achieve general artificial intelligence remains largely unsolved. A general artificial intelligence would be able to solve many diverse tasks and change its goals flexibly and rationally in response to contextual cues. We do not know how humans do this. Reconstructing the processes that underlie general intelligence poses a challenge to both AI and computational accounts of mind and brain.

As editors, we see *The Routledge Handbook of the Computational Mind* as fulfilling three aims. First, we see the Handbook as a 'time capsule' of current threads, marking points of departure and of continuity in relation to seminal treatments. Since the Handbook crystallizes most of the important trends and ideas we can identify today, it will be a helpful resource for those researchers that will look back at the historical trajectory of the field in a couple of decades or so. Second, we see the Handbook as a volume 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, the foundations of computational cognitive neuroscience, and AI/computer science.

## Acknowledgements

We would like to thank a number of individuals for helping to make this volume possible. Fahad Al-Dahimi 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. Most important of all, the authors for providing thoughtful, bold, and valuable contributions and for their constructive responses to our comments and patience through 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: Havard 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. (1936–1937). "On computable numbers, with an application to the *Entscheidungsproblem*". In: *Proceeding of the London Mathematical Society, series 2* 42, pp. 230–265.

— (1950). "Computing machinery and intelligence". In: *Mind* 49, pp. 433–460.