in a sufficiently extensive and systematic manner, which might give the impression that it is ultimately not of primary importance. If, as we have said, the historical thesis constitutes one of the two levels upon which this book is developed, it is surprising how no general chapter has been devoted to the alleged philosophical tradition of pragmatic genealogy; where to discuss, for instance, the reasons why this has remained unseen through the years. Instead, we are faced with a series of chapters in which, individually, methodological similarities of varying strength are detected, but whose overall historical nexus remains elusive to us. Tracing similarities a posteriori in the light of a given systematization is not in itself illicit, but it is not sufficient on its own to constitute a historical account.

Another theme presented in several places but partially unaddressed is conceptual engineering. Conceptual engineering is explicitly referred to by the author in several places (17, 30, 193, 208), and of course, it is integral to one of the book's main themes: reverse conceptual engineering. In light of this, we would expect a close exploration of the relation between these two philosophical enterprises throughout the book. Unfortunately, we must settle for a few rather general passages, such as the one about how pragmatic genealogy encourages responsible conceptual engineering (41). In the absence of a detailed examination of the methodological assumption of these two projects, it is not even clear whether they are compatible and integrable.

In any case, Queloz's book is still a vigorous attempt to undertake a methodological and rigorous approach to genealogy, an effort that appears to be decidedly well-directed and capable of yielding valuable results. We now have only to look forward to developments in a methodological direction and an applicative one.

Independent researcher FRANCESCO ALBENZIO

Lieto, Antonio, *Cognitive Design for Artificial Minds*.
New York: Routledge, 2021, pp. xiv + 119.

The collaboration between artificial intelligence (AI) and cognitive science is a long-lasting debated topic and it is very deeply intertwined with the theoretical foundations of these two disciplines. Even though AI and cognitive science are different fields, with different aims, methods, and applied results, they share at least two things, speaking from a very wide perspective: 1) the object of research: intelligence and cognition; 2) a general interdisciplinary and transdisciplinary approach. If for some respects the former claim is correct, and therefore intelligence and cognition can be considered as two partially overlapping notions, the latter is a sort of necessary condition for the birth of both: AI in the mid-twentieth century and cognitive science a couple of decades later. Nevertheless, it was through interdisciplinarity that these two fields could give rise to a common target, being AI from the very beginning dedicated to the simulation of "every aspect of learning and other features of intelligence"[1] and cognitive science to the study of thought

[1] From the Dartmouth proposal of 1955 and printed as McCarthy, J., Minsky, M.L., Rochester, N., & Shannon, C.E. 2006, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, August 31, 1955, *AI Magazine*, 27, 4, 12, DOI: 10.1609/aimag.v27i4.1904

and mental phenomena by putting together aspects of psychology, philosophy, linguistics, neuroscience, anthropology, and computer science, especially AI.

One may wonder why AI should not be considered as a fully cognitive discipline, rather than an engineering and technological one, given that its aim is to simulate every feature of intelligence. This is related to the ambiguity of the notion of simulation. To simulate a performance of a task that is considered to require normally human intelligence is different from simulating the underlying mechanisms and processes enabling the intelligent behavior and the cognitive performance. Only in the latter sense the notion of simulation has been adopted by cognitive science and, in return, cognitive science has become (also) a computational discipline. The distinction between a more engineering approach and a more psychological one to AI is not new and is part of the evolution of the discipline since AI was mainly symbolic driven,[2] but the more recent approaches to AI has renewed the connection between AI and the study of principles, processes, and mechanisms upon which intelligence is based. Many of the new approaches are biologically and neurologically inspired, situated, evolutionary, dynamical, and embodied, so their biological plausibility is at the core of this new approach as much as in the new approaches to cognitive science.[3] Within this new framework Lieto speaks about a rebirth of a collaboration between AI and cognitive science, a collaboration that is grounded on the old ideas of simulation and computational modeling of cognitive capabilities.

The computational cognitive science that uses cognitive modeling involves some problems, among which the main one is the problem of model. What makes a computational model a cognitive one? What are the right and relevant constraints to build a model that is not merely a system producing the same performance in specific tasks as the humans do? As the author states, "'functional' systems (in the sense explained in the book) cannot be considered artificial models of cognition if they are not additionally equipped with 'structural constraints'" (93). This is effective if one wants to explain how mind and brain work (the main aim of the cognitive/psychological AI), but also if the overall goal is to achieve systems that are capable of a suitable interaction with human beings. It is not by chance that these issues are addressed especially in some recent AI trends, such as, for example, robotics (in particular, social robotics[4]), explainable AI, and artificial life.

Starting from these premises, the focus of Lieto's proposal is on cognitive architectures, a notion that was introduced by Newell in his attempt to define a unified theory of cognition.[5] They are abstract models between the high-level cognitive capabilities and their neural/bodily implementation, so they are at an intermediate level and their characterization as an integrated mechanism is what allows to build a computational counterpart of them in an artificial system. In

[2] See for example Winston, P. 1984, *Artificial Intelligence, 2nd Edition*, Reading: Addison-Wesley.
[3] Cordeschi underlines the fact that new AI, with new models associated to the research projects of cybernetic period, is, in many cases and from this respect, the same as a new cognitive science. See Cordeschi, R. 2008, "Step Toward the Synthetic Method: Symbolic Information Processing and Self-Organizing Systems in Early Artificial Intelligence modeling", in Husbands P., O. Holland, and M. Wheeler (eds.), *The Mechanical Mind in History*, Cambridge, MA: MIT Press, 219-58.
[4] On this topic see Dumouchel, P. and L. Damiano 2017, *Living with Robots*, Cambridge, MA: Harvard University Press.
[5] Newell, A. 1990, *Unified Theories of Cognition*, Cambridge, MA: Harvard University Press.

other terms, a cognitive architecture is a model of one or more cognitive capabilities *and* its software implementation in a computational cognitive model. The more interesting cognitive architectures are, clearly, the more general ones, i.e. the ones modeling the cognitive capabilities at the highest degree of integration among intelligent features. The intermediate nature of cognitive architecture makes the problems of relevant constraints of modeling a crucial one to achieve an actual model of cognitive processes. In fact, the problem of right model is *the* problem of computational cognitive science using AI systems, as the assumption that the relevant constraints can be identified is the strongest one, from a methodological and epistemological point of view, to achieve both a "working" cognitive artificial systems and an explanation of the cognitive process.[6]

The cognitive architectures analyzed in the volume are probably the most well-known: SOAR and ACT-R,[7] starting from which many models have been developed in the last forty years. It is worth it to mention that they both started as symbolic architecture, but at least in the case of ACT-R many models developed within this general framework are hybrid, i.e. they mix symbolic and subsymbolic processes. One of the main features of many cognitive architectures is that they have a modular structure, which they derive from a well-established idea of mind that is typical of the classical, symbolic cognitive science and philosophy associated to it, especially by Fodor.[8] According to the modularity of mind view at least a part of cognition is carried out by modules, that is mental or neural structures with a specific function. Even though the modularity of a cognitive architecture is not strictly committed with modules that are characterized by the properties required by the theory, a modular structure is very well suitable to be described in a symbolic, discrete, and functional way, and in this way implemented in a software structure. For this reason, it appears to be even more convenient from a methodological point of view than from an epistemological one. A mechanistic integrated system is easily describable as a modular structure, which, in addition, fosters the possibility to build artificial systems with a hybrid way to process information, as it seems it should be the case. Or, at least, this is the view stated by Lieto.

The choice of SOAR and ACT-R is not by chance. They are two cognitive architectures in which knowledge representation is crucial and a very relevant part of the architecture. The knowledge level, to use a terminology by Newell, of both, however, is problematic for some respects, in particular for the limits that Lieto finds in "the limited size and the homogeneous typology of the encoded and processed knowledge" (65). If the former is roughly self-explanatory, the latter refers specifically to a semantic capability, i.e. the capability to categorize. Psychological research of the last fifty years has highlighted a big variety of this capacity even in the same cognitive agent, that is the human being. Heterogeneity means, therefore, flexibility, and the core of the author's proposal is a cognitive architecture

[6] And this is separate from the psychological and/or biological plausibility of the constraints. For a discussion on this see Cordeschi, R. 2002, *The Discovery of the Artificial. Behavior, Mind and Machines Before and Beyond Cybernetics*, Dordrecht: Kluwer.
[7] For a wide review of cognitive architectures see Samsonovich, A.V. 2010, "Toward a unified catalog of implemented cognitive architectures (review)", in Samsonovich, A.V., K.R. Jóhannsdóttir, A. Chella and B. Goertzel (eds.), *Biologically Inspired Cognitive Architectures 2010: Proceedings of the First Annual Meeting of the BICA Society*, Frontiers in Artificial Intelligence and Applications, 221, 195-244.
[8] Fodor, J.A. 1983, *The Modularity of Mind*, Cambridge, MA: MIT Press.

using a hybrid knowledge base that is able to process jointly different form of categorization and different kinds of categorized knowledge in form of complex structures of concepts: the DUAL PECCS.

The core of DUAL PECCS as a "cognitively inspired categorization system" (71) is a hybrid knowledge base, in which concepts are represented both according to the classical theory of concepts (a list of features of the concept itself, which are the necessary and sufficient conditions for a thing to be regarded as a member of the category expressed by the concept) and to the prototype/exemplar theories (using typical information about the concept):

> From a reasoning perspective, one of the main novelties introduced by DUAL PECCS consists of the fact that it is explicitly designed according to the flow of interaction between commonsense categorization processes (based on prototypes and exemplars and operating on conceptual spaces representations) and the standard rule-based deductive processes (operating on the ontological conceptual component) (73).

Conceptual spaces representation and ontologies are available and up-to-date tools to representing knowledge in an artificial system, so this can be considered an extension of cognitive architectures such as SOAR and ACT-R in their standard diagram but still in line with them. It is not surprising that the focus of the cognitive design approach is seen by the author in a development and an improvement of knowledge representation encompassing different theories of concepts to have a flexible behavior and performance in the artificial system from the point of view of knowledge. One of the main reasons of the birth of last decades approaches to AI has been the hard issues arisen by the "rigid" knowledge representation systems of AI in the 70s and 80s, and the general problem of how implementing common sense and background knowledge in an AI system, which cognitive architectures such as DUAL PECCS try, at least partially, to address. Lastly, even more interesting is the mention of a mutual influence of the implemented system and the experimental cognitive settings to which it is inspired, in the sense that the system performance can give some insights, in return, to the experimental research on the examined cognitive capability. According to the author, "this kind of result is exactly the type we look for in the context of a computationally grounded science of the mind" (75), and it is easily attributable also to the old and long-lasting tradition of the cognitive/psychological AI.

A last remark is needed about the notion of plausibility, as it is at the core of the modeling methodology in AI cognitive systems. The author stresses "the irrelevance, with respect to the 'plausibility' issue, of the level of abstraction adopted to model a given cognitive behaviour" (47). This position is somewhat controversial, as it is not approved by everyone. According to different approaches to cognitive modeling someone states that the right level of abstraction is the symbolic/logical/functional one, whereas others believe that the right level is the subsymbolical/neural/bodily one. The debate on such an issue has been foundational in AI and cognitive science development from an epistemological standpoint. Of course, it is related to the successful results of different approaches in modeling different cognitive capabilities along the wide range of what is meant to be cognitive. Lieto's proposal on plausibility—that is already claimed by

Cordeschi among others, as we said earlier—is deserving as an attempt to go beyond this debate and to treat every different approach with the same relevance, thus justifying hybrid artificial systems also from their structural point of view:

> the notions of both cognitive and biological plausibility, in the context of computational Cognitive Science and computational modelling, refer to the level of accuracy obtained by the realization of an artificial system, with respect to the corresponding natural mechanisms (and their interactions) they are assumed to model. In particular, cognitive and biological plausibility of an artificial system asks for the development of artificial models (i) that are consistent (from a cognitive or biological point of view) with the current state-of-the-art knowledge about the modelled phenomenon and (ii) that adequately represent (at different levels of abstractions) the actual mechanisms operating in the target natural system and determining a certain behaviour (47).

The question about what elements in the structure of the natural system give rise to the behavior to be modeled is very consequent from these statements and the most relevant one concerning the epistemic and explanatory value of the model. Starting from the list of criteria to characterize biologically plausible robotic models proposed by Webb (2001),[9] Lieto provides his own list (called Minimal Cognitive Grid) that is more synthetic also to catch a more neutral plausibility dimension in evaluating the explanatory power of a model and that is based upon three main issues: the ratio between functional and structural elements in designing a model, its potential generality, and the performance match requiring relevant features in the natural system behavior such as errors and execution time.

The Minimal Cognitive Grid together with a general discussion of evaluating methods of artificial systems (and many examples and proposals of future line of related research) is one of the two main innovative contributions of the book as a study on the philosophy of artificial intelligence and cognitive science. The other one is the renewed strength that is given to the view that consider AI, at least as a relevant research opportunity, in the wide and multifarious range of its approaches as a cognitive discipline in its fundamentals, methods, and goals.

University of Bologna FRANCESCO BIANCHINI

Conant, James and Chakraborty, Sanjit (eds.), *Engaging Putnam*.
Berlin: De Gruyter 2022, pp. viii + 372.

Hilary Putnam has surely been a thinker of the first magnitude in the last quarter of the 20th century, providing first-class contributions to many fields in philosophy. Such contributions belong to subdisciplines like philosophy of science, philosophy of language, philosophy of mind, philosophy of mathematics, logic, epistemology, and ethics. Putnam's work has been so influential in many debates in these areas because of his readiness to change his mind when faced with compelling arguments, whether from himself or from other thinkers. Along the way, he has displayed an outstanding collection of different views and ideas—and many

[9] Webb, B. 2001, "Can Robots Make Good Models of Biological Behaviour?", *Behavioral and Brain Sciences*, 24, 6, 1033-50, DOI: 10.1017/s0140525x01000127