Book Reviews

Queloz, Matthieu, *The Practical Origins of Ideas: Genealogy as Conceptual Reverse-Engineering*.
Oxford: Oxford University Press, 2021, pp. xiv + 304.

A brand-new genealogical season seems to be starting, but despite this growing popularity, there is still a lack of common ground on what genealogy is and what it stands for, and an alarmingly vast variety of conceptions is still available on the market. Moreover, being in the middle of the notorious analytic-continental divide, cultural as well as philosophical misunderstandings abound.

Before Matthieu Queloz, no recent author had ever addressed the question of genealogy as a philosophical method in such detail. The interest aroused by *The Practical Origin of Ideas* is not unexpected, since it makes available a well-conceived conception of philosophical genealogy, whose perspectives and meta-philosophical ambitions are clear and defined, though open-ended and plural. Moreover, his work taps into the manifold of genealogical conceptions in circulation, both, and most evidently, from genealogies traceable to the influence of Bernard Williams,[1] as well as those inspired by the Foucauldian tradition.[2] The author undertakes two distinct but closely related operations: the methodological exposition of what he calls "pragmatic genealogy" and the rediscovery of a hitherto ignored historical tradition of this method. We thus realize that great authors of the past such as Hume and Nietzsche, and more recently Edward Craig, Bernard Williams and Miranda Fricker, can be plausibly assigned to this tradition. The two projects are mutually enlightening: through the presentation of the method, it is possible to bring out instances of it, which in turn allows us to test its qualities (18-19).

The book's first three chapters are devoted to laying out pragmatic genealogy's theoretical framework and methodological assumptions. The first chapter moves from some questions and suspicions concerning our most abstract ideas. We inherit venerable ideas, such as those of truth, justice and knowledge, the practical purpose of which is often unclear to us; nevertheless, our actions as individuals and as human communities are guided by these same ideas. The method of pragmatic genealogy allows us to reveal what such ideas do for us, a result we achieve through the production of a peculiar historical-philosophical artefact: a rational and historical (sociological, psychological) narrative that explores how we have developed them. To be more precise, the *explananda* of this method are *conceptual practices*, that is, practices "[…] essentially shaped by sensitivity to conceptual norms or reasons—take away the idea in terms of which those norms and reasons are articulated, and the practice collapses" (3).

I will try to summarize the nature of this method for the benefit of the rest of the review. The outcome of pragmatic genealogy should not be thought of as a succession of historical facts, but as a model analogous to those in science,[3] which provides us with a perspicuous view of the conceptual practice examined. The

[1] Fricker, M. 2007, *Epistemic Injustice: Power and the Ethics of Knowing*, Oxford: Oxford University Press.
[2] Cf. Koopman, C. 2013, *Genealogy as Critique: Foucault and the Problems of Modernity*. Bloomington: Indiana University Press, for more on these two different conceptions.
[3] An idea that has precedents: Cf. Kusch, M. 2009, "Testimony and the Value of Knowledge", in Haddock, A., Millar, A. and Pritchard, D. (eds.), *Epistemic Value*, Oxford: Oxford University Press.

Argumenta 9, 1 (2023): 209-222
ISSN 2465-2334

DOI 10.14275/2465-2334/202317.boo

model in question is dynamic, representing a changing object, and emerges in two stages: a *fictionalizing* stage and a *historicizing* stage. The former requires "a maximally ahistorical setting",[4] which Craig, from whom Queloz draws inspiration, suggestively calls the "state of nature".[5] In this setting, we represent the traits of a conceptual practice (a proto-practice) whose function corresponds to the most basic conceivable function of the conceptual practice we intend to explain. By gradually increasing the complexity of the factors involved in this toy-society, it is possible to observe, step by step, the modification of the practice in response to ever-changing needs. In this way, it is possible to break down, analyse, compare and, above all, present in sequence those instrumental relations inherent in conceptual practices that in real life we can observe only synchronically. In the second stage, we move from an ideal model to a model based on the actual history of a human community: the conceptual practice under investigation is thus historicized. It is a matter of incorporating historical needs and pressures into our model and showing how that practice changes in response to them. If in the first stage, it is possible to detect the *practical needs* underlying the practice under scrutiny, the second stage shows us the *historical contingencies* that shaped the proto-practice into what it is today.

According to the author, this explanatory procedure, besides being a clear example of philosophy as model building, is analogous to a reverse engineering operation. With clear reference to the influential philosophical project known as conceptual engineering, he calls pragmatic genealogy an instance of *reverse conceptual engineering*. The second chapter is thus devoted to the presentation of seven virtues of reverse conceptual engineering, to which are added three distinctive benefits of pragmatic genealogy as a form of conceptual engineering: explanation without reduction, normative significance, and the facilitation of responsible conceptual engineering. In the next chapter, Queloz proceeds to examine the strengths of his favoured method, as compared to other forms of reverse conceptual engineering (above all the *paradigm-based explanation*). In particular, he identifies two kinds of conceptual practices that would be hardly analysable without pragmatic genealogy: *self-effacingly functional* practices and *historically inflected* practices. The former has a rather elusive functional requirement: for the practice to be properly functional, the agents must not have access to its function when they engage in it. The latter are those current practices in which the link to the basic needs they were serving when they arose has not been conserved.

The most hermeneutically inspired chapters, in which the author aims to bring to light the hidden philosophical tradition of pragmatic genealogy, are devoted to Hume and Nietzsche. Queloz comes to Hume's aid, in the fourth chapter, defending him from the accusation of producing a merely conjectural form of history. Similarly, in the fifth chapter, he refutes the charge addressed to Nietzsche, who allegedly traced a scarcely documented historical genealogy: Queloz shows how to correctly understand their purposes through the lens of pragmatic genealogy. In addition to the exegetical insights contained in these chapters, Queloz highlights some peculiarities of his method by drawing on the genealogies of the two authors reviewed; the possibility of vindicatory genealogy is exempli-

[4] Cf. Fricker 2007: 108-109.
[5] Craig, E. 1990, *Knowledge and the State of Nature: An Essay in Conceptual Synthesis*, Oxford: Oxford University Press.

fied by Hume's treatment of the virtue of justice, while Nietzsche's work is summoned in support of the possibility of employing this method to avoid what the German thinker thought was the philosophers' ancient defect of thinking ahistorically.

The chapters concerning Craig, Williams and Fricker (§6, §7, and §8, respectively) allow Queloz to substantiate his historical thesis and showcase further benefits of his proposal. As Williams once warned, "the state of nature is not the Pleistocene",[6] and the chapter on Craig further clarifies what the state of nature is and what its implications are. Queloz takes the occasion to argue in favour of the compatibility between Craig's approach, incorporated in his method, and the principles of factivity and non-analysability of knowledge of the widespread *Knowledge first* epistemological conception.

The chapter on Williams is in my view pivotal to this book. *Truth and Truthfulness* is still considered a significant work today; despite this, the aspects that were most important to the author in writing this book are rarely considered. Queloz offers, perhaps for the first time, a well-documented clarification and a strenuous defence of the author's intent and method. We find here one of the most representative examples of pragmatic genealogy, one that is not only extensive but also paradigmatic, since it is the genealogy of a self-effacing practice, the best-suited field of action of this method. From Williams we learn how an exclusively instrumental use of practices related to truthfulness along with access to the function of these would profoundly destabilize them to the point of collapse: if all individuals expected truthfulness in the practices of others while reserving for themselves the possibility of not being truthful for their own benefit, this would soon result in the collapse of these practices: we could not expect truthfulness from anyone. Concealing their own function is vital for these practices to remain stable. As a result, Williams' vindicatory genealogy leads us to an apparently controversial result: it shows how it is possible to value intrinsically a conceptual practice based on an abstract and venerable concept while at the same time continuing to value this practice instrumentally. In support of Williams, Queloz defends the compatibility between valuing a conceptual practice intrinsically and valuing it instrumentally.

Chapter eight is devoted to the genealogy contained in Miranda Fricker's influential book *Epistemic Injustice*. Here, Queloz has a chance to show that pragmatic genealogy can be proposed as an ameliorative project of our practices and not merely as a descriptive survey. Taking an ameliorative outlook is, in a few words, about trying to change our current practice to what we believe it should be. This perspective has attracted great interest within conceptual engineering, devoted among other things precisely to exploring the possibility of modifying our representational devices, such as concepts. However, pragmatic genealogists can also pursue an ameliorative approach, as exemplified in Fricker's work. Indeed, the retro-engineering of conceptual practice allows us to identify the developments that resulted in a practice that we believe is not the best possible. As Fricker's genealogy clearly shows, this opens the way for an ameliorative process. She brings in a political dimension precisely at the exit from the State of Nature: it consists of the creation of social groups and the consequent phenomena of social categorization. Then, she shows how the testimonial injustice that still abounds

[6] Williams, B. 2002, *Truth and Truthfulness: An Essay in Genealogy*, Princeton: Princeton University Press, 27.

today is the result of pressures opposed to a virtuous division of epistemic labour, inviting us to cultivate the virtue of testimonial justice.

Having finished his close examination of the work of past genealogists, Queloz turns back to the exposition of his method. In the ninth chapter, the normative ambitions of pragmatic genealogy are defended: Queloz presents and responds to four increasingly specific objections that sum up the most common criticisms addressed to normatively ambitious genealogical explanations. First, the charge of genetic fallacy is dismissed. Queloz presents two different forms of the genetic fallacy. The former cannot threaten his method; the latter is committed only by inferring something about the justification of a conceptual practice from irrelevant information about its formation process, which is entirely avoidable in a pragmatic genealogy. He then proceeds to describe two kinds of conceptual practices in which the formation process carries normative weight. The second charge, which focuses on lack of continuity, is avoided altogether, probably because it applies only to far more traditional genealogies. Queloz's genealogy does not assume that there must be continuity between the conditions under which a conceptual practice arose and those that survive today, but on the contrary, is designed to reveal it. He also rejects the claim that pragmatic genealogy can only deal with practices that emerged in connection with anthropological universals, which would severely narrow its scope: he shows in some detail how pragmatic genealogy can also deal with extremely local and contingent practices. The last objection, which points the finger at the arbitrariness in the attribution of needs on which Queloz's method is based, is partly overturned and partly accepted. This method makes it possible to account for the attributions of needs since these must be systematically traced back to basic and increasingly less contestable needs. However, a central role is indeed played by the genealogist's point of view, but this is a welcomed aspect of this method, which does not assume that there is an extra-subjective point of view for such matters.

The last chapter is spent on some meta-philosophical considerations. Here two possible approaches that pragmatic genealogy can encourage are introduced: a Socratic inquiry grounded in pragmatic inquiry and the practice of philosophy as a humanistic discipline. The latter approach, evidently Williamsian, reveals how good genealogical practice requires the maximum integration of insights gained from the other humanistic disciplines and in the social sciences, abandoning the idea of a pure philosophical inquiry independent of other forms of knowledge.

The Practical Origins of Ideas reflects Queloz's erudition and his well-rounded knowledge of the field of inquiry, as well as his remarkable clarity and care in exposition. As already mentioned, Queloz brings new life to the thought of the late Bernard Williams, an author of whom he is an eager connoisseur given how confidently he masters his vast and various philosophical production

The weaknesses of this book are mainly architectural. The author has made a very hard but understandable choice in the economy of the text, by not producing *ad hoc* instances, choosing instead to exploit past genealogies that he has reread (or rediscovered) as pragmatic genealogies. This constitutes a burdensome constraint because it does not allow the choice of more didactic examples and ties their exposition to previous exegetical passages.

In addition, the historical thesis regarding the tradition of pragmatic genealogy, although presented with abundant interpretative suggestions, is not treated

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

versions thereof. This variety can be difficult to track for common readers and sometimes even for scholars.

The present collection, *Engaging Putnam*, edited by James Conant and Sanjit Chakraborty, is a major attempt to keep alive various relevant threads in Putnam's legacy and to honour an absolutely leading figure in contemporary philosophy. They are not shy to acknowledge the difficulties in an enterprise like this, with so many arguments and views changed within a few decades—a philosopher that has been considered a "moving target" (16-20). However, this ensemble of views is in an important way tied together by a central thread in Putnam's efforts, the issue of realism understood as our struggle to grasp the crucial role that a mind-independent reality plays in our intellectual endeavours. The book has two introductions—one devoted to celebrating Putnam's greatness and uniqueness in the contemporary scene, and another to present the contents of the collection—and twelve chapters by philosophers whose work has been heavily influenced by Putnam's. The list includes renowned figures such as Yemima Ben-Menahem, Tim Button, Roy Cook, Mario De Caro, Maximilian de Gaynesford, Gary Ebbs, Sanford C. Goldberg, Tim Maudlin, Martha C. Nussbaum, Duncan Pritchard, Joshua R. Thorpe, and Crispin Wright. Almost all the chapters address from a specialist's perspective some particular view or argument by Putnam. Hence, this is not just an honorary book: the authors celebrate Putnam's legacy by trying to engage with his views in a critical way. In this review there is not enough space to duly cover all the papers included. I extend my apologies for concentrating on the contributions that better fit my personal appreciation of Putnam's work and/or spare my limitations of competence.

I start with Thorpe and Wright's essay on a topic of great relevance for Putnam's role in recent philosophical discussions: the controversial proof for the view that we are not brains in a vat (BIV).[1] Thorpe and Wright engage in a commendable goal: to figure out the main lessons from this argument and the ensuing 35 years of worldwide discussion. This is a very important goal, since the significance of the proof has "remained stubbornly controversial" (63). Because of this fact, the authors raise important questions: "Does the proof work? If so, what exactly does it show? And of what, if any, significance, metaphysical or epistemological, is the result?" (63). They lay out the argument as follows: "(1) If you were in the VAT scenario, you could not refer to BIVs. However: (2) You can refer to BIVs (since, of course, your word "BIV" refers to BIVs). Therefore: (3) You are not in the VAT scenario" (65). They discuss it first at the level of reference (65-66) and declare that the proof here works by means of the semantic externalism defended in terms of the Twin-Earth thought experiment. However, they argue that the status of premise (2) remains controversial: is it not question-begging for the overall argument? "[D]on't you have to know that you are not in the VAT scenario before you can know that you can refer to BIVs—and thus know exactly the thing that the VAT argument is supposed to prove?" (66). Then they proceed to read the argument at the level of concepts (66-67). Here the argument goes as follows: "(1*) If you were in the VAT scenario you could not have any concept of a BIV. But: (2*) You do have a concept of a BIV. Therefore: (3*) You are not in the VAT scenario" (67). This version is also supported by semantic externalism, now concerning conceptual content, and works as much as the former does—with the

[1] Putnam, H. 1981, *Reason, Truth, History*, Cambridge: Cambridge University Press.

same doubts concerning the (question-begging) status of premise (2) of the referential version. They then directly address this controversy (68-88). First of all, they show that the argument shares problems with McKinsey's argument,[2] enabling a thinker to gain contingent socio-linguistic knowledge from the armchair—this paradoxical conclusion is taken as evidence that even though these arguments may be formally valid, they fail to transmit justification to their conclusions.[3] Second, given these problems with the warrant of transmission it follows that, even though we do not conclude that the proof has failed, we face another issue concerning what it is that the argument is supposed to prove—it seems that, except for a sense in which the VAT argument succeeds, it depends on the fact that a VAT could not make the argument because this presupposes an unavailable mastery of the English language and because BIVs fail to refer to BIVs in their VAT language. Third, according to the authors, the many new sceptical versions of the thought experiment fail in the end to make the VAT argument unsuccessful, even though answering the sceptic was not Putnam's primary goal.[4] Finally, the main goal of the VAT scenario was to illustrate how metaphysical realism was not incompatible with errors in the ideal theory and indeed with the conception of an Ideal Error—the authors here show how a problem of the VAT scenario is its inability to see alternative options like Davidson's[5] to this unwarranted conclusion, as these permit to highlight significant differences between "metaphysical realism, understood as throughout this discussion, and Ideal Error" (87-8).

Another chapter which delves into Putnam's ground-breaking work is the one written by Goldberg, addressing the compatibility of semantic externalism with our understanding of the first-person perspective (107-129). Goldberg characterises semantic externalism, both for linguistic meaning and for mental content, as the acceptance of the following principles:

> LE [Linguistic Externalism] For all languages L and speakers S of L, there are some expressions e of L for which the standing meaning of e as used by S does not supervene on S's bodily states (107).
>
> AE [Attitude Externalism] For all subjects of the propositional attitudes S, there are some attitudes A of S's which are such that the fact that S instantiates A does not supervene on the facts constituting S's bodily states (108).

Goldberg then addresses the second topic, which is the first-person perspective, i.e. our epistemic perspective on the world, by distinguishing two conceptions: a *spatial* view and an *informational* view. According to the spatial conception, "to have a point of view—an epistemic perspective on the world—is to occupy a particular spatial location at every moment at which one exists" (109). According to the informational conception, "to have a point of view […] is to be such that one's cognitive life can be represented as an ever-evolving stock of information resident

[2] McKinsey, M. 1991, "Anti-Individualism and Privileged Access", *Analysis*, 51, 1, 9-16.
[3] "If one specific kind of epistemic basis for the premises of a valid argument is such that it would be *undermined* by doubt about its conclusion, then one cannot rationally be open-minded about the status of that conclusion yet simultaneously avail oneself of that basis to accept the premises" (73).
[4] See also Pritchard's chapter on this issue (263-64).
[5] Davidson, D. 1986, "A Coherence Theory of Truth and Knowledge", in Lepore, E. (ed.), *Truth and Interpretations*, Oxford: Blackwell, 307-19.

"in" one's information-processing system" (109). These options are compatible with each other: we can admit that the information which we access and process depends on the locations we find ourselves in at certain given moments (109-10). Goldberg adds further assumptions to this scenario, like the following: "the informational system just is a physical system that traces a spatial position through time" (110); and "novel empirical information" reduces to what has "causal impact on the physical system" (110). By putting these assumptions together, we can claim that one's point of view can be understood in terms of the location occupied, the initial state of the system, and all "*the physical goings-on within that system*" concerning its "*impacts*" with the world (110). While this conception is *prima facie* reasonable, it has a problem with AE: this picture of a first-person perspective only concerns causal relevance, while AE acknowledges the relevance of objects/other subjects in one's environment to characterise metaphysically one's mental life. According to Goldberg, this observation is the starting point of one greater difficulty, because AE challenges the usual conception of the autonomous epistemic subject (110-11). AE puts constraints on one's mental life: the concepts that form the contents of our attitudes cannot be specified independently of the subject's environment (111). Goldberg here affirms that many of us are tempted to say that there are dimensions of our mental lives that somehow escape AE's constraints (111). For example, whereas concepts are determined according to externalist credentials, "conceptions" may be more subjective, i.e. they can contain errors and idiosyncrasies, generating contexts which evade strict externalism. Goldberg reads Putnam's externalism as understanding this subjectivism as mostly wrong: conceiving of things cannot be specified independently of the world and the community a subject belongs to. But this puts the very idea of the autonomous epistemic subject in jeopardy (111). A new feature that may be useful and "tempting" in thinking about points of view is the idea that one's epistemic perspective on the world is metaphysically (though not causally) independent of the world itself (MIPOV). MIPOV seems plausible from the angle of introspection, that is, regarding "the nature of one's self-knowledge of […] the materials that constitute […] one's attitudes" (112), and gains traction also from considerations revolving around the idea of a conception. Without enough clues about how "conceiving" works, we would fail to capture how one takes the world to be (112). A problem is that such conceiving relies on a capacity to discern the content-relevant features of one's mental life from the armchair (112). But if this is the case, AE fails to plausibly account for the subject's point of view. Goldberg identifies the considerations concerning introspection as the main rationale for this conclusion after discussing the argument for it (114-15). As said, MIPOV exploits the concept of a "conception": an epistemic "perspective" on the world is captured by how one "takes things to be" (115). Goldberg argues that the level of conceptions is a level of description of the subject's mind that is metaphysically independent of how things are (116). This depends on an argument that exploits our ability to "hold the appearances fixed" while "varying the underlying reality" (116). At least in these circumstances, how a subject conceives of reality is metaphysically independent of that reality: "S's point of view can be invariant over how things are in the world; so any construal of her point of view that fails to appreciate this is deficient" (119). AE fails to appreciate how the construction of a point of view entails the ability to keep appearances fixed in the face of variations in how things are. Goldberg presents this argument as the crucial case against externalism in current debates. However, according to Goldberg, Putnam

already offered reasons to refute MIPOV, and so there is a challenge for the anti-externalist. Discussions revolving around introspection vs. AE have shown how externalism is compatible with discerning the commitments involved in representing things in a certain way from the armchair (121-22). Goldberg offers an analogous move for MIPOV's defence based on the contrast concepts/conceptions:

> Even in the restricted set of cases in which a subject accepts or presupposes that how things seem to her is indicative of how they are, how things seem to her—how they appear to her to be—can be held fixed, even as we radically vary the nature of the world around her (123).

Goldberg argues that we aim to represent objective kinds "as the objective kinds that they are" and this is a claim that can be endorsed even by Putnam's critics. This becomes the basis of an argument showing that "for any concept whose individuation is 'externalist', the subject's conception of that concept must be construed externalistically as well" (124).

Nussbaum's chapter addresses Putnam's relationship with Aristotle's legacy, dealing with some anti-reductionist lessons that became important in Putnam's later years.[6] The first lesson concerns the philosophy of mind and the way in which Putnam abandoned functionalism about mental states—i.e. the idea that mental states are identified in terms of the functional role they play in someone's cognitive economy. This ground-breaking idea permitted us to understand "abstract" computations as connected with a "material" substrate in a way inspired by the relationship between software and hardware. Nussbaum reconstructs how Aristotle's influence had a role in this important change of mind: it was in an Aristotelian spirit that Putnam at a certain point came to realise that the intentional level of mental states could not be reduced to the computational level required by machine functionalism. According to Putnam, the complexity of certain intentional states cannot be wholly explained in terms of computations, leaving aside the relations of such states with (sets of) objects in the real world (237). Another lesson with a distinguished Aristotelian flavour, according to Nussbaum, concerns the directional intentionality of thought and language. Putnam stated the superiority of Aristotle over Wittgenstein as a guide to this problem (238). Putnam started to wonder how Aristotle's idea of an isomorphic resemblance between the form of an object and the relative idea in one's mind anticipates a central insight of Wittgenstein's Tractarian picture theory of meaning. But these resemblances do not go too far: causal theories of reference put such insights quickly in jeopardy. Here, the idea of "not logically equivalent different descriptions of the same event" enters the scene. Therefore, the causal connection exploited by the causal theory of reference is alone insufficient and lacks an account of form (e.g. given Putnam's model-theoretic argument).[7] At this point, Aristotle and Wittgenstein take again the centre stage as both defend a particular notion of "form" (238). Putnam finds Aristotle's notion by far more useful than Wittgenstein's in dealing with the dispute with the causal theorist of reference. This choice is based on the worldly roots of Aristotelian metaphysics (while Wittgenstein's notion of form is abstract): according to Putnam, "[t]he idea that logic could do

[6] Also Ben-Menahem's chapter addresses the issue of reductionism (289-308).
[7] Putnam, H. 1981, *Reason, Truth, History*, Cambridge: Cambridge University Press.

all the work of metaphysics was a *magnificent* fantasy, but fantasy it surely was".[8] Another superior aspect of Aristotle's notion is its everyday (i.e. non-technical) character. An account like this is, however, exposed to objections. The first goes like this: our everyday representations sometimes go badly wrong, so these should not be inserted as criteria "in the mind in order for reference to be secured" (239). Putnam replied that the requirement of having the essential metaphysical properties always embedded in our everyday representations is too strict for getting reference right: "[p]eople successfully referred to water without knowing its atomic structure" (239). Another problem was Aristotle's idea that species have timeless essences, which is at odds with current biology. To this observation, Putnam replied by pointing out that even if timeless essences are hard to defend, certain features of them, such as "the ordinary synchronic notion of species" are still useful and indeed "indispensable" (239). Scholars now certify that Aristotle was not as rigid in defending "timeless essences" as medieval interpretations stated. Nussbaum concludes with another lesson concerning ethics that leaves also room for hints of Putnam's personality, providing a remarkable portrait (242-48).

The above chapters are just some highlights which can give the reader an approximate idea of what a great book this is. All the chapters would have deserved a full presentation as they tackle pivotal problems such as the a priori in philosophy of science, realism in philosophy of mathematics, scepticism in epistemology, free will, and naturalism, the ethical value of literature, and many more. This collection of papers on Putnam's work honours him by paying tribute to the central issues of his philosophy, without dodging going deep into the most controversial arguments, and often ending up with overt criticisms or noteworthy disagreements.

University of Cagliari PIETRO SALIS

[8] Putnam, H. 1995, *Words and Life*, Cambridge, MA: Harvard University Press, 71.