

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's⁵ 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.