Mario De Caro
Università Roma Tre & Tufts University

PUTNAM ON THE MIND‑BODY PROBLEM

Abstract. *This article discusses Hilary Putnam's views on the mind-body problem, by locating them in the general context of a satisfying pluralistic naturalism that he tried to articulate throughout his entire philosophical career. The first attempt in this direction was* computational functionalism, *his version of psychological functionalism centered on the analogy between mind/body and software/hardware, which (differently from David Lewis and others) he came to think of as an empirical hypothesis. That was a very successful proposal; however, later Putnam abandoned it and embraced what he called "liberal functionalism". The reason for this change of mind was twofold: on the one hand, Putnam reached the conclusion that computational functionalism was incompatible with his views on semantic externalism; on the other hand, he began to think that mental states, besides being compositionally plastic (i.e., two entities can be in the same psychological state without being in the same physical state), are also computationally plastic (i.e., two entities can be in the same psychological state without being in the same functional state). In conclusion, I will argue that "liberal functionalism" opened an interesting perspective for a successful non-reductive version of naturalism.*

Key words: *mind-body problem; Hillary Putnam; functionalism; plasticity*

1. From computational functionalism to liberal functionalism

Hilary Putnam was an omnivorous philosopher: paraphrasing Terence's famous words, one could say that nothing philosophical was alien to him. From philosophy of mind to epistemology, logic to philosophy of language, history of philosophy to metaphysics, philosophy of religion to ethics, he offered contributions that were always brilliant and often seminal. However, as is well known, Putnam's huge philosophical work was marked by frequent changes of mind; therefore, one may think that his philosophical development was devoid of continuity. This interpretation, however, would be mistaken, since several well-defined major threads and unchanged goals unified his entire philosophical production – and this is particularly true of his reflection on the mind-body problem, which is the subject of this paper. Therefore, in order to understand the trajectory of Putnam's thought regarding the mind-body problem, one has to consider three other main themes which remained constant in most of his work:

1. The search (which began at the beginning of the Sixties) for an encompassing form of *liberal naturalism,* able to do justice to the scientific worldview, on the one hand, and the manifest worldview with his ineliminable normative components, on the other hand;

2. The pursuit of a *scientific realism* immune from the insurmountable difficulties of the so-called "*metaphysical realism*" (the idea that there is exactly one true and complete description of the way the world is);

3. An uninterrupted allegiance to *semantic externalist* (which started at the beginning of the Seventies);

1.1. Computational functionalism

Computational functionalism was Putnam's first accomplished attempt to answer the mind/body problem.[1] The target that Putnam had in mind when he developed his functionalist view (or "computational functionalism", as he came to call it later) was the so-called "mind-brain type identity view" (proposed by Smart, Place, and Feigl), according to which mental types of events are identical to physical (that is, cerebral) types of events. Putnam offered a famous argument against that view. Take a human's mental event, such as feeling pain. Now, imagine an octopus having pain. The octopus is in the same mental state of the human, but certainly its physical state is very different from the human's physical state (in fact, the octopus has a brain that is anatomically very different from ours).[2] Moreover, besides octopuses, also a huge number of animals that feel pain have brains and nervous systems very different from ours and from each other (and nowadays we even have some evidence that even vegetables may feel pain).[3] But there is more: we can imagine that also robots or aliens physically very different from us could be able to feel pain – and it may well be that these entities are possible. So, if it is true that even a robot, made of silicon instead of carbon, or an alien, made of who knows what, could feel pain, it means that there is a potentially unlimited number of physical bases of pain. Consequently, there cannot be a type-identity between the mental and the physical, because the physical type in question is unavoidably open-ended.

An alternative view, which Putnam started to elaborate in 1960, was based on the view that the mind should be interpreted as a Turing machine or as a piece of software running on some hardware (the brain). The mind, in this sense, is a program hardwired in a physical basis (for humans, the brain) and it defines all the mental states, which are seen as intrinsically computational; and, more specifically, each mental state is functionally defined by its causal inputs and outputs, independently of its physical base. In this light, if a mental

---

1 See the last seven essays in H. Putnam, *Mind, Language, and Reality. Philosophical Papers,* vol. II, Cambridge University Press, Cambridge 1975.

2 "In an octopus, it is not clear where the brain itself begins and ends. The octopus is suffused with nervousness" (P. Godfrey-Smith, "The Mind of an Octopus", *Scientific American Mind*, 28, 1, 2017, pp. 62–69; retrieved at https://www.scientificamerican.com/article/the-mind-of-an-octopus/).

3 Frank Kühnemann of the Institute for Applied Physics in Bonn has found out that "when a leaf or a stem is cut off, the plant 'cries out' in pain by releasing the gas ethylene over its entire surface": see https://www.dw.com/en/when-plants-say-ouch/a-510552–1.

state of a human and one of a robot play the same causal role, the human and the robot are in the same computational (i.e., mental) state.

1.2. Liberal functionalism

At the beginning, Putnam interpreted computational functionalism as valid a priori (as long as there minds exist, of course), but later he came to see it as a (strongly corroborated) empirical hypothesis. However, finally, he saw insurmountable difficulties with computational functionalism as such and abandoned it altogether. This is how Putnam summarized how his attitude toward this view changed:

> In "Minds and Machines" I assumed that the brains of both robots (pretend there are intelligent ones!) and humans can be described as computers. I suggested, but didn't commit myself to, the idea that the mental states of those robots and those humans could be identified with what I called the "logical states" of their brains, meaning by that the states described by their programs. I called them "logical states" to emphasize that the *physical description* of those states was irrelevant; if a robot "brain" and a human brain have the same program, then the human and the robot have the same mental states, if that idea is right. Subsequently, I committed myself to this identification of mental states and computational states as an "empirical hypothesis". Very soon, I found difficulties with the identification of mental states and computational states—difficulties that led me to various reformulations.[4] Eventually, I found I couldn't reconcile this identification with my advocacy of externalist and anti-individualist semantics in "The Meaning of 'Meaning'", and I finally I discarded it as "science fiction".[5]

As Putnam hints in this passage, the main reason for which he abandoned computational functionalism was that he realized its incompatibility with another view he had been defending since the beginning of the Seventies: semantic externalism. Putnam presented the latter view by appealing to the famous "Twin Earth" thought-experiment, which started the "externalist revolution" (to paraphrase John Heil 1992, 24). Here is how Putnam summarizes that thought experiment:

> Imagine a planet like earth—call it "Twin Earth"—on which the liquid that fills the lakes and rivers, that people drink, etc. is not H_2O but a different compound XYZ, with similar superficial characteristics. The Twin Earthers are supposed to be our "Doppelgangers"; some of

---

4   Putnam describes these reformulations in his autobiographical essay in S. Guttenplan (ed.), *A Companion to the Philosophy of Mind*, Blackwell, Oxford 1993, pp. 507–513.

5   H. Putnam, "Corresponding to Reality", in H. Putnam (eds.), *Philosophy in an Age of Science*, ed. by M. De Caro and D. Macarthur, Harvard University Press, Cambridge (MA) 2012, pp. 72–90; quotation at pp. 39–40.

them even speak English. Also, I imagine the year to be 1750, hence prior to the chemical composition of either water or twater (Twin Earth 'water') being known. The English speaking Twin Earthers naturally call twater "water" (and the French-speaking ones call it "eau", and the German-speaking ones call it "Wasser"). The linguistic intuition of the great majority of people who have considered this thought experiment is that upon learning that Twin Earth "water" doesn't consist of H_2O at all, we Earthers would say "it isn't really water". The word "water" has a different meaning on Earth and on Twin Earth. Twin Earthian Oscar's word "water" and Earthian Oscar's word "water" are homonyms, but not synonyms. They do not have the same meaning—not even if Earthian Oscar and Twin Earthian Oscar happen to be microphysical duplicates![6]

The famous slogan that Putnam derived from this view was "Meanings ain't in the head!"; however, the slogan was too prudent, as Putnam repeatedly said later, since it should have rather be "Thoughts ain't in the head!". The externalist view that was supported by the Twin Earth experiment concerned our causal interactions with the physical world: so it could be called "physical externalism". However, Putnam also defended another version of externalism, "social externalism", on the basis of the idea of the "linguistic division of labor", according to which the necessary and sufficient conditions for individuating the referents of a general name (such as "elm" or "aluminum") are "all present in the linguistic community considered as a collective body; but that collective body divides the 'labor' of knowing and employing these various parts of the 'meaning' of [that general name]"[7].

The reason why semantic externalism is incompatible with computational functionalism is not difficult to understand, even if Putnam resisted several years before drawing such conclusion. As said, according to computational functionalism, thoughts are internal to the mind on the individuals; according to semantic externalism, instead, thoughts reach out the individual minds, since they involve our transactions with the external world (natural and social). Writes Putnam:

> I had to give up "functionalism," for example, that is, the doctrine that our mental states are just our *computational* states (as implicitly defined by a "program" that our brains are hard-wired to "run"), because that view is incompatible with the semantic externalism that years of thinking about the topic of reference had eventually led me to develop. If, as I said in "The Meaning of 'Meaning'," our intentional mental states aren't in our heads, but are rather to be thought of *as world-involving abilities*, abilities identified by the sorts of transactions

6   H. Putnam, "The Development of Externalist Semantics", in H. Putnam, *Naturalism, Realism, and Normativity*, ed. by M. De Caro, Harvard University Press, Cambridge (MA) 2016, pp. 199–212.

7   H. Putnam, "The Meaning of Meaning", in Id., *Mind, Language and Reality,* cit., pp. 215–271; quotation at pp. 227–228.

with our environment that they facilitate, then they aren't identified simply by the "software" of the brain.

Putnam justified his abandonment of computational functionalism also in another way. In fact, he came to realize that mental states "are not only compositionally plastic, that is, capable in principle of being realized in different sorts of hardware, but *computationally plastic*, that is, capable of being realized in different sorts of software."[8] Mental attitudes such as "believing (or fearing or hoping) something" can be mapped onto, or realized by, a software in many different ways.[9]

However, Putnam did not entirely abandon functionalism: he rather reinterpreted it in an externalistic spirit, and called the resulting new view "liberal functionalism". The idea behind the new view was that the mind is a system of object-involving abilities, which still are functions – but not functions that are merely hardwired in the brain of a speaker (as it was for computational functionalism), but functions that are intrinsically "transactional", since from the start they involve the natural and social environment in which that speaker is located.

> I still believe that our so-called "mental states" are best thought of as *capacities to function*, but not in the strongly reductionist sense that went with the model of those states as "the brain's software." They are, so to speak, "long-armed" functional states—their "arms" reach out to the environment, and their identity depends, as Ruth Millikan has stressed, on their evolutionary history.[10]

2. Putnam on naturalism and realism

Hilary Putnam strong refusal of the reductionist "mind-brain type identity view" and his attempts at formulating an adequate non-reductive functionalist view of the mind show that since the early stages of his philosophical career he had been looking for a satisfying non-reductive form of realistic naturalism.[11] What he aimed at was a view that, while deeply

---

8   H. Putnam, "From Quantum Mechanics to Ethics and Back Again", in *Philosophy in an Age of Science*, cit., pp. 29–30.

9   See H. Putnam's *Representation and Reality*, Cambridge (MA), MIT Press, 1988, and Id., *Self-Portrait*, in S. Guttenplan (ed.), *A Companion to the Philosophy of Mind*, Blackwell, Oxford 1993, pp. 507– 513.

10   H. Putnam, "From Quantum Mechanics to Ethics and Back Again", in *Philosophy in Age of Science*, cit., pp. 51–71; quotation p. 51

11   See H. Putnam, *Naturalism, Realism, and Normativity*, cit.; see also M. De Caro, "Putnam's Philosophy and Metaphilosophy", introduction to that volume, pp. 1–18. Forms of non-realistic naturalism have also been developed (think of Bas van Fraassen's or John Dupre's views, which deny that we can know that the atoms and the other unobservable entities exist).

respectful of the results obtained by the natural sciences (by refusing all supernaturalist interferences), did not assume that, at least in principle, such sciences could explain everything that could be explained.

Putnam's first accomplished attempt to formulate a satisfying non-reductionist form of naturalism was his "internal realism," which took form in 1976.[12] The core of that conception was an epistemic view of truth (truth in idealized epistemic conditions) inspired by C.S. Peirce and Michael Dummett. This view identified truth with justification in idealized epistemic conditions, which Putnam interpreted in a soft way:

> If I say "There is a chair in my study", an ideal epistemic situation would be to be in my study with the lights on or with daylight streaming through the window, with nothing wrong with my eye-sight, with an unfocused mind, without having taken drugs or been subjected to hypnosis, and so forth, and to look and see if there is a chair there.[13]

The main reason for which Putnam developed that view was his refusal of "metaphysical realism" (a view that had attracted him in the previous couple of decades), according to which reality can be completely described in exactly one way and that way precisely and ultimately fixes ontology. It should be noted, however, that even during his internal realism period Putnam never entirely abandoned scientific realism. First of all, he always thought that the theoretical terms of our best scientific practice do refer to real entities, even if such entities are in principle unobservable (i.e., electrons and black holes), and this means that he always rejected all forms of scientific antirealism, such as instrumentalism, conventionalism, operationalism, and relativism. Second, when he moved away from his pre-1976 physicalistically-oriented realism toward internal realism, Putnam was motivated by his desire to shape a satisfying philosophical realism—that is, a realism able to accept simultaneously (1) the approximate and revisable correctness of the scientific worldview and (2) the approximate and revisable correctness of the ordinary worldview, which Putnam then believed (and never stopped believing) was threatened by the reductionist conceptions of reality. In this light, at least since 1976, Putnam rejected all positions that are unable (or worse, do not even try) to do justice, at the same time, to science and to the ordinary view of the world. Arguably, Putnam's painstaking and uninterrupted efforts to shape a version of naturalistic realism able to acknowledge the partial and revisable verisimilitude of both the ordinary and the scientific images of the world is one of his most relevant bequests to the next generations of philosophers.

Then, starting in 1990, Putnam abandoned internal realism; and that happened for two main reasons. First, he realized that the epistemic conception of truth was deeply inadequate. A convincing supporting example

---

12   The best presentation of internal realism is in H. Putnam, *Reason, Truth, and History*, Cambridge University Press, Cambridge 1981.

13   H. Putnam, *Realism with a Human Face*, Harvard University Press, Cambridge (MA) 1990, p. vii.

of why truth is not epistemically constrained is a conjecture such as "There is no life outside the earth" – which may well be true but, if it is so, it would be unverifiable even in ideal epistemic conditions. In abandoning the epistemic view of truth, however, Putnam realized that he did not need to go antirealist in order to refuse the dogmatic view that he had called "metaphysical realism". His new aim was, in fact, to develop "a modest non-metaphysical realism squarely in touch with the results of science".[14]

The second reason that convinced Putnam to abandon internal realism was the so-called "no-miracles argument", which he had developed in 1973, but whose relevance regarding the issue of realism he fully appreciated much later. This argument is based on the idea that the only way of accounting for the great explanatory and predictive success of the best theories of modern science is to acknowledge that these theories are true (or approximately true) in regard to the natural world and that they refer to real entities, even when those entities are unobservable. From the point of view of antirealism, on the contrary, the fact that science works so well in offering comprehensive explanations and extremely precise predictions of observable phenomena is an inexplicable mystery, if not a sheer miracle.[15] Consequently, according to Putnam, we should consider our best scientific theories as approximately true and the entities such theories refer to as real. Unsurprisingly, antirealists have tried to attack the miracle argument in various ways, but in my view, Putnam and others have responded to those arguments in satisfying ways.[16]

Therefore, according to Putnam – apart from his internal realist period – scientific theories can be true (or approximately true) even in case we cannot ever verify them. However, even if he was a scientific realist, Putnam refused the strict naturalist view – and endorsed the liberal naturalist one – for two main reasons. First, because of the phenomenon he called "conceptual relativity," which means that some theories can be *cognitively equivalent*, even if prima facie they appear to be incompatible. (This phenomenon could less equivocally be called "descriptive equivalence," since the original term may suggest a connection with relativism and antirealism that is entirely inappropriate.) As Putnam convincingly argued, in some scientific fields such as mathematical physics, this phenomenon is ubiquitous:

> To take an example from a paper with the title "Bosonization as Duality" that appeared in Nuclear Physics B some years ago, there are quantum mechanical schemes some of whose representations depict the particles in a system as bosons while others depict them

---

14 H. Putnam, *Ethics without Ontology*, Harvard University Press, Cambridge (MA) 2004 p. 286, n.1.

15 During his internal realism years Putnam defended both the no-miracles argument and the epistemic view of truth: he solved this obvious tension by abandoning the latter view in 1990 (analogously, as we have seen, between 1960 and the mid-1980's he tried to reconcile computational functionalism with semantic externalism, until he realized that they were incompatible).

16 H. Putnam, "Corresponding to Reality", cit.

as fermions. As their use of the term "representations" indicates, real live physicists—not philosophers with any particular philosophical axe to grind—do not regard this as a case of ignorance. In their view, the "bosons" and "fermions" are simple artifacts of the representation used. But the system is mind-independently real, for all that, and each of its states is a mind-independently real condition, that can be represented in each of these different ways. And that is exactly the conclusion I advocate...[These] descriptions are both answerable to the very same aspect of reality...they are "equivalent descriptions".[17]

The second reason for which Putnam refused strict naturalism is more interesting for our purposes. This is the fact that, in his view, the ontology of the world cannot be limited to the entities and properties described by physics:

I do indeed deny that the world can be completely described in the language game of theoretical physics; not because there are regions in which physics is false, but because, to use Aristotelian language, the world has many levels of form, and there is no realistic possibility of reducing them all to the level of fundamental physics.[18]

One of Putnam's favorite examples in this sense was that, depending on what our interests are, we can correctly and usefully describe a chair in the alternative languages of carpentry, furniture, design, geometry, or etiquette. Each of those descriptions is useful in its specific way, without being reducible to any of the others. Moreover, there is no fundamental and unifying theory of what being a chair is, so to speak. And this is true of a vast amount of entities (possibly of all of them, with the exception of the entities of microphysics), since they can all be described in different ways; and this is not just because of conceptual relativity, but also because things have different properties that belong to different ontological regions. A poem is real, for example, but certainly its properties cannot be accounted for by any natural science; or moral (or immoral) actions exist, but not in the ontological region of physics – even if, for Putnam, being a poem or being a moral action are properties that globally supervene on physical properties.

According to Putnam, the old ontological project of providing a unified inventory of the universe, which would supposedly encompass the referents of all possible objective statements – a project of which contemporary metaphysical realism is a very clear expression – has made us wandering in Cloud Cuckoo Land for too long.[19] And this means, for Putnam, that Ontology with a capital "o" is a dead project. However, another form of ontology (one with a lower-case initial) is still possible, i.e., the search for the entities our best theories and practices commit us to. The latter project, however, cannot be carried out if one is driven by the ideological bias that there has to be one, and only one, true

---

17   *Ibid*., p. 64.

18   *Ibid*., p. 65.

19   H. Putnam, *Ethics without Ontology*, cit.

theory of the world; nor can it be carried out without noticing that there are different mutually irreducible ontological levels. And it is a pragmatic question which level is relevant to a particular discursive practice. In this light, Putnam's liberal naturalism incorporated *causal pluralism* (in his view causation and explanation are inextricably interconnected notions)[20] and *the refusal of the fact-value dichotomy* (since, in his view, values and normativity are ubiquitous).[21]

Putnam's view of the mind changed in parallel with his attempts at defining a satisfying form of non-reductive naturalism. In this sense, Putnam remained a functionalist in a broad sense since he continued to understand the mind in terms of its "functions", both internal and external, although not now characterized in computational terms. His later conception of the mind as a "structured system of object-involving abilities",[22] however, built upon his long-standing commitment to semantic externalism by taking seriously that there is no interface between the mind and world in perception or conception:

> The identification of naturalism with such "reduction programs" as the program of reducing the intentional to the non-intentional, or dispensing with intentional and normative notions entirely is a mistake, and I have been explaining how that mistake led me, at one time, to abandon the very realist intuitions with which I started. Some naturalists are reductionists, to be sure, and reduction programs *have* sometimes succeeded, but counting oneself as a naturalist does not require one to subscribe to reduction programs that are, as far as we can now tell, utterly unrealistic. The liberalized functionalism I advocate is an antireductionist but naturalist successor to the original, reductionist, functionalist program. For a liberalized functionalist, there is no difficulty in conceiving of ourselves as organisms whose functions are, as Dewey might have put it, "transactional", that is environment-involving, from the start.[23]

According to the later Putnam, then, a feasible liberal functionalism about the mind has to be conceived in the context of a general liberal naturalism that reconciles what science tells us about the world with the irreducibility of the intentionality of the mental. And I think that this idea is one of the most important ones that this great philosopher left us with.[24]

---

20   H. Putnam, *The Threefold Cord. Mind, Body, and World*, Columbia University Press, New York 1999, pp. 137–150.

21   H. Putnam, *The Collapse of the Fact/Value Dichotomy and Other Essays*, Harvard University Press, Cambridge (MA) 2002, and Id., "The Fact/Value Dichotomy and Its Critics", in *Philosophy in an Age of Science*, cit., pp. 283–298.

22   H. Putnam, "Replies", *The Philosophy of Hilary Putnam*, monographic issue of *Philosophical Topics*, 2, 1 1992, p. 256.

23   Putnam, "Corresponding with Reality", cit., pp. 82–83.

24   I am greatly indebted to Hilary Putnam for the innumerable conversations I had with him in regard to the issues discussed in this article.