True or Rational? A Problem for a Mind-Body Reductionist

**Introduction[[1]](#footnote-0)**

The problem presented in this essay touches upon an issue of reductionism in philosophy of mind, represented by various forms of physicalism. A whole range of reductionist physicalisms (hereafter I use words “physicalism” and “reductionism” interchangeably, except in a passage, where the so called non-reductionist physicalism is discussed briefly) has to answer a question which poses a serious difficulty shaped in the form of a dilemma.

The dilemma goes as follows: a physicalist must accept that either physicalism is false or it lacks rational grounding, since the question “How can be the truth and the rational justification of physicalism compatible?” is unanswerable in a way expected by such a philosopher. As I try to show, these two features: the rational grounding and the truth of this belief exclude each other and as a result a reductionist must face an unsolvable problem concerning the possibility of their simultaneous assertion. Unsolvable, as every time one tries to assert reductionism’s truth, it proves to be self-undermining. The argument for this thesis has been provided by various formulations of an objection originating from Karl R. Popper’s remarks on materialism (Popper, Eccles 1983) and all together they pose a danger for the identity theory, epiphenomenalism, behaviourism and the mind-body functionalism. A physicalist might try to answer this objection leading them to the conclusion that their belief is either false or irrational (which is an idle task). They may also try to adopt other, non-reductive approaches, yet resulting in their ceasing to a be a reductionist.

In the first part of this essay I present the argument against physicalism and distinguish it from other arguments frequently proposed against it (such as the inverted qualia argument) as well as from other dilemmas (just to mention now the famous Hempel’s dilemma). Then I go on to discuss possible answers of a reductionist and show their irrelevance in some cases and non-reductionist consequences in others. Finally, I sum it up in short conclusions.

**The dilemma for a reductionist – the neopopperian argument against physicalism**

The dilemma has been partially presented in *The Self and its Brain,* written together with John C. Eccles, where Popper argues that the identity theory is incompatible with rationalism. Among various remarks showing inconsistency of this or other forms of materialism (such as that: *(...) the radical physicalist must adopt radical behaviourism and apply it to himself: his theory, his belief in it, is nothing; only the physical expression in words, and perhaps in arguments - his verbal behaviour and the dispositional states that lead to it - is something.* [Popper, Eccles 1983, 60]), he quotes John B. Haldane’s objection against materialism. It goes as follows: *(...) if materialism is true, it seems to me that we cannot know that it is true. If my opinions are the result of the chemical processes going on in my brain, they are determined by the laws of chemistry, not of logic.* (Haldane as cited in Popper, Eccles 1983, 75). As Popper points out, the existence of computers makes the argument easily refutable, as their functioning based on laws of physics on no account contradicts functioning according to laws of logic. Later on, however, he goes on to develop Haldane’s argument in the form of a dialogue between an interactionist and a physicalist and reaches a conclusion that the physicalist lacks an external point of reference while assessing his beliefs. For the interactionist such an anchor is provided by the objects of the World 3 (such as logic in general), whose existence is denied by the physicalist. Consequently, the physicalist’s stance is irreconcilable with rationalism (Popper, Eccles 1983, 81; see the discussion in Pawłowski 2019).

Similar objections are proposed in Edward Feser’s discussion of Friedrich A. von Hayek’s causal functionalism and arguments against it presented by Popper (Feser 2006a). According to Popper, as Feser notices, functionalism is self-undermining, since it ignores a core fact that distinctive properties of human language are its argumentative and descriptive functions (in contrast to the signal one, characterizing for example bees’ system of communication): if one reduces a proposition’s meaning to its functional role between inputs and outputs of a given (computer or neural) system, a person’s having it has nothing to do with a logical inference from other beliefs. And then none of these propositions is rationally justified, thus also a proposition asserting physicalism’s truth (Feser 2006a, 307-310). Feser supports this approach with another example: if causal powers of certain beliefs do not depend on their contents or meanings (as they are rooted in electrochemical properties of neural processes), we cannot insist that our thoughts “Socrates is a man” and “all men are mortal” serve as rational justification of the thought “Socrates is mortal” due to their meaning. Was their content different, like “Fido is a dog” (but causal powers the same), nothing would change. Thus no belief is rationally justified, and *a fortiori* also the belief that physicalism is true (Feser 2006b, 150-154).

Mariusz Grygianiec offers a brief reconstruction of the argument against the identity theory in a form of a syllogism (Grygianiec 2016). Analogously, it can be also presented in this framework as an objection against functionalism, which I proposed in (Pawłowski 2019):

1. If functionalism is true, then every mental state (including all beliefs) is identical[[2]](#footnote-1) to a functional state.
2. If every belief is identical to a functional state, then every inference comes down to passing from one functional state of a brain to another, given definite causal connections (laws) bounding inputs and outputs of the system and described by a theory T.
3. If every inference comes down to passing from one functional state of a brain to another, given definite causal connections bounding inputs and outputs of the system and described by a theory T, then these states possess their causal powers only because of their functional properties (not owing to their meanings or contents that they are bound with), ascribed basing on this theory.
4. If these states possess their causal powers only because of their functional properties (not owing to their meanings or contents that they are bound with), ascribed basing on a theory T, then there is no such belief that could serve as a justification for a different one (since only functional connections, bounding inputs and outputs of the system define which conviction is entailed by which).
5. If there is no such belief that could serve as a justification for a different one, then no belief can be rationally justified.
6. If no belief can be rationally justified, then also functionalism cannot be rationally justified.
7. Conclusion: if functionalism is true, then it cannot be rationally justified.

Conclusion: functionalism is false or cannot be rationally justified (Pawłowski 2019).

Worth noticing that the same applies not only to the identity theory (when the first premise will be the following: *If materialism is true, then every mental state (including all beliefs) is identical to a neurophysiological state* [compare Grygianiec 2016]), but also to epiphenomenalism, because it explicitly postulates mental states’ lack of causal powers (compare Jackson 1982), and to behaviourism as a simpler version of functionalism. As a consequence, a mind-body reductionist has to answer the question, how he or she can believe in the truth and the rational justification of their position at the same time. This, according to that argument, is impossible, since either physicalism is true or cannot be rationally grounded (its truth entails its irrationality). Although a physicalist might employ various strategies to avoid facing this agonizing conclusion, as will be discussed later on, none of them remains valid.

It could be also noted that a similar argumentative strategy is taken by the opponents of the (more generally understood) metaphysical naturalism. They convince that it leads to its own self-refutation, just to mention Alvin Plantinga (2002) and Thomas M. Crisp (2016). Analogous objection was also provided by C. S. Lewis, against whom Elizabeth Anscombe famously argued, accusing him of confounding reasons and causes (Anscombe 1981). Here, however, it should be pointed out that the reductionism, being a stronger thesis, claims that mental states are “nothing more and above” than the “so-and-so” (physical states, natural states, etc.). Consequently, the critique need not be referred to that case.

At the same time, this dilemma posing a danger for different forms of physicalism (the identity theory, epiphenomenalism, behaviourism and functionalism) should be distinguished from other arguments against it. For example, similarity to another well-known argument used against physicalism, the Hempel’s dilemma, is rather superficial. Hempel’s dilemma entails that either physicalism should be defined in terms of contemporary physics and then most presumably is false or in terms of future or ideal physics and then is trivial (Hempel 1969; Stoljar 2017). Similarly, Feser argues that scientism is either false, because its ambition to explain the whole reality only with a use of scientific methods is unattainable, as they must always rest on some prescientific (philosophical) assumptions, or trivial, when under “science” one understands also these assumptions (Feser 2014, 10-13). Here, however, the alternative lies not between this belief being false or trivial, but rather between the falsity and the lack of rational justification. In both cases the latter part of the alternative (triviality/irrationality) includes the truth of physicalism. However, what makes them different are obvious character in the first, and irrationality of such truth in the second case.

Moreover, the dilemma “the falsity or (the truth and) the irrationality” is for sure a way different argumentative strategy against physicalism than most objections proposed in philosophical debate. For the sake of brevity I refer to arguments considering probably the most significant form of physicalism nowadays, functionalism. While the inverted qualia argument and other objections such as the what-is-it-like-to-be argument by Thomas Nagel (1974) or the Mary argument proposed by Frank Jackson (1986) try to show that functionalism fails to explain the whole richness of the reality’s aspects, Ned Block’s argument proves that it cannot guarantee the possession of consciousness to these, and only these structures to which it should be ascribed (it cannot exclude the “liberalism” and the “chauvinism” at the same time) (Block 1993), and the Kripkean objection against functionalism posed by Joseph Levine (1993) simply attempts to prove its falsity, the here presented dilemma offers a much more interesting solution. Functionalism (and with it some other reductionist approaches) might be true, yet for an undeniably high price: admitting it, ambitions to maintain rational approach must be abandoned. On the other hand, the desire for rationality entails its falsity.

**(The impossibility of) physicalist replies**

Confronted with the syllogistic version of the argument, a physicalist might attempt to answer it in many ways. The first solution is to try to challenge the validity of this inference through overruling one of the premises.

It might be argued that the premise 1) is too wide: it should not be spoken about all mental states, but only about some of them. Ned Block, for example, makes a restriction that only “narrow” mental states should be involved, namely those ones, whose truth-conditions can be found only inside a given person (Block 1993). The stipulation is intended to guarantee that the Twin Earth critique of Hillary Putnam (1973) will not be applicable here. As a result of it, the entire inference would collapse. Block offers a counterexample (meant to prove “wide” mental states’ irrelevance in this field) according to which my duplicate, created by the exchange of all atoms would be functionally identical to me, yet he would lack some of my memories. In such cases it should be denied that all beliefs are identical to some functional states and thus the premise 1) should be refuted (see Pawłowski 2019).

However, as I refer (Pawłowski 2019), these two replies, Putnam’s counterexample of the Twin Earth and Block’s narrow mental states do not have to be accepted: it might be shown that it is simply the case that the world in which water is H2O offers a different input than in the world in which water is XYZ, while considering identical functional systems. Even if these convictions generate the same outputs, they will be different functional states, since inputs *de facto* differ. It seems that one does not have to “save” the initial assumptions of functionalism from Putnam with such a reply, simply because there are much easier counterarguments. Moreover, the accurate duplicate example is in turn rather untenable, as it is in no case sure, whether the duplicate would be characterized by psychological continuity with me or not. Thus it is also not sure if he really would not have the same beliefs as me. One does not have to accept this reply. For instance, Derek Parfit argues that this is exactly the case that my duplicate would be psychologically continuous with me (Parfit 1984, 239; see Pawłowski 2019). Consequently, both problematic claims: the one proposed by Putnam and the one defended by Block (which tries to answer the Twin Earth example) seem dubious. At the same time, a functionalist, willing to explain the phenomenon of mind, should not simply ignore some mental states as irrelevant: functionalism must be a general theory of mind or it will not be such a theory at all. Thus, however controversial the premise 1) could seem, it is fully justified and reasonable.

While the premises 5) and 6) are rather uncomplicated tautologies and the premise 2) is an uncontroversial description of the functionalist position (as presented in [Block 1993] or [Levine 1993]), a physicalist might argue that one should not so easily accept the premises 3) and 4) and the requirement of justification and rationality of a belief following from them. They might claim that this requirement is too restrictive, similarly as it is proposed in various discussions with sceptics, who can be criticized for their inflated expectations towards cognitive certainty (see Wittgenstein 1975). Rational justification, a physicalist will say, is exactly this: the fact of a belief’s accurate inference from other beliefs, basing on causal powers described by a given theory T with regard to appropriate inputs and outputs. There is nothing more behind it and whoever demanding more from it expects the impossible.

A simple counterexample might be proposed to this objection. If meanings and contents of beliefs play no role, since their rationality and the level of justification are rooted only in causal powers of a given functional system, no mistake would be possible. Was my brain so designed (natively or resulting from a surgery), that all operations “2+2=” would lead to the result “5”, everything would be perfectly fine. Similarly to previously discussed examples by Feser with Socrates and Fido, we need meanings or contents of mental states to be able to speak about their rational inferences. Without an external point of reference other than the causal powers (also causal powers of the others’ brains), such as standards of rationality, we would not have any point of reference. Thus the critique of inflated standards of rationality leads to its understated level.

Being unable to rejoin the argument, a reductive physicalist may adopt another view. They might be inclined either to a weaker or a stronger approach. The weaker is represented by the so called non-reductive physicalism, generally exemplified by views accepting the principle of supervenience - mental states’ dependence on physical states as their “base states” - yet without an attempt to reduce the former to the latter. However, one does not have to remind various difficulties non-reductive physicalism encounters, just to mention the causal exclusion problem (see Kim 1998), to spot the irrelevance of this strategy. Adopting non-reductive physicalism, a physicalist simply ceases to be a reductionist. Similarly with a stronger view, that is eliminativism - an attempt to eliminate mental states as objects of the non-scientific common-sense psychology. Whenever a reductive physicalist tries to become an eliminativist and excludes mental states as non-existing, he or she no longer presents a reductive attitude.

Finally, a physicalist might attempt to ignore the results of the dilemma and overrule the principle of truth, correspondingly with Friedrich Nietzsche’s remarks from *On Truth and Lies in a Nonmoral Sense*:

*In some remote comer of the universe, poured out and glittering in innumerable solar systems, there once was a star on which clever animals invented knowledge. That was the haughtiest and most mendacious minute of "world history" - yet only a minute.* (Nietzsche 1976, 42)

- or follow Richard Rorty’s liberal ironist’s attitude and reconcile with the impossibility of a fully objective vocabulary describing the world (Rorty 1993). Then, they will say, physicalism is only a useful tool for the description of the world. They might do that, but no longer can their beliefs be described as rationally justified.

Consequently, any time a reductive physicalist will try to avoid the consequences of the presented dilemma, he or she will have to cease being a reductionist or a rationalist – that is accept one of the question’s results.

**Conclusions**

The dilemma, originating from Popper’s writings and creatively developed by other authors, poses a great danger for mind-body reductionists. Its conclusion that reductive forms of physicalism force their representatives to admit either the falsity or the lack of rational justification of their beliefs is bounding, because all attempts to face it are idle. The inference is logically correct and based on reliable premises: thus any attempt to ignore its results requires abandoning physicalist reductionism. One option is adopting non-reductive physicalisms (or trying to eliminate the concept of mental states), another – falling into nonrational beliefs.

A physicalist then, confronted with the question “How can you believe in the truth and rational justification of your view at the same time?” is unable to reply it, since the argument-dilemma proves these two features to be incompatible. Any sensible answer leads to a change in the intellectual position and abandoning reductionism.

**References:**

Anscombe, G. E. M. 1981. „A Reply to Mr C. S. Lewis's Argument that "Naturalism" is Self-Refuting“. In: *The Collected Philosophical Papers of G. E. M. Anscombe, Volume Two, Metaphysics and the Philosophy of Mind*. Oxford: Basil Blackwell, 224-232.

Block, N. 1993. “Troubles with Functionalism”. In: *Readings in the Philosophy of Science*, ed. by A. I. Goldman, Cambridge: MIT Press, 231-253.

Crisp, T. M. 2016. “On Naturalistic Metaphysics”. In: The Blackwell Companion to Naturalism, ed. by K. J. Clark, London: Wiley Blackwell, 61-74.

Feser E. 2006a. “Hayek the cognitive scientist and philosopher of mind”. In: *The Cambridge companion to Hayek*, ed. by E. Feser. Cambridge: Cambridge University Press, 287−314.

Feser E. 2006b. *Philosophy of mind. A beginner’s guide*. Oxford: Oneworld Publications.

Feser E. 2014. *Scholastic metaphysics: a contemporary introduction*. Piscataway: Transaction Books.

Grygianiec, M. 2016. Die popperschen Herausforderungen für den Materialismus. *Logos i Ethos*, vol. 2(42), 103-115.

Hempel, C. 1969. “Reduction: Ontological and Linguistic Facets”. In: *Essays in Honor of Ernest Nagel*, ed. by S. Morgenbesser et al. New York: St Martin's Press, 179-199.

Jackson, F. 1982. Epiphenomenal Qualia. *The Philosophical Quarterly*, 32(127), 127-136.

Jackson, F. 1986. What Mary Didn’t Know. *The Journal of Philosophy*, vol. 83(5), 291-295.

Kim, J. 1998. *Mind in a Physical World: an Essay on the Mind-Problem and Mental Causation.* MIT Press.

Levine, J. 1993. Materialism and Qualia: the Explanatory Gap. *Pacific Philosophical Quarterly*, vol. 64, 354-361.

Nagel, T. 1974. What Is It Like to Be a Bat? *The Philosophical Review*, vol. 83(4), 435-450.

Nietzsche, N. 1976. “On Truth and Lie in an Extra-Moral Sense” (frag.). In: *The Portable Nietzsche*, sel. & trans. by W. Kaufmann. London: Penguin Books, 42-47.

Pawłowski, M. 2019. Neopopperowski argument przeciw funkcjonalizmowi. *Filozofia Nauki*, vol. 3(27), 77-86.

Plantinga, A. 2002. “Introduction”. In: *Naturalism Defeated? Essays on Plantinga’s Evolutionary Argument against Naturalism*, ed. by J. K. Beilby. Ithaca: Cornell University Press.

Popper, K. R. & Eccles, J. C. 1983. *The Self and Its Brain*. London: Routledge & Kegan Paul.

Putnam, Hilary. 1973. Meaning and Reference. The Journal of Philosophy, Vol. 70(19), 699-711.

Rorty, R. 1993. *Contingency, Irony and Solidarity*. Cambridge: Cambridge University Press.

Stoljar, D. 2017. "Physicalism". In: *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), ed. by E. N. Zalta. Online:

<https://plato.stanford.edu/archives/win2017/entries/physicalism>.

Wittgenstein, L. 1975. *On Certainty*. Ed. by G. E. M. Anscombe & G. H. von Wright, trans. by G. E. M. Anscombe & D. Paul. Oxford: Basil Blackwell Ltd.

1. I would like to thank Dr hab. Mariusz Grygianiec, whose helpful remarks directed my attention to some problems mentioned in the paper. [↑](#footnote-ref-0)
2. One does not have to differentiate here between the type identity and the token identity. [↑](#footnote-ref-1)