https://doi.org/10.15407/fd2023.03.093
UDC 141.1:167.2

Dmytro SEPETYI, Doctor of Sciences in Philosophy,
Associate Professor at the Department of Social Studies
of Zaporizhzhia State Medical and Pharmaceutical University,
26, Mayakovskoho A., Zaporizhzhia, 69035
dmitry.sepety@gmail.com
https://orcid.org/0000-0003-2110-3044

THE KNOWLEDGE ARGUMENT AGAINST MATERIALISM AND THE STRATEGY OF PHENOMENAL CONCEPTS

Materialism/physicalism that generally dominates in the contemporary analytic philosophy is challenged by fairly powerful anti-materialist arguments, notably the zombie argument (most influentially defended by David Chalmers) and the knowledge argument (the most widely discussed version of which was advanced and defended by Frank Jackson). These arguments highlight the explanatory gap from the physical (which, if materialism is true, should constitute everything that exists, including consciousness) to phenomenal mental states, the principal impossibility to explain the latter by the former, and from this conclude that phenomenal consciousness is not physical, and so materialism is false. Materialist philosophers attempt to neutralize these arguments in several ways, the most influential of which is the strategy of phenomenal concepts. This article analyzes the main points of this debate with a focus on the knowledge argument, examines and responds to the main objections to the knowledge argument — that it should be mistaken because the alternative is epiphenomenalism, which is unacceptable; that no new knowledge but only new capacities and/or acquaintance are involved; that the knowledge is the same but in different forms; that the knowledge argument affects only type physicalism but not token physicalism. The case is made that psychophysical identities assumed by a posteriori physicalism are unexplainable in principle, and the postulation of brute unexplainable psychophysical identities glossed over by the strategy of phenomenal concepts amounts to dogmatic commitment (motivated by scientism) to a view despite its apparent falsity and its unintelligibility (the impossibility to explain how it can be true), made less unpalatable by offering an ad hoc theory about the mindbrain arrangement that makes us unable to see how the view can be true. As opposed to this, the position of the supporters of the knowledge argument and the zombie argument can be seen as guided by the principle of rational trust in obviousness and our capacities of judgement.

Keywords: *materialism, physicalism, dualism, knowledge, phenomenal concepts.*

Citation: Sepetyi, D. (2023). The Knowledge Argument Against Materialism and the Strategy of Phenomenal Concepts. *Philosophical Thought*, 3, 93—110. https://doi.org/ 10.15407/fd2023.03.093

In the contemporary analytic philosophy, the general domination of materialism/physicalism is undermined by pretty weighty anti-materialist arguments, such as the zombie argument (Kirk, 1974a; 1974b; Chalmers, 1996; 2010) and the knowledge argument (Jackson, 1982; 1986; Robinson, 1982; 2016; Fumerton, 2013). These arguments highlight the explanatory gap (Levine, 1983; 2001) from the physical (which, if materialism is true, should constitute everything that exists, including consciousness) to phenomenal mental states, the principal impossibility to explain the later by the former, and from this conclude that phenomenal consciousness is not physical, and so materialism is false. Materialist philosophers attempt to neutralize these arguments in several ways, the most influential of which is the strategy of phenomenal concepts. This article analyzes the main points of this debate with a focus on the knowledge argument, examines the main objections to the knowledge argument and responds to these objections. The case is made that psychophysical identities assumed by *a posteriori* physicalism are unexplainable in principle, and the postulation of brute unexplainable psychophysical identities glossed over by the strategy of phenomenal concepts amounts to dogmatic commitment (motivated by scientism) to a view despite its apparent falsity and its unintelligibility (the impossibility to explain how it can be true), made less unpalatable by offering an *ad hoc* theory about the mindbrain arrangement that makes us unable to see how the view can be true. As opposed to this, the position of the supporters of the knowledge argument and the zombie argument can be seen as guided by *the principle of rational trust in obviousness and our capacities of judgement*.

1. The knowledge argument

The authorship of the argument is usually ascribed to the Australian philosopher Frank Jackson, who formulated its "classical" version in the article "Epiphenomenal *qualia*" (Jackson, 1982). However, it is not quite right to attribute the argument to any one philosopher: in the very 1982, Howard Robinson formulated a nearly identical argument (Robinson, 1982), and pretty similar reasoning can be found in a number of earlier philosophical works. Probably, the earliest close predecessor was Leibniz's "mill argument":

"... perception and that which depends on it cannot be explained mechanically, that is, by means of shapes and motions. And if we suppose that there were a machine whose structure makes it think, feel, and have perception, we could imagine it increased in size while keeping the same proportions, so that one could enter it as one does with a mill. If we were then to go around inside it, we would see only parts pushing one another, and never anything which would explain a perception. This must therefore be sought in the simple substance, and not in the compound or machine" (Leibniz, 2014: p. 157).

Other examples of reasoning along the lines of the knowledge argument can be found in works of several philosophers of 20th century: Charlie Dunbar Broad (1925: pp. 70-72), Bertrand Russell (1927: p. 389), Brian Anthony Farell (1950: p. 183), Herbert Feigl (1967: pp. 139-140), Paul Meehl (1966: pp. 151-158), Thomas Nagel (1974).

In a general form, the argument can be formulated as follows:

1) Subjective experiences – sensations, perceptions, etc. – have a specific qualitative character – how it feels, what it is like for the experiencer (for example, what it is like to feel toothache as distinct from other pains, or from tickle or any other sensation; what it is like for a person to see red rather than green, or rather than to smell a rose).

2) No matter how full our knowledge of physical (physiological) entities, properties, processes involved with some experience, this knowledge does not contain (imply) the knowledge of the specific qualitative character of the experience – of how it feels. A person can know whatever she likes about everything physical involved with an experience while knowing nothing – and having no resources to know anything – about the experience itself, what it is like for the experiencer. So, the knowledge of the physical, however full, does not contain (imply) the knowledge of experiences, their subjective qualitative character.

3) Hence, experiences are distinct from anything physical, that is, they are non-physical.

For example, we can know whatever we want about physical processes involved with sonar perception of bats; however, no such knowledge will enable us to know what it is like to be a bat – that is, what it is like to have a bat's sonar experiences (Nagel, 1974). Jackson proposed to imagine a person (woman named Mary) who never had colour visual experiences (from the very birth, she lived in a special room in which everything looked black-and-white) but has all the physical (including physiological) knowledge relevant to colours and colour experiences: the corresponding frequencies of electromagnetic lightwaves, photon energies, physiological processes evoked in retina and brain, etc. No such knowledge could enable Mary to know how it feels (what it is like) to have colour experiences of green or red (Jackson, 1982; 1986). Likewise, Robinson proposed to imagine a deaf scientist who is a world leading expert on the physical (including physiological) processes involved with hearing (Robinson, 1982: p. 3). Or we can imagine a person incapable, from birth, of having olfactory experiences; however much she knew about the physics-chemistry-physiology of olfactory perceptions, she cannot know what the smell of a rose is like.

Now let us consider the main objections against the knowledge argument.

2. Frank Jackson's surrender and the objection from the unacceptability of epiphenomenalism

It should be noted that Frank Jackson has eventually (in the mid of 1990-ies) renounced the knowledge argument and joined the materialist camp; however, on his own explanation (Jackson, 2004), he did it not because he saw some serious flaws in the argument but because other considerations convinced him that *materialism must be true* (and hence, *something must be wrong* with the argument that entails the falsity of materialism, even if we have no good idea as to what it is). However, this conversion was an effect of Jackson's initial commitment to the tenet of *the causal*

closure of the physical, which claims that all physical events, if caused at all (we make this reservation in order to leave place for the possibility of purely accidental events, perhaps of the kind involved in quantum mechanics), are caused entirely by other physical events; no physical events are ever caused, entirely or partially, by some nonphysical factors. This commitment left Jackson with a choice between materialism, which claims that phenomenal consciousness is somehow identical with some physical processes in the brain (or their functional aspect), and epiphenomenalism, which denies such identity but bereaves consciousness of any causal powers with respect to physical events, including human behaviour and, in particular, our talk or writing about our phenomenal mental states. In 1980-ies, Jackson preferred epiphenomenalism because the knowledge argument convinced him that materialism is false; after 1996, Jackson thought that arguments against epiphenomenalism are weightier.

Besides epiphenomenalism, Jackson accepted a view that reductive physicalist explanation, although impossible for subjective qualities of experiences, *qualia*, is possible for cognitive mental states/processes. Accordingly, he held that thinking, unlike *qualia*, is ontologically reducible to a physical (neurophysiological) basis. In conjunction with the view that *qualia* are causally idle (have no effect on any physical processes), this produced an unacceptable severance between *qualia* and thinking about them. If our thinking about *qualia* does not depend causally on *qualia* but is constituted by causally all-sufficient physical processes, then it follows that we would think about *qualia* in exactly the same way even if *qualia* did not exist! This absurdity made Jackson eventually to abandon epiphenomenalism and join materialists.

By the way, a similar view – and similar vacillation in the face of the dilemma that forces the choice between alternatives that seem each apparently false – was characteristic not only of Jackson but of some other recent philosophers as well, – in particular, of the "father" of the zombie argument, Robert Kirk. Kirk made nearly the same way with respect to the zombie argument as Jackson with respect to the knowledge argument (Kirk, 2005).

However, the absurdity of the severance between *qualia* and thinking about them need not be blamed upon dualism. Rather, the likely culprit is the physicalist reductionism about thinking (cognitive physicalism), which was taken by Jackson and Kirk for granted.

One thing you should not do when discussing arguments against materialism is taking for granted *cognitive physicalism* – the view that thinking can be adequately reductively explained in materialist (functionalist) terms, and that *only* the physicalist reducibility of sensual-perceptual experiences-qualia (such as pain, tickle, visual, auditory or olfactory experiences) is problematic. Cognitive physicalism limits our choice to that between materialism and the most insipid variety of dualism (epiphenomenalism *cum* cognitive physicalism), while leaving out its more vigorous and defensible varieties. On the side of non-materialists, to assume cognitive physicalism is to give up into the uncontested domain of materialism the larg-

est and most important part of the human mental realm, and then try to fight for miserable leftovers. This endeavour would be hopeless because materialists would have all the advantages, and non-materialists – all the disadvantages of the acceptance of the materialist view about cognitive mental states.

Howard Robinson aptly criticised this approach to the knowledge argument:

"Those who like Jackson ... think that physicalism can be correct for everything but qualia are in an inconsistent position. The knowledge argument should not be cast in the form "physicalism can work for all other mental states but not for qualia", but in the form "even if it might look as if functionalism will work for less clearly introspectible states, such as thoughts, Mary's case shows that it will not work for qualia, and we can see from this that it does not work for thought – at least, a certain category of thought... – either. "" (Robinson, 2004: p. 72)

On the most common-sense and intuitively appealing view, our thinking, understanding, and volition are just as subjective by their nature as an experience of pain or of green colour. Whether or not they qualify as *qualia*, it is advisable for a non-materialist to put them on the same – phenomenal, subjective, non-physical – side.

Likewise, the choice between epiphenomenalism and materialism is a spurious dilemma, because we can decline both, together with the tenet of the causal closure of the physical.

3. The objection from capacities and/or acquaintance

The objection from capacities and/or acquaintance (Lewis, 1988; Nemirow, 2007; Tye, 2009) was advanced, as a rule, in the context of Jackson's version of the knowledge argument, the thought experiment with Mary in the white-and-black room. In this experiment, the claim that Mary, while having full knowledge of the physical facts relevant to colour perception, does not know, and cannot know, what colour visual *experiences* are like (what it is like for a normal human being in normal conditions to see something red or green), was supported by the proposition to imagine that eventually, after decades of "colour isolation", Mary gets into normal condition and finds out what it is like to see colours – what visual colour experiences are like. Jackson's opponents objected that what Mary acquires is not any new knowledge but some special relation of acquaintance or some new capacities (to recognise, recollect, etc.); accordingly, they claimed that before this colour initiation Mary did not lack any knowledge – she lacked only these capacities and/or acquaintance.

Defenders of the knowledge argument – such as Frank Jackson in 1980-ies (1986), Martine Nida-Rümelin (2004; 2007), Torin Alter (2007), Howard Robinson (1993; 2016), Richard Fumerton (2013) – responded that although it is quite possible and plausible that when acquiring qualitatively new experiences a person acquires acquaintance or new capacities, there is obviously more involved. Viz., there is something it is like for a person to have these experiences, how it feels. It is really the case – it is a fact – that it feels the way it does rather than somehow otherwise. And this is a special fact that can be known only through acquaintance with an experience, by experiencing it, and does not follow from any physical facts.

Subjective-phenomenal qualities of experiences are parts of reality, and they are lacking in the materialist (physicalist) picture of the world. Hence, this picture is incomplete, and materialism is false.

Another response to the objection from capacities and/or acquaintance is that even if Mary, in addition to full physical knowledge concerning colours and colour perception, acquires *isolated* colour experiences, and so gets acquainted with them and acquires the corresponding capacities of recognition, recollection, etc., she still cannot know how blue heaven, or a red rose, or green grass normally looks (what it is like to see it). Imagine, for example, that Mary's room has, instead of the constant white-and-blue illumination, the changeable colour illumination – on Mondays things look shades of gray, on Tuesdays – shades of green, on Wednesdays – shades of blue, on Thursdays – shades of red, etc. (Sepetyi, 2017: pp. 44-45). Or imagine that Mary was once shown, in normal illumination, green triangles, red rectangles, and blue circles without being told the names of these colours (Nida-Rümelin, 1998: pp. 52-56; 2004: pp. 243-247; 2007: p. 309). Mary can give these colours some "private" names, and think of her experiences of triangles, rectangles, and circles (or Tuesdays, Wednesdays, and Thursdays) in these terms. However, she still will not be able to know how a red rose or green grass looks when normally illuminated.

These variations help us to see that with acquaintance with new experiences, Mary acquires new *propositional knowledge*, and we can divide the acquisition of this knowledge into two stages. On the first stage, Mary acquires *phenomenal concepts* of red, green, and blue colours and propositional knowledge that there are experiences with such subjective-phenomenal qualities and she is capable of experiencing them (has mental dispositions to experience them in certain conditions). After naming these experiences, Mary can formulate the corresponding propositions[1]. Later, on the second stage (in the initial, Jackson's "classical" version of the knowledge argument these two stages are not separated), when Mary sees the world in the normal illumination, she acquires further propositional knowledge, about which of these experiences normal people in normal conditions have when they see green grass, blue heaven, and other coloured things, and what public names correspond to her "private" names of colours. All this new knowledge (acquired on both stages) cannot be derived from any knowledge of physical facts, is not entailed by it.

4. The objection of the same knowledge in a different form

Other opponents of the knowledge argument admitted that after the end of "colour isolation" Mary acquires new propositional knowledge but claimed that this new knowledge is not *knowledge of new facts* but knowledge of some already known (physical) facts *in a new form* (Horgan, 1984; Churchland, 1984; Tye, 1986; Loar, 1997; Lycan, 1990; Papineau, 2007).

[1] Here, I disagree with Nida-Rümelin, who suggested that on this stage, Mary does not acquire any propositional knowledge but acquires only phenomenal concepts that enable her to aquire new propositional knowledge later, when she will see natural things (heaven, grass, etc.) in the normal illumination (Nida-Rümelin, 1998: pp. 63, 66).

To this objection, it can be responded that it is incogent because full knowledge of fundamental physical facts (which, according to the conditions of the knowledge argument, is accessible to Mary) should entail any knowledge, in any form, of all ontologically derived facts on higher levels of description; so that a person who knows all the relevant facts at the fundamental level and has mastered the semantics of the terms at higher levels of description, will be in principle capable to make out any facts at higher levels of description. The principal impossibility to derive some facts (about consciousness) from physical facts means that the former are not ontologically derived from (constituted by) the latter, and so materialism is false.

As an example of the objection of the same knowledge in a different form, let us consider its version advanced by Paul Churchland (1984). Churchland suggests that we can know something about a thing under one description but not know this about this thing under another description and not know that it is the same thing. For example, I could know that there is a boxing world champion named Muhammad Ali and could know something else about a person named Cassius Clay, while not knowing that Muhammad Ali is Cassius Clay. If so, I would not know that Cassius Clay is a boxing world champion, although I know that Mohammed Ali is a boxing world champion, and although "Muhammad Ali" and "Cassius Clay" refer to the same person (Churchland, 1984: p. 32). However, the analogy is impotent against the knowledge argument, because the latter involves not merely *some* knowledge of physical facts relevant to certain experiences but *full* knowledge of such facts. Obviously, if I knew all facts about a person named "Cassius Clay", I could not fail to know that he is a boxing world champion, and that he has the assumed name "Muhammad Ali"! And if I could knew all physical facts about Cassius Clay, and if I understand the meanings of such terms as "boxing world champion", "name", "assumed name", I would be, in principle, in a position to know that Cassius Clay is a boxing champion and is Muhammad Ali, because boxing championship and naming are matters of specific human behaviour, which is constituted by some complex physical movements, and of the social conventions as to which behavioural patterns count as boxing championship and naming.

5. Does token physicalism fares better than type physicalism?

Amir Horowitz and Hilla Jacobson-Horowitz (2005) argue that the knowledge argument, even if it succeeds as an argument against type physicalism, fails against token physicalism.

It is worth noting that the names "type physicalism" and "token physicalism", and some corresponding explanations are misleading in suggesting that type physicalism assumes that there are identities between types of physical and mental states, whereas token physicalism assumes that there are identities only between tokens of physical and mental states. Such a straightforwardly-token physicalism could hardly be coherent. If you hold, as a physicalist, that all that exists is physical, then you cannot consistently deny that all qualitative differences in the world are constituted by some physical differences, and so there should be some mapping

of types of mental states into types of physical states. The real point of token physicalism seems to be multiple realisability: a certain mental state can be realised in a number of different physical ways. This applies to different sentient creatures: with human beings, pain can be realised by one kind of physical (physiological) states in the brain, with frogs – by a different kind of physical (physiological) states, with sentient robots that can (if physicalism is true) be constructed in some future – by an even more different kind of physical states. This can apply also to the same brain: perhaps even with human beings, pain is sometimes realised by one kind of physical states, and sometimes by another.

The so called token physicalism (which would be better described as multiple realisability physicalism) still involves sort of type identities. If a certain mental state can be realised in a number of different physical ways, then surely, there are physical states that are realisations of this mental state, and there are physical states that are not realisations of this mental state. Therefore, the mental state can be described as identical with *the physical type that is the disjunction* of all types of physical states that are realisations of this mental state. However, the disjunction can be indefinitely (in principle, even infinitely) long and complicated, so that we can be unable of giving a precise and full specification of a mental state in physical (or physiological) terms.

Now the point of Horowitzs is that if a phenomenal mental state, such as phenomenal red (the subjective quality, "what it is like", of the visual experience of seeing red), can be multiply physically realized, then one (Mary in her black-and-white room) can know all about particular realisations of such a phenomenal mental state but be unable to know what this phenomenal mental state is like. This is possible because "the knowledge of the fundamental physical properties does not guarantee *the possession of the concepts* of the higher-order properties" (Horowitz, Jacobson-Horowitz, 2005: p. 58).

That is, of course, true, but fails to invalidate the knowledge argument. There are at least two ways to decline Horowitzs' objection, each sufficient on its own.

First, consider the variations of the knowledge argument of the kind advanced by Nida-Rümelin (1998: pp. 52-56; 2004: pp. 243-247; 2007: p. 309), where Mary does possess the relevant phenomenal concepts (of phenomenal red, green, and blue) under other (private) names while not knowing anything about physical properties that are responsible for the corresponding experiences. In that case, when presented with physical properties responsible for colour experiences and said what colours they are responsible for, Mary will still be unable to know what it is like to see red, or green, or blue.

Second, consider how we acquire other physically respectable concepts that involve multiple realisability. One possibility is that we acquire a concept by being given a definition, which would be an explicitly stated disjunction of all possible realisations. Another possibility is that we acquire a concept by being presented with, or given descriptions of, several instances, and we are capable of grasping some similarity between these instances and of using that similarity to recognise

other instances that we can meet in future as instances of the same kind (subsumable under the same concept). In the extreme case, we can be presented with only one instance, being said that it is X, and form the concept of Xs as things sufficiently like that instance. However, if Mary's case was like that, she could certainly acquire phenomenal colour concepts in her black-and-white room, in one or another of these ways, and so be able to know what it is like to see red, green, or blue. However, she cannot.

6. Howard Robinson: Leibniz's Law as a challenge for physicalism

The objections against physicalism so far discussed are based on the typical construal of the knowledge argument as one that proceeds from an epistemological premise (about what we know or do not know) to an ontological conclusion (about the non-physical nature of phenomenal mental states). However, Howard Robinson, in the book *From the Knowledge Argument to Mental Substance* (2016), argued for another construal: the genuine premise of the knowledge argument is the fact of the existence of mental states having specific subjective qualitative character that is not at all like any physical properties and cannot imaginably be constituted by such properties:

"... the properties ascribed to neural states by science and the qualitative properties revealed in the WIL {"what it is like"} of experience seem entirely different, and, by Leibniz's Law, things with different properties cannot be identical. So the problem can be put as follows: The physicalist claim that phenomenally conscious states are identical to, or in some other way "nothing over and above", physical states or processes, seems to run up against Leibniz's Law. The challenge for the physicalist is to say what, if anything, it could be about a physical state or process that constituted it as, or made it count as, or made it to be, an experience with phenomenal character without imputing to it the kind of non-physical property that lead to the Leibniz Law problem" (Robinson, 2016: p. 77).

As Robinson and other opponents of physicalism amply argued, so far, materialists failed to meet the challenge. And it is arguable that this cannot be done, in principle.

To see this, let us compare the situation with psychophysical identities postulated by physicalists and the situation with known unproblematical, "respectable" identities, such as water≡H_20, heat≡the average kinetic energy of molecules, Hesperus≡Phosphorus≡Venus. All such identities can be explained by a reductive story

• either about how the observable (macro)properties (by which we identify water, heat, etc.) can be constituted by properties, structures, and dynamics of their constituents (such as molecules, atoms, etc.) on a lower, usually microscopic, level (these microconstituents and their properties are not directly observable, but they are hypothesised exactly for the purpose, and in such a way, that they could explain the observable phenomena);

• or about how a thing can *appear* in two or more different ways (can have two or more distinct *appearances*) in different circumstances: because the appearances are different, we can mistakenly think that we have to do with two different things, whereas in fact, there is only one thing that manifests different sets of its properties, or perhaps even the same properties in different perceptual modes (for example, a shape by vision and by touch); anyway – the differences are eventually not in the thing itself but in the perceptual mental states it evokes in the observer – the very perceptual states by means of which a person identifies the thing and its properties.

However, such explanatory reductive stories are impossible for phenomenal mental states, and without them the psychophysical identities postulated by materialists are unintelligible.

The impossibility of the explanatory story of the first kind is the hard problem of consciousness: for any physical entities that have no phenomenal mental states (in particular, the microconstituents of which our brains are composed), however they are arranged in space (whatever spatial structures they form) and however this arrangement changes with time, it seems obvious that these structures and dynamics do not entail there being subjective experiences and experiencers.

Functionalism does not solve the problem. For the argument that for any multitude of purely physical entities and events (that involve nothing irreducibly subjective on the fundamental level), however it is ordered in space and time, its presence does not entail the presence of subjective mental states (and the mental subject whose states they are), it does not matter whether we describe this multitude in low-level physical terms or some higher-level functional terms. If "function" is understood in a way that does not involve subjective mental states, as structural elements of some abstract system of relations between physical processes in the organs of sense and perception (sensory "input") and physical movements of the parts of the body that constitute behaviour (behavioural "output"), or as contributions to behavioural patterns and maintenance of the organism's integrity, whereas behaviour and organisms are considered as constituted by purely physical elements, structures, and patterns of dynamics (such as movements of hands, legs, lips, etc.), then physical realisation of these functions does not entail there being anything subjective (phenomenal).

What about the explanatory story of the second kind? Can we explain the identity of a phenomenal mental state with some physical state (some spatial structures and dynamics of the microconstituents of the brain) by saying that it is the same state X that appears to us in two different ways? No, we cannot. The reason is that the explanation of the kind "one thing – several appearances" has any sense only if consciousness (a conscious mind) is already there as the mental subject to which the thing appears one way or another. If there is (1) a conscious mind capable to experience different appearances and (2) a thing that can affect the mind (and so appear to it) in different ways, it is clear that the interaction between (1) and (2) entails and explains there being different appearances. However, you cannot explain in this way the existence of the mental subject (the conscious mind)

capable of experiencing appearances. In other words, in the explanations "one thing – several appearances", consciousness is a necessary part of *the basis of explanation (explanans)* and so it cannot be *what is explained (explanandum) on the purely physical basis*.

So, an intelligible explanation of how consciousness can be something physical is impossible in principle. The materialist is posed with the unattractive choice between two desperate alternatives: either deny/ignore the existence of consciousness, in the usual sense (eliminativism, open or disguised as functionalism) or dogmatically hold that phenomenal mental states are identical with some physical states of the brain (or with some functions realised by physical states of the brain), despite it seeming obvious that they are not, and despite these postulated identities being unexplainable.

On a closer scrutiny, these two options – either eliminativism, explicit or implicit in analytic (*a priori*) physicalism, or commitment to brute unintelligible and unexplainable identities, characteristic of *a posteriori* physicalism – turns out to be not much different. As Philip Goff (Goff, 2011) explained, the claims that phenomenal concepts are *opaque* and that their referents are in fact, *despite what they seem to be*, some physical or functional states entail that thinking of phenomenal mental states in terms of phenomenal concepts reveals nothing of what these states really are; on the contrary, phenomenal concepts hermetically conceal the real (physical or functional) nature of their referents. For example, our phenomenal concept of pain is the concept of the subjective quality of pain experiences – what it is like, how it feels when it pains; but in fact, (according to *a posteriori* physicalism) pain is nothing like what it seems to be – it is some physical of functional brain state. So, phenomenal concepts are entirely misleading – the states to which they refer are not at all like these concepts present them. So, phenomenal mental states as they are presented by phenomenal mental concepts – that is, pains, joys, colour experiences, smells, etc. how we normally think of them – do not exist; all there really is in their place are some physical brain states, or functions realised by physical brain states. This differs from eliminativism only in the letter, if at all.

7. Brute necessities, unintelligible identities and the strategy of phenomenal concepts

The arguments of the opponents of materialism, such as the knowledge argument and the zombie argument, succeed at least in two respects: first, by means of thought experiments and similar reasoning (Mary's story, thinking of a bat's sonar perception, conceiving a world of phenomenal zombies, etc.) they highlight and manifest the radical difference between, on the one hand, phenomenal mental states and, on the other hand, physical entities, properties, relations and anything that can be intelligibly constituted by them; second, they show the failure of materialistic attempts to explain how phenomenal mental states can be identical with, or constituted by, physical (nonmental) entities, properties, and relations. Moreover, they show that this failure has systematic character that makes it very

unlikely that a satisfactory materialistic explanation of phenomenal mental states (their subjective quality, how it feels, what it is like for the experiencer) can appear in future.

The anti-materialist arguments invoke and sharpen the obviousness that whatever physical structures and dynamics (and corresponding functions), their presence does not entail that there are subjective experiences and experiencers, mental subjects (such as our selves). Although some materialists tried to counter this point by invoking Kripkean *a posteriori* necessities, this move is invalid, as Kripke himself (1972: pp. 327-342), and later Chalmers (1996: pp. 38, 56-69; 2010: pp. 166-184), Kirk (2005: pp. 14-17), Levine (1983; 2001: pp. 45-49), and others explained. This leaves for materialists, as a way of escape short of eliminativism (the denial of the very existence of consciousness as the realm of subjective qualitative mental states), to postulate *a posteriori* necessities of some other, non-Kripkean kind – brute, unexplainable necessities due to brute, unexplainable identities. However, such postulation looks a desperate *ad hoc* move and refusal to admit the obvious: properties that are obviously entirely different – so much so that the claim that they are identical is ununtelligible[2] – are declared, in spite of everything, identical, without explaining how it is possible.

In the contemporary materialistic philosophy, the postulation of unexplainable brute psychophysical identities is usually not quite straightforward but implicit in the influential approach called *the strategy of phenomenal concepts*. This makes the account of psychophysical identities more sophisticated, so that these identities look less "brute", although still unexplained and unexplainable. In effect, the strategy is used to deflect the request for explanation of psychophysical identities by providing instead a sophisticated account of *why, although materialism is true, psychophysical identities are unexplainable*. It explains why phenomenal and physical properties, while being the same, seem entirely different, so much so that we cannot even understand how they can be the same.

The strategy is based on the assumption that can be described as *Inferential Isolation*: phenomenal concepts refer to *physical properties* but are "inferentially isolated" from *physical/phenomenal concepts*. That is, although phenomenal mental states are identical with some physical states (processes) of the brain or functions realised by these physical states (and so the physical entails the phenomenal), our mindbrain is organised in a specific way that inferentially isolates phenomenal concepts from physical and functional ones, and so makes it seem to us that the phenomenal and the physical are entirely different and makes us unable to see how the later can be identical with, or constitute the former.

However, the strategy does not really meet the objections against *a posteriori* physicalism with its postulations of brute identities. All it does is shunning explana-

[2] Cf.: E.J. Lowe: "I can make no clearer sense of the idea that a conscious mental state might just *be* a physical state than I can of the idea that a physical object might just *be* a natural number. ... the thesis that mental states "just are" (identical with) physical states is simply unintelligible." (Lowe, 2008: pp. 22-23)

tion of how phenomenal mental states can be some physical/functional states by offering instead some explanation of how our brains can make us think that phenomenal mental states are not physical, whereas in fact they are physical.

Consider, for example, explanations by one of the leading theorists of this approach, David Papineau. He admits that "mind-brain identity claims strike us as false":

"It seems undeniable that most people have a strong intuition of mind-brain distinctness—an intuition that pains are something extra to brain states, say. This intuition is prior to any philosophical analyses of the mind-brain relation, and indeed persists even among those (like me) who are persuaded by those analyses that dualism must be false" (Papineau, 2007: p. 135).

However, Papineau thinks that this is not "tantamount to denying physicalism", because "physicalists should allow that physicalism *seems* false, not that it *is* false"; "physicalists should maintain that we have an *intuition* of mind-brain distinctness but that this intuition is mistaken" (Papineau, 2007: p. 135).

Anticipating the charge that such a defence is *ad hoc*, Papineau denies this and explains that materialists not merely assert the falsity of the intuition of mind-brain distinctness but can "offer some explanation of why we should all have such a persistent intuition of mind-brain distinctness even though it is false". In fact, "[t]here are a number of possible ways of explaining away the intuition of distinctness, especially for physicalists who recognize phenomenal concepts" (Papineau, 2007: p. 135).

However, are such explanations sufficient to absolve materialists' assertion of what seems to be obviously false (that phenomenal mental states are identical with some physical brain states) from the charge of adhocness? I think that they are not, for the following reasons.

The followers of the strategy of phenomenal concepts claim that although mind and brain, the phenomenal and the physical, are in fact identical, we have a strong persistent – but mistaken – intuition that they are distinct. They recognise that any satisfactory physicalist position should offer some explanation of this (Papineau, 2007: p. 135). However, this requirement can be understood in two very different ways. First, a physicalist is required to offer a satisfactory explanation of two things: (1) how the intuition that mind and brain are distinct can be mistaken, that is, how phenomenal mental states can be identical with some physical states (of the brain) and (2) how can it be that we have the persistent intuition that mind and brain are distinct, on the assumption that this intuition is mistaken. Second, a physicalist is only required to offer the explanation of (2).

With the first construal, the crucial thing is (1), which means providing an intelligible explanation of how phenomenal mental states can be identical with, or constituted by, some physical states of the brain. This would amount to solving the hard problem of consciousness. However, the strategy of phenomenal concepts is devised exactly to relieve a physicalist from the burden of offering such explanation. So, the strategists conveniently replace the hard problem with (2). Instead of explaining how the entities with purely physical qualities, having nothing subjective

(phenomenal) about them, when arranged in a certain spaciotemporal order, can constitute subjective (phenomenal) mental states, the strategists propose to assume that they do so in some incomprehensible way, and then explain this incomprehensibility by some theory about the systematically misleading conceptual organisation of our mindbrains.

However, such "explanation" is a very dubious achievement, and it is *ad hoc*. Moreover, the strategy has the character of *reinforced dogmatism*[3], for one can defend in this way any claim, however obviously false, and deflect any criticism. The method is: for any statement X, however implausible and unintelligible,

(1) insist that X is true;

(2) claim that the conceptual organisation of the human mind is deficient in some way that makes it seem to us that X is false and makes it incomprehensible for us how X can be true;

(3) offer some theory of such conceptual organisation. (The tenability of the theory offered does not matter much: if one such theory fails to withstand criticism, the strategy does not really suffer – you can always modify the theory or propose another one or just hope that a satisfactory theory will be found later: surely, there is a lot of ways how the conceptual organisation of the human mindbrain can be deficient in such a convenient for X manner.)

Eventually, the defence of materialism by means of the postulation of brute psychophysical identities (usually excused by the strategy of phenomenal concepts), which is distinctive of *a posteriori* physicalism (as distinct from *a priori* physicalism, which denies the obvious in a different way – by denying that phenomenal concepts are neither physical nor functional ones), follows the simple defensive strategy that can be expressed by the principle: if a claim X seems obviously false, but you do not want to abandon X, then deny the obvious. Insist that our mind is so organised that X seems false and we cannot understand how it can be true, but, nonetheless, X is true[4]. As opposed to this, the supporters of the knowledge argument and the zombie argument are guided by *the principle of rational trust in obviousness and our capacities of judgement*:

If it seems that X, and this seems obvious, and there is no intelligible explanation of how it can be that non-X, it is reasonable to hold that X. In particular, if

[3] I use the phrase "reinforced dogmatism" following Karl Popper, in the sense: a method that makes one's system proof against any sort of criticism (Popper, 1945: p. 38; 1962: p. 327).

[4] The attitude that underlies the commitment to materialism by most contemporary analytic philosophers was very plainly stated by Daniel Dennett. In his most famous book, *Consciousness Explained*, he writes that he adopts "the apparently dogmatic rule that dualism is to be avoided *at all costs*". Why so? Because "*accepting dualism is giving up*" (Dennett, 1991, p. 37) – giving up the aspiration of explaining consciousness in the way other natural phenomena are explained in natural sciences. That is, the fundamental assumption is that everything that exists *must* be explainable in the natural-scientific way. To allow that there can be something that is not so explainable (does not fit the Procrustean bed of scientism) is absolutely inadmissible – it is an awesome defeat, surrender that should be avoided *at all costs*. Of course, if this is one's attitude, all arguments are useless.

things (or properties) *seem* entirely different (*and this seems obvious*), and there is no intelligible explanation of how they can be identical, it is reasonable to hold that they are different, not identical.

The point of such arguments as the knowledge argument is exactly that materialism is unintelligible, and it can hardly be rational to hold that a view that is unintelligible is true nonetheless. The unintelligibility of materialism was the explicit conclusion of Thomas Nagel's classical paper "What is it like to be a bat?": "physicalism is a position we cannot understand because we do not at present have any conception of how it might be true" (Nagel, 1974: p. 446). Nagel, however, was at that time unwilling to reject physicalism, so he made a reservation "at present", which suggests the hope of developing *in future* a conception that would make physicalism intelligible. However, if a position X is unintelligible *at present*, it is hardly rational *at present* to hold that X is true, relying on the promissory note that X will be perhaps somehow made intelligible *in future*. In fact, the promissory note was never paid, and Nagel eventually came to the conclusion that physicalism is false (Nagel, 2012).

A later Papineau's article (2011) provides a further material that corroborates our analysis in two respects: first, it manifests the materialistic commitment along the lines "unintelligible (unbelievable) but true"; second, it identifies the principle of the causal closure of the physical and cognitive physicalism as bedrocks of materialism.

Papineau admits as much as that materialist philosophers "don't fully believe their materialism"; although in theory they assert the identity of phenomenal mental states with some physical states of the brain, intuitively they "don't really believe the two states are identical" (Papineau, 2011: p. 12). So materialists find themselves in an awkward situation – "to find yourself continually judging at an intuitive level something that you are theoretically committed to denying". While professing materialism, "at the same time they will continue to experience an intuitive conviction that materialism is false" (Papineau, 2011: p. 14).

Despite the awkwardness of such a split of personality, Papineau thinks that holding materialism is rationally justified. Why? Because he thinks that although we (even those who profess materialism) are intuitive dualists and cannot free ourselves from intuitive dualistic convictions, the explanation of why this is so must be materialistic. If so, the truth of dualism has nothing to do with the formation of our intuitive convictions that dualism is true. And hence, the fact that we have such convictions cannot serve as evidence to the truth of dualism (Papineau, 2011: pp. 13-14).

In this argument, the decisive premise is that the (true) explanation of the psychological fact of our intuitive dualism must be materialistic, – that is, the fact is entirely due to physical factors (some neurophysiological goings-on in our brains), and not in any way dependent on the truth of dualism. This premise, in its turn, Papineau justifies by the appeal to the principle of the causal closure of the physical and the physicalist construal of beliefs. It is clear that *if* my beliefs are some physical brain states and *if* nothing non-physical does not affect these brain

states, then my intuitive convictions (beliefs) that materialism is false are fully explainable by material (physical) factors. And if so, these intuitive convictions provide no support for dualism, because I would have these convictions anyway, whether dualism is true or false.

However, both premises to which Papineau appeals, the causal closure of the physical and the cognitive physicalism, are not at all obliging for an opponent of materialism. The inefficiency of this argument is especially obvious against interactionist dualism, which explicitly rejects the causal closure tenet. Papineau admits this indirectly and confusedly, in the queer form of the statement that "dualism requires epiphenomenalism", because "[t]o suppose otherwise requires denying the causal closure of the physical, a step which few contemporary dualists are prepared to take" (Papineau, 2011: pp. 13-14). This queer formulation is deficient in two ways. First, certainly, dualism does not require epiphenomenalism – there are two branches of dualism – epiphenomenalism, which holds, together with materialism, that the physical is causally closed, and interactionism, which explicitly denies the causal closure of the physical. Second, it is just not true that few contemporary dualists are prepared to deny the causal closure of the physical. Perhaps, Papineau's claim reflects the dominant tendency in his closer academic environment[5], but generally by far most of the prominent contemporary and recent defenders of dualism – Karl Popper, John Eccles, John Beloff, John Foster, Howard Robinson, Richard Swinburne, William Hasker, James Moreland, Henry Stapp, to list some – are/were interactionists.

REFERENCES

Alter, T. (2007). Does Representationalism Undermine the Knowledge Argument? In T. Alter and S. Walter (Eds.), *Phenomenal Concepts and Phenomenal Knowledge* (pp. 65-76). New York, Oxford: Oxford University Press.

Broad, C.D. (1925). *The Mind and Its Place in Nature.* London: Routledge and Kegan Paul.

Chalmers, D. (1996). *The Conscious Mind,* New York: Oxford University Press.

Chalmers, D. (2010). The Two-Dimensional Argument Against Materialism. In Chalmers, D. *The Character of Consciousness* (pp. 141-205). New York: Oxford University Press.

Churchland, P. (1984). *Matter and Consciousness.* The MIT Press.

Dennett, D. (1991). *Consciousness Explained.* Little, Brown and Co.

Farrell, B.A. (1950). Experience. *Mind,* (59), 170-198.

Feigl, H. (1967). *The "Mental" and the "Physical": The Essay and a Postscript.* Minneapolis: University of Minnesota Press.

[5] The debates were turning mostly around the arguments and ideas advanced by philosophers who, while making the case that materialism has serious, perhaps insurmountable, problems with explanation of phenomenal mental states, still accepted (as the title of a book by one of these philosophers, (Kim, 2005), suggests) "physicalism or something near enough". That "something near enough" usually involved epiphenomenalism and cognitive physicalism. In particular, two philosophers-dualists whose arguments were most amply discussed, early Frank Jackson (1982; 1986) and early David Chalmers (1996), accepted these positions. Later, both changed their views: Jackson converted to physicalism, Chalmers became neutral between epiphenomenalism, interactionism, and Russellian monism.

Fumerton, R. (2013). Knowledge, Thought, and the Case for Dualism. Cambridge: Cambridge University Press.

Goff, P. (2011). *A Posteriori* Physicalists Get Our Phenomenal Concepts Wrong. *Australasian Journal of Philosophy,* (89(2)), 191-209.

Horgan, T. (1984). Jackson on Physical Information and Qualia. Philosophical Quarterly, (32), 127–136.

Horowitz, A., Jacobson-Horowitz, H. (2005). The Knowledge Argument and Higher-Order Properties. *Ratio,* 18(1): 48-64.

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

Jackson, F. (1986). What Mary didn't know? *Journal of Philosophy,* (83), 291-295.

Jackson, F. (2004). Looking Back on the Knowledge Argument. In P. Ludlow, Y. Nagasawa and D. Stoljar (Eds.), *There's Something about Mary* (pp. xv-xix). Cambridge, London: The MIT Press.

Kim, J. (2005). *Physicalism, or Something Near Enough*. Princeton University Press.

Kirk, R. (1974a). Sentience and Behaviour. *Mind,* (83(329)), 43-60.

Kirk, R. (1974b). Zombies v. Materialists. *Aristotelian Society Supplementary Volume*, (48(1)), 135-152.

Kirk, R. (2005). *Zombies and Consciousness*. Oxford, New York: Oxford University Press.

Kripke, S. (1972). Naming and Necessity. In D. Davidson & G. Harman (Eds.), *Semantics of Natural Language* (pp. 253-355). Dordrecht: Reidel.

Leibniz, W. (2014). The Monadology. In L. Strickland (Ed.), *Leibniz's Monadology: a new translation and guide* (pp. 14-33). Edinburgh University Press.

Levine, J. (1983). Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly,* (64), 354-361.

Levine, J. (2001). *Purple Haze: The Puzzle of Consciousness*. Oxford University Press.

Lewis, D. (1988). What Experience Teaches. *Proceedings of Russellian Society,* (13), 29-57.

Loar, B. (1997). Phenomenal States. In N. Block et al. (Eds.), *The Nature of Consciousness* (pp. 597-616). Cambridge, Mass. and London: MIT Press.

Lowe, E.J. (2008). *Personal Agency*. New York, Oxford: Oxford University Press.

Lycan, W.G. (1990). What is the Subjectivity of the Mental. Philosophical Perspectives, (4), 109–130.

Meehl, P.E. (1966). The Complete Autocerebroscopist. In P. Feyerabend and G. Maxwell (Eds.), *Mind, Matter, and Method: Essays in Philosophy and Science in Honor of Herbert Feigl* (pp. 103-180). Minneapolis: University of Minnesota Press.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, (82), 435-450.

Nagel, T. (2012). *Mind and Cosmos: Why the Materialist Neo-Darwinian Conception of Nature is Almost Certainly False*. New York, Oxford: Oxford University Press.

Nemirow, L. (2007). So this is what it's like: a defense of the ability hypothesis. In T. Alter and S. Walter (Eds.), Phenomenal Concepts and Phenomenal Knowledge. New Essays on Consciousness and Physicalism (pp. 32-51). Oxford: Oxford University Press.

Nida-Rümelin, M. (1998). On Belief About Experiences. An Epistemological Distinction Applied to the Knowledge Argument Against Physicalism. *Philosophy and Phenomenological Research*, (58 (1)), 51-73.

Nida-Rümelin, M. (2004). What Mary Couldn't Know: Belief About Phenomenal States. In P. Ludlow, Y. Nagasawa and D. Stoljar (Eds.), *There's Something about Mary* (pp. 241-267). Cambridge, London: The MIT Press.

Nida-Rümelin, M. (2007). Grasping Phenomenal Properties. In T. Alter and S. Walter (Eds.), *Phenomenal Concepts and Phenomenal Knowledge* (pp. 307-338). New York, Oxford: Oxford University Press.

Papineau, D. (2007). Phenomenal and Perceptual Concepts. In T. Alter and S. Walter (Eds.), *Phenomenal Concepts and Phenomenal Knowledge* (pp. 145-166). New York: Oxford University Press.

Papineau, D. (2011). What Exactly is the Explanatory Gap? *Philosophia*, 39, 5-19.
Popper, K. (1945). *The Open Society and Its Enemies. Volume 2. The High Tide of Prophecy: Hegel, Marx, and the Aftermath*. London: George Routledge & Sons.
Popper, K. (1962). *Conjectures and Refutations*. London and New York: Routledge & Kegan Paul.
Robinson, H. (1982). *Matter and Sense. A Critique of Contemporary Materialism*. Cambridge: Cambridge University Press.
Robinson, H. (2004). Dennett on the Knowledge Argument. In P. Ludlow, Y. Nagasawa and D. Stoljar (Eds.), *There's Something about Mary* (pp. 69-73). Cambridge, London: The MIT Press.
Robinson, H. (2016). *From the Knowledge Argument to Mental Substance*. Cambridge University Press.
Russell, B. (1927). *The Analysis of Matter*. London: Kegan Paul.
Sepetyi, D. (2017). *Consciousness as Subjectivity: The Mistery of Self. Book 1. Zombies, Computers, and Absolute Spirit*. [In Ukrainian]. Zaporizhzhia: Prosvita. [=Сепетий 2017]
Tye, M. (1986). The Subjective Qualities of Experience. Mind, (95(377)), 1-17.
Tye, M. (2009). *Consciousness Revisited. Materialism without Phenomenal Concepts*. Cambridge, London: The MIT Press.

Received 08.07.2023