Causality and the Nature of Mental States

By

Ivan Milić

Submitted to

CENTRAL EUROPEAN UNIVERSITY

Department of Philosophy

In partial fulfillment of the requirements for the degree of

Master of Arts

Supervisor: Professor Hanoch Ben-Yami

Budapest, Hungary

June 2010

Abstract

Behaviorist doctrine has been renounced with severe criticisms and followed by functionalism which turned out to be the leading theory in the philosophy of mind. In light of this fact, the aim of my essay is rather controversial. I intend to argue for conceptual behaviorism by discarding functionalist theory on the number of issues.

In the first chapter, I outline the main features of functionalism: topic-neutrality, typical causes and effects, inner mental states and the multiple realizability thesis. The anti-functionalist arguments against these features: general arguments, cognitive arguments, triviality arguments and the multiple realizability arguments are evaluated in the second chapter. By accepting the soundness of these objections, I use them as a guide in defending conceptual behaviorism in the philosophy of mind. Finally, in the third chapter, I argue for conceptual and criteria behaviorism, by arguing against rival alternatives.

'Causality and the nature of mental states' emphasizes two spotlights and points of divergence for behaviorism and functionalism. Both theories stress the importance of causality with respect to mental states, although they differ with regard to their description of causal model. The contentious issue about the nature of mental state revolves around the debate whether we should account for functionalist 'inner' mental state or not. I will suggest that the question of the nature of mental states should be answered by the conceptual analysis which embeds the manifestation of mental states in behaviorist terms.

Introduction

The outbreak of psychological behaviorism in the 1920s replaced the introspectionism with its Cartesian dualist program. The behaviorist manifesto planned to reconstruct the science of the mind by studying the behavior of organisms and its classical conditioning with auxiliary principles were hoped to play the role in behavior equivalent to the role of Newton's laws in physics.

The revolution in psychology announced the revolution in the philosophy of mind. In 1935, Carl Gustav Hempel proposed the argument for logical or analytic behaviorism. This position came to be defended by two subsequent seminal works, Gilbert Ryle's 1949 *The Concept of Mind* and Ludwig Wittgenstein's 1950 *Philosophical Investigation*. Although not admitting to be behaviorists, their insights had a profound effect for behaviorism: Ryle's idea about the ghost in machine, as an instance of category mistake and Wittgenstein's 'beetle in the box' thought experiment attempted to dismiss with occult episodes of Cartesian 'inner' processes.

Until the 1960s, philosophy of mind was largely behaviorist in nature. Beside analytic behaviorism, metaphysical behaviorism has been flourishing as well, propounded by Quine among others. However, each strand of behaviorism was countenanced by number of objections, and with the rise of functionalism, which responded to these problems, behaviorism lost on popularity. Inheriting some behaviorist ideas, functionalism characterized mental states as causally mediating subject's inputs and outputs. Yet, functionalism encountered serious difficulties of its own concerning the mental causation (Kim, 1989), its multiple realizability thesis (Kim, 1992; Shapiro, 2000), avoidance of qualia (Nagel, 1974; Jackson 1982; Block, 1990, 2006), etc.

In what follows I will evaluate these arguments, along with the general arguments, cognitive arguments, and triviality arguments. By suggesting that these objections undermine functionalist position, I will propose its replacement and advocate the return to Wittgensteinian conceptual behaviorism.

1. Functionalism in the Philosophy of Mind: The Sirens' Song

The leading conception in the philosophy of mind, as well as in psychology and cognitive science (Heil, 2000: 89), functionalism gained wider support over its rivals due to its following features:

(i) Topic-neutrality

(ii) Typical cause and effect

(iii) Inner mental state

(iv) Multiple realizability

Although functionalism is not a unified doctrine but similarly to behaviorism comes in many flavors, it can be said that features (i)–(iv) represent views of all of its strands: the machine state functionalism (Putnam, 1997), psychofunctionalism (Pylyshyn, 1984), teleological functionalism (Sober, 1999), and analytic functionalism (Armstrong, 1993; Lewis, 1966, 1972, 1991)[1].

1. 1 Topic-neutrality

An account is topic-neutral if it does not refer to the nature or 'real essence' of an object, state or process it purports to explain. For instance, if we analyze our concept of water as 'whatever structure underlies the transparent liquid drunk by humans and which falls as rain' (Robinson, 1982: 27), our analysis is topic neutral since our explanation did not invoke hydrogen and oxygen.

[1]Analytic functionalism is sometimes referred to as causal functionalism (Armstrong, 1993), causal-theoretical functionalism (Kim, 1999) and conceptual functionalism (Block, 2000).

Motivated by Occam's razor, U. T. Place (1956) and J. J. C. Smart (1959) were one of the first to propound topic-neutral analysis in the philosophy of mind. As Smart famously formulated:

> When a person says 'I see a yellowish-orange after-image' he is saying something like this: "*There is something going on which is like what is going on when* I have my eyes open, am awake, and there is an orange illuminated in good light in front of me" (Smart, 1959: 142).

In this way, mental states came to be characterized without reference to their nature. They might have been identical to brain states, silicon chips, hydraulic tubes, or immaterial substance, such as a soul.

Although Place and Smart were not functionalists themselves, topic-neutrality became part of the functionalist program as well. Eventually, functionalism was rendered compatible with (1) type-type identity, (2) token-token identity and (3) dualism, since its analysis of mental states was shown to be true if both something physical or something non-physical came to be the real essence of mental states (Robinson, 1982: 41).

Robinson reminds that topic-neutrality is 'philosophically illuminating' since one can refer to a process whose nature we don't know (Robinson, 1982: 33)[2]. Yet, the majority of functionalists do not seem to be impressed enough by this possibility since most of them tend to use functionalism as a framework to fit their materialist views (Armstrong and Lewis being the most famous proponents of this fashion). However, functionalists' motive is not solely a

[2] However, topic-neutrality seems to emerge only with respect to reference but not sense as the behavioral part seems to be present (Robinson, 1982: 33).

matter of preference of materialist metaphysics; it is also a strategy to avoid problems facing their theory why they get more inclined to it[3].

The functionalist program tended not only to avoid characterizing the nature of mental predicates, but some functionalists (notably analytic functionalists) aimed to eliminate the mental predicates *in toto*. Namely, the 'Ramseification', a method proposed by Frank Ramsey and adopted by David Lewis, purports to show how to formulate functional definitions of mental states without mentioning mental predicates. It starts with an underlying theory T of a certain mental state, e.g., belief that it is raining, whose highly simplified form would state that:

> (T) For any *x*, if *x* sees or hears the rain and is normally alert, *x* believes that it is raining. If *x* believes that it is raining, *x* will carry an umbrella.

As the underlying theory T, we used commonsense psychology in this case (which is a theory employed within analytic functionalism) although we might have used scientific psychology as well (in which case, we would pursue psychofunctionalism). After formulating (T), we proceed to introduce predicate variables in place of mental states 'being normally alert' and 'believes that it is raining'.

> (T*) There exist mental states M_1 and M_2 such that for any *x*, if *x* sees or hears the rain and *x* is in M_1, *x* goes into M_2. If *x* is in M_2, *x* will carry an umbrella.

Finally, we introduce existential quantifier and translate (T*) as [T (M_1, M_2)] to get:

> (T**) $\exists\, M_1\, M_2\, [T\, (M_1, M_2)]$

[3] Kim's solution to the qualia problem makes reference to the neural states (Kim, 1999: 114).

The outcome of Ramsey-Lewis method, a sentence (T**), does not contain psychological expressions and is intended to explain a mental state in terms of causal relation (Kim, 1998: 105-8).

1. 2 Typical cause and effect

Some authors have argued that a description of mental states in terms of behavior or dispositions to behave is bound to vary in indefinitely many ways, since our dispositions to act depend on the multitudinous circumstances (Heil, 2000: 63). For instance, if I believe that I am running late for a bus, I may decide to hurry, wait for another bus, hitch-hike, or call a taxi. Moreover, to each of these behaviors, one can add a set of other behavioral dispositions. For this reason, it is claimed, the behavioristic descriptions are necessarily 'open-ended' (Lowe, 2000: 42–3).

The 'open-ended' lists have a suspicious explanatory power. Namely, it is difficult to see how such lists are supposed to provide explanation we already don't have. Four types of behavior we mentioned with respect to the same mental state do not have anything in common except the fact that they are all ascribed to a person having a mental state in question. It appears that the only way that to find out that such a mental state will be related to these types of behavior is by virtue of knowing it already. Therefore, 'open-ended' lists are of no use in explaining mental states.

It was precisely because the open-ended lists were of no avail, I believe, that functionalists decided to 'close' them by including '*typical* causes and effect' of a mental state.

> The definitive characteristic of any (sort of) experience as such is its causal role, its syndrome of most typical causes and effects. (Lewis 1966: 17)

Alternatively, 'typical causes and effects' are referred to as the 'characteristic ways' for the input, interactions with other mental states and output (Lowe, 2000:45).

Two lines of thought have been proposed in order to introduce the typical causal links: one was to say that mental states are *de facto* typical, by alluding to many simple mental states, such as being in pain or seeing red. The other way to achieve this was to group the open-ended lists of dispositions into coherent sets of inputs, associated mental states and outputs, which would be typical[4].

Arguably, the open-endedness was not avoided in any of these cases. With respect to the first case, as I will discuss in section 2.1, it can be proven that there are some mental states which do not have typical causal links. With regard to the other case, I will try to show that the number of coherent sets of causal links, to which functionalists adhered, is nonetheless open-ended. I will argue for this claim in section 2.2.

1. 3 Inner mental state

As we have seen in section 1.1, the belief that it is raining was characterized in terms of input and output. The behaviorist analysis, carried out entirely in these terms, was considered to be inadequate by functionalists. Their discontent rested on the objections which can be illustrated as follows.

(1) Subsequent to playing any three notes on a piano to a person with absolute pitch, you are amazed by the correctness of her *report* about the notes you played. What you

[4] Such an attempt will be discussed in more detail in the section 2.1

come to believe is that her extraordinary talent cannot be reduced to what she *says*. According to the behaviorist, however, there is nothing over and above her talent than such behavioral dispositions. For this reason, behaviorist explanation appears to be 'disappointingly slight' (Armstrong, 1993:64).

(2) Approaching the front door, you hear a person talking quietly. Since her voice is low, you assume that she *believes* that other people fell asleep. But your assumption, although correct, is incomplete because if her belief caused her to speak in a low tone, it is likely that her inner mental states also consisted of her *thinking* that a low tone will not wake up anybody, her *wish* not to wake up anybody etc. (Lowe, 2000: 46)

To circumvent the difficulties described in (1) and (2), functionalists decided to introduce internal states which would be correlated with both input and output of mental states (Kim, 1998: 112). An argument against the functionalist's internal state will be proposed in section 2. 2.

1. 4 Multiple realizability

The multiple realizability thesis was introduced by Hilary Putnam in the 1960's and largely used as part of the argument against the identity theory (Bickle, 2006)[5]. Putnam's idea was that 'the *same* Turing Machine may be physically realized in a potential infinity of ways' (Putnam, 1997b: 415). Supposedly, its building blocks can vary *ad libitum*: the automata in question might be built out of flip-flops, relays or vacuum-tubes (Putnam, 1997b: 414) while its psychological predicates will nevertheless remain the same. This was taken to be true for

[5] However, such an argumentation is not relevant to my undertaking here. I will first present multiple realizability as an important feature of functionalism, whereas its criticism will be taken up in the second chapter.

both humans and animals. For instance, the physical-chemical "correlate" of a given mental state might be different in case of mammal and an octopus, but they might share the same psychological predicate nevertheless (Putnam, 1999: 32). Consequently, it has become a part of common wisdom that mental states are multiple realizable by various physical systems, be it brain, hydraulic tubes or even immaterial substance, such as a soul. The common analogy with two computers, which can run the same software albeit being built of different hardware, makes the whole account more persuasive (Lowe, 2000: 48).

Since the identity theory proposed characterization of mental states in terms which appeared to be an exclusively human trademark, not admitting the variety of physical and non-physical realizers of mental states, it was labeled 'chauvinism'; by the same token, the multiple realizability thesis was decisive in making functionalism anti-chauvinist. Nonetheless, two types of counterarguments emerged.

First, some authors have shown that the anti-chauvinism leaves the possibility of trivialization. Roughly, given the functionalist picture, one is obliged to espouse *liberalism*, i.e., the view that things which initially would not be recognized as having mental states would turn out to have them. This type of argument will be discussed in section 2.3.

Second, the multiple realizability thesis (MRT) itself can be shown to be trivial. MRT can be formulated as:

> (MRT) Two tokens are different realizations of the same kind if and only if they differ in causally relevant properties.

As I will argue later, 'causally relevant properties' of a certain kind are those properties which contribute to factors which individuate that kind. For instance, a causally relevant property for a knife is its sharpness, whereas its color is not such a property. Thus, if two knives differed only in their color, we would not be inclined to admit that they are

different realizations of this kind. I will point out that the role of color in the example with knife is analogous to the role of physical realizer with respect to the mental states. This argument will be discussed in detail in section 2. 4.

2. The Criticism of Functionalism: Wreck of the Ship

Despite the initial plausibility of functionalism, philosophers of mind have proposed various types of arguments against it over the past few decades.

(*i*) Qualia arguments

According to this group of arguments, sensory mental states such as 'seeing red' or 'feeling pain' have a certain sensational content, a specific 'what it is like' which cannot be captured by a causal characterization of functionalist type. The best known qualia arguments are instances of the knowledge argument pursued by Nagel (1974) and Jackson (1982) as well as inverted qualia arguments (Block, 1990) and 'absent' qualia arguments (Chalmers, 2002) and (Block, 2006).

(*ii*) Holistic arguments

Holistic arguments emphasize the fact that a functionalist characterization of mental state is holistic. In order to explain a mental state, the functionalist needs to describe it in terms of the role it plays in a psychological theory, along with the number of other mental states that enters the picture. Arguably, the context of description is too large to allow functionalism to describe mental states which would be common to all the humans, let alone all the animals. In other words, when the functionalist defines a mental state, e.g., a certain belief by reference to some desires and knowledge, he runs into problems that other people who do not have these desires and knowledge will lack a mental state in question. It goes without saying that the same holds for the animals (Putnam, 2001).

(*iii*) Cognitive arguments

Cognitive arguments against functionalism are arguments which attempt to criticize functionalist account of cognitive states such as a belief or understanding, rather than to focus on sensory states, such as being in pain or being thirsty. We will discuss cognitive arguments in section 2.2; in particular, we will examine Ben-Yami's argument (1999), Zhu's reply (2006) and three cognitive arguments against Zhu's objections.

(*iv*) Triviality arguments

Triviality arguments contend that the functionalist's claim that a complex physical system demonstrates a given functional organization is either trivial or is not as informative as we usually take it to be. In section 2. 3 we will look closely to Godfrey-Smith's triviality argument (2009), as well as Hinckfuss' and Searle's objections.

(*v*) Multiple realizability arguments

This sort of arguments undermines the multiple realizability thesis. One of the most interesting proposals against the multiple realizability thesis was defended by Shapiro which we be discussed in section 2. 4.

2. 1 General Arguments

Arguments against functionalism can be broadly divided into two groups: general and specific arguments.

The five types of arguments I outlined in the previous sections are specific arguments which target some more specific claims about the multiple realizability thesis, sensory and cognitive mental states etc. General arguments, on the other side, attack more general features

of functionalism such as the role of causality, the insistence on typical conditions in a causal model etc. In what follows, we will present one general argument and argue for its plausibility.

2. 1. 1 The Hallucination Argument

One can hardly escape the impression that counterarguments against functionalism are often based on imagined scenarios: zombies, weather watchers, or people with inverted spectrum. I believe that problems with functionalism emerge on a more basic level.

Functionalists attempted to overcome difficulties of behaviorist open-endedness by positing '*typical* causes and effects' or more precisely, typical physical stimuli, typical ways in which mental states interact with other mental states and typical behavioral outputs. In what follows, I intend to show that there are cases which resist such characterization.

When functionalism is introduced, its main ideas are applied to rather simple mental states. Usually, functionalism is tested on feeling of pain or seeing red (Lowe, 2000: 45) with respect to sensory states, or on the simple beliefs such as the belief that it is raining, with regard to cognitive ones. These simple mental states are the basis of the promise that functionalism can cash out *every* mental state in the proposed way. In addition, the functionalist promises to establish his views by employing the '*typical* causes and effect' clause. Although the notion of the typical can be put to objection of being a notion with vague boundaries, this is not a criticism I would like to develop here. Rather, I wish to investigate whether there are *typical* causes and effects with regard to *all* mental states. Naturally, if functionalism is to be a coherent theory it would have to hold such a view.

In this section, I will argue that the mental state of hallucination resists a characterization in terms of typical causal links. Hallucinations can be broadly divided into two types[6]. One type concerns the hallucinations of which one is *aware* as such when having them. For instance, if I hallucinate that I see a blue elephant in my room I will be aware that I hallucinate it. On the other hand, there are hallucinations of which we are not aware as such *when* having them. For example, the hallucination of the everyday object will not make us think that it is not in our visual or auditory field, since we would not have reason to suppose that we do not perceive it in the first place.

I will argue against the functionalist account of the latter type of hallucinations, namely the hallucinations which we do not recognize as such. In arguing against functionalism, I will deal with three following questions:

(*i*) Do hallucinations have a typical *cause*?

(*ii*) Must there be a typical *interaction* with other mental states?

(*iii*) Must there be a typical *effect*?

With respect to (*ii*) and (*iii*) I will show that there *are* typical interactions with other mental states and typical behavioral outputs, but that these typical interactions are not any different from those of veridical perception. My argument will rely on the claim that the notion of typical implies the notion of uniqueness. Thus, if certain features are typical of hallucinations they are supposed to belong only to hallucinations, and not to veridical perceptions. To insist that two different mental states can share all their typical causes and effects would be to misconceive the meaning of the word typical.

[6] For the sake of simplicity, let us omit the cases in which the hallucinations are caused by drugs.

With respect to (*i*), my line of argumentation will be different. I will show that the hallucination has no typical cause whatsoever. If I am right, this will yield two problems for functionalism. First, the functionalist account will be threatened by an obvious incoherence since it contends that that there are 'typical causes and effects' for all mental states.

One line of response would be to require the possibility that *certain* mental states such as this kind of hallucination have no typical causes[7]. On this view, the hallucinations would be characterized as being different from veridical perceptions since veridical perceptions *have* typical inputs. Since the functionalist only needs to show that two different types of mental states can be distinguished on the basis of its typical inputs, the interactions with other mental states and outputs, the problem for functionalism disappears.

But even if allow for this, functionalism will face the new difficulty. Namely, functionalists have introduced typical causes and effects with intention to 'close' the open-ended lists, which were meant to account for mental states. Hence, acknowledging that there are some mental states without typical causes and effects would bring functionalism open-ended lists. This is highly unacceptable result for functionalists since it would collapse into behaviorism.

(*i*) Do hallucinations have a typical *cause*?

As I have suggested, the typical cause exists in the related case of veridical perception, but not with regard to hallucination. For instance, whenever I *perceive* my hand, it is in front of me, but the very fact that I *hallucinate* my hand means that it is *not* in my visual field at the moment I hallucinate. Even under the supposition that there are cases when people always hallucinate *x* when seeing *y* (as its typical cause), this is true of at most few cases of

[7] Such a claim is perfectly compatible with the behaviorist point of view.

hallucinations and not all of them. Thus, I conclude that at least some hallucinations have no typical cause.

(*ii*) Must there be a typical *interaction* with other mental states?

Imagine that on one occasion you have a veridical perception of a spoon, whereas on another occasion you are merely hallucinating it[8]. Other things being equal, the difference between the two cases with regard to the second link in the causal chain is arguably *nothing*. The mental states which will be triggered when you perceive a spoon will be the same as when you hallucinate it, since you would still believe that you perceive the spoon. In other words, when hallucinating a spoon nothing will suggest that it is not the case of veridical perception. Therefore, the link (*ii*) does not help us to distinguish between these two mental states.

(*iii*) Must there be a typical *effect*?

In the case of the effect or behavioral output of hallucinations, one can argue that there is no unique typical effect either. For example, the behavioral output cannot be distinguished in the case of veridical perception and of hallucination which we do not recognize as such. Whether we perceive a spoon or only hallucinate it, we could equally reach out for it, believing it to be in front of us. Also, if asked to report what we see before our eyes, two reports would not differ either. My conclusion, therefore, is that 'typical causes and effects' clause cannot be applied to the case of hallucination for two reasons.

First, there is no unique (*ii*) *typical* interaction with other mental states and there is no unique (*iii*) effect of hallucination, which would distinguish this mental state from the veridical perception. Yet, this is not a problem for the account of veridical perception since

[8] It is important to note that there is nothing extraordinary about the object of this hallucination. It has the properties we usually assign to it.

veridical perception *has* (*i*) a typical input. Notwithstanding, it *is* the problem for the hallucination since it does not have (*i*) a typical cause which forces functionalism to adopt the open-ended lists of causes, which makes it to collapse into behaviorism.

2. 2 Cognitive Arguments

"Cognitive arguments" against functionalism in the philosophy of mind are arguments which attempt to criticize functionalist analysis of cognitive states such as belief or understanding, rather than the functionalist explanation of sensory states, such as being in pain or being thirsty. In what follows, I will discuss Ben-Yami's cognitive argument against analytic functionalism.

Before developing Ben-Yami's argument, I propose the classification principle, its two corollaries and a definition of token-similarity (section 2.2.1). The plausibility of these claims is highly important as I intend to develop three arguments which show that Zhu violates these claims. In the following section, I elaborate Zhu's first objection. It contends that Ben-Yami's argument is based on an incomplete picture of functionalism, and indicates that it leaves functionalism unscathed as soon as we offer the complete functionalist description of mental states. I point out that Zhu's first objection is false by mounting the completeness argument, which shows that since completeness is a context-relative notion it is possible to have a context in which description of a mental state would be complete but still lacking elements from functionalist classification. Zhu's objection, I suggest further, conflicts with both the classification principle and a definition of token-similarity (Section 2.2.2). In the following section, I develop Zhu's second objection according to which epistemic condition of a belief yields uniform description of that mental state. I refute this objection by claiming that Zhu misconceives the nature of belief and I propose a counterexample in which two tokens of the

same belief turn out to be utterly different with regard to their epistemic condition (Section 2.2.3). In the following section, I summarize Zhu's third objection according to which any two tokens of the same mental type necessarily share their actual and counterfactual causal links. I disprove this objection by proposing the inflation argument which shows that Zhu's claim depends on his restriction of the relations between actual and counterfactual causal links to those of contrariness and contradictoriness. By showing that the relations outside Aristotle's square of opposition are legitimate as well, I dismantle Zhu's final objection (Section 2.2.4). In conclusion, I sum up the results.

2. 2. 1 Ben-Yami's Argument and the Classification Principle

A convincing cognitive argument against the functionalist account of belief was developed by Hanoch Ben-Yami (Ben-Yami, 1999). Recently, this argument has been challenged by Jing Zhu (Zhu, 2006) who set out to show that Ben-Yami's argument is unfounded. In the following four sections I intend to reconstruct Ben-Yami's argument, elaborate Zhu's criticism and develop three cognitive arguments against analytic functionalism.

In this section, prior to the presenting of Ben-Yami's argument, I will propose the classification principle, its two corollaries and a definition of token-similarity. I intend to argue for the plausibility of these claims in order to develop three arguments which show that Zhu violates these claims.

The reason I aim to formulate the classification principle, rather than some other principle, derives from the fact that analytic functionalism, Ben-Yami's argument and Zhu's objections all discuss the classification of mental states. Since classification of mental states within functionalism makes use of their explanatory features, our classification principle will

need to articulate the relation between explanatory features and the classification of tokens under the types. The classification principle goes as follows:

(P) A token is classified under a type if and only if its explanatory features coincide with the explanatory features of the type.

Two corollaries of (P) we will be using in the paper are the following:

(1) If tokens *A* and *B* share *all* explanatory features, then they are tokens of the same type.

(2) If tokens *A* and *B* share *no* explanatory features, then they are tokens of different types.

Having proposed the classification principle and its two corollaries, let us present a definition of token-similarity. Since similarity is not a case-independent notion (Godfrey-Smith, 2009: 290) we will have to propose its context-dependent definition. Echoing Geach's treatment of relative identity (Geach, 1967: 3), let us say that the sentence '*x* is similar to *y*' is short for '*x* has the similar *F* with *y*' where *F* stands for an *explanatory feature* (to be explained shortly). Accordingly, we can formulate a definition of token-similarity in the following way:

(S) If tokens *A* and *B* share more explanatory features than *B* and *C*, then *A* and *B* are more similar than *B* and *C*.

By way of example, let us see what this notion amounts to. In color theory, for instance, explanatory features are hue, saturation and value of colors. These explanatory features are used to classify colors – e.g., a token of color will be classified as light blue if and only if its explanatory features coincide with explanatory features of the type light blue (these being 240^{o} for hue, 90% for saturation and 80% for value). With respect to corollaries (1) and (2) if two tokens share all three explanatory features, they will be classified as the same color. Conversely, if they do not share any such feature, they will be classified as distinct colors.

With regard to our definition of token-similarity, it follows that the more explanatory features a color *A* has in common with a color *B*, the more similarly *A* and *B* will get classified. Also, the less explanatory features *A* and *B* have, the less similarly will they be classified. Both the classification principle and a definition of token-similarity seem rather intuitive and unproblematic.

As in color theory, analytic functionalism provides three explanatory features as well. Its explanatory features are (*i*) stimulus input plus other mental states, (*ii*) interaction with other mental states[9], and (*iii*) behavioral output. It is by virtue of these explanatory features that functionalists classify mental states. As Lewis famously characterized it, a mental state is:

> [A] state apt for being caused in certain ways by stimuli plus other mental states and apt for combining with certain other mental states to jointly cause certain behavior. (Lewis, 1991: 230)

Just like the color theory, analytic functionalism seems compatible with the classification principle. For instance, any two tokens of experiencing pain will have in common three explanatory features: (*i*) tissue damage, (*ii*) a desire to relieve the pain and (*iii*) wincing or crying. However, although one could provide many examples which show that analytic functionalism is compatible with classification principle, its two corollaries and a definition of token-similarity, we will try to locate a conflict between analytic functionalism and these claims in the following sections. This will yield the conclusion that analytic functionalism is untenable position and that Zhu's objections are insufficient for its defense.

In his paper 'An Argument Against Functionalism', Ben-Yami argued against the classification model of analytic functionalism. He observed that some cognitive states, such as belief, prove this classification model to be insufficient. Namely, he showed that there are

[9] It should be noticed that (i) contains mental states as causes whereas (ii) deals with mental states as effects. (cf. Kim, 1998: 104)

cases in which two tokens would not share explanatory features (*i*)-(*iii*) but would nevertheless be regarded as tokens of the same mental type. If this was the case, it would follow that analytic functionalism is incompatible with corollary (2) which states that: 'If tokens *A* and *B* share *no* explanatory features, then they are tokens of different types'. Having seen that corollary (2) is highly plausible, this would imply that analytic functionalism is invalid.

To illustrate this point, Ben-Yami proposes the following example. Let *A* and *B* be tokens of the belief that it is raining. The belief *A* can be explained in terms of three explanatory features: (*i*) the sight of the rain (*ii*) the desire to smell the rain and the belief that the awning is fitting, which leads us to (*iii*) open the window. On the other hand, *B*'s causal network may consist of different explanatory features: (*i*) the sound of the rain, (*ii*) the desire to keep the water out and the belief that the awning is broken, which leads us to (*iii*) close the window. Naturally, Ben-Yami raises the following question:

> Why are two states [*A* and *B*], which were caused by very different causes and in their turn caused very different mental states and behaviours, the same type of belief? (Ben-Yami, 1999: 320)

Ben-Yami's question suggests that analytic functionalism cannot accurately classify tokens of the same mental state under the same type and indicates that functionalist causal network is an unacceptable model of classification.

2. 2. 2 The completeness argument

In this section, I will analyze Zhu's response to Ben-Yami's argument, and propose the completeness argument which, I believe, conclusively shows that Zhu's defense is insufficient and that analytic functionalism is incompatible with the classification principle and a definition of token-similarity.

In his paper 'In Defense of Functionalism', Zhu attempts to show that Ben-Yami's argument is invalid. His first objection begins with a claim that Ben-Yami's account of the first causal link, namely, physical stimuli plus other mental states, is *incomplete*. Both the sight of rain in the case of *A* and the sound of rain in the case of *B* are "only *part* of the causes that generate the belief as the effect" (Zhu, 2006: 96). Such a partial description must be completed by adding 'certain other mental states which embody the agent's knowledge about raining, to cause the belief that it is raining' (Zhu, 2006: 96). Zhu's proposal is justified, as he relies on Lewis repeated claims on this subject (Lewis, 250: 1972 and Lewis, 1991: 230) although he never mentions which 'certain other mental states' he has in mind. Nevertheless, we may think of some: the *memory* of the sight/sound of rain, the *capability* to compare these memories with the current rainfall etc. The exhaustive list of such features would provide us the complete description of functionalist causal link (*i*). Once we arrive at such a complete description, Zhu insists, *A* and *B* will have the common causal links instead of different ones which Ben-Yami pointed out. In this way, the functionalist type of classification will be safe from Ben-Yami's argument, which was persuasive only because its description of functionalism was incomplete.

On the face of it, Zhu's objection is legitimate since Ben-Yami did not take into consideration Lewisian 'certain other mental states' when describing the first causal link, but only the physical stimuli. However, Zhu's objection is questionable because of his other

claim, namely, that given the *complete* functionalist description of *any* two tokens of the same mental type, they will turn out to be common or uniform. His inference, that completeness implies uniformity, presupposes that the notion of completeness is case-independent and applicable to a description of any token whatsoever. But, as I intend to show, the notion of completeness is context-dependent which is why Zhu's inference does not hold.

To illustrate the point that the notion of completeness does not imply uniformity of mental states, suppose that two persons, of which one is congenitally blind and the other congenitally deaf, have the belief that it is raining. According to Zhu, the *complete* functionalist description of their beliefs would need to include very similar explanatory features. More precisely, it would follow that the congenitally blind and the congenitally deaf person would have almost uniform *first* causal link, since Zhu is here interested primarily in it (Zhu, 2006: 96). But this is far from being an accurate view. In fact, congenitally blind and congenitally deaf persons would not share any causal link[10]: the first person would rely on the sound of the rain, memory of that sound and the comparison of previous sounds with the current one, whereas the second person would depend on the sight of the rain, memory of that sight and the comparison of previous sights with the current one. Therefore, they would both have a complete functionalist description of the first link in the causal chain, although they would not be uniform, since they do not have anything in common. Therefore, the completeness does not imply the uniformity.

According to the classification principle, or more precisely, its corollary (2), if tokens *A* and *B* share *no* explanatory features, then they are tokens of different types. As we have seen in the previous example, the two tokens of belief shared no explanatory feature albeit they were regarded as the tokens of the *same* type. Therefore, the functionalist model of

[10] Of course, a congenitally blind and congenitally deaf person would share *some* causal links: knowledge about rain, tactile information about it etc. Nonetheless, these common features are insufficient to classify these two mental states as being identical.

classification is wrong. And although Ben-Yami's example was not appropriate because of his incomplete description, his main idea that functionalist model of classification is insufficient is valid.

In the vein of the previous objection, Zhu attacks Ben-Yami's account of the causal links (*ii*) and (*iii*). He insists that Ben-Yami's account of these causal links is incomplete as well.

> Likewise, the belief that it is raining must work with certain other mental states to jointly cause other mental states and behaviour. If I did not have the *desire* to keep water out, I would not close the window even though I believe it is raining; I open the window not only because I *believe* it is raining, but also because I *know* the awning will keep the water out, and I *want* to smell the rain. If I believe it is raining I may come to think that the farmers will complain, because I also *know* there has already been too much rain this year and too much rain does no good for agriculture; or I might come to think that they will stop complaining because I *know* there hasn't been enough rain yet and insufficient rain also does no good for agriculture (Zhu, 2006: 96–7).

Zhu's objection is followed by the conclusion that "normally, a certain type of mental state is seldom the effect of a single cause, nor does it in turn sufficiently cause other mental states or behaviour" (Zhu, 2006: 97). In the vein of the previous attempt to defend functionalism, Zhu restates that as soon as the complete account of the nature of causal links is given, Ben-Yami's argument poses no threat to functionalism. But, just like Zhu's previous objection, this one also seems to takes for granted that the notion of completeness is case-independent and that any complete description implies the uniform description. Unlike with the first objection, where we showed that analytic functionalism violates corollary (2), let us

now present an example in which this strand of functionalism conflicts with the definition of token-similarity.

Such an example would have to propose a complete description of mental states *A*, *B*, and *C* such that a token *A* has less in common with a token *B* of the *same* mental type then with a token *C* of *different* mental type. For instance, let us say that *A* and *B* are the tokens of the belief that it is raining, whereas *C* is a token of the belief that it is snowing. In order for functionalism to be compatible with a definition of token-similarity, no token should ever be more similar to a token of different mental type than to a token of the same mental type. But we can coherently set up the examples so that *A* is described by (*ii*) a *desire* that rain enters the house, and having *known* that the awning will not keep the water out, you (*iii*) *open* the window. Further, we can explain *B* by (*ii*) the *desire* to keep the water out, and because one *knows* that the awning will not keep the water out, one (*iii*) *closes* the window. Since *A* and *B* are utterly different it is evident that we can set up the conditions so that *C* could have more in common with *A* than *A* has in common with *B*. Yet, according to a definition of token-similarity, this should not happen, because *A* and *B*, as tokens of the same type, must be more similar than *A* and *C*, the tokens of a different type. Since functionalism violates a definition of token-similarity, I conclude that it is invalid.

We have shown that analytic functionalism conflicts with corollary (2) since it aims to classify *A* and *B* under the same type, although they do not share explanatory features. Also, we demonstrated that analytic functionalism violates the definition of token-similarity, since tokens of different types, *A* and *C*, share more features in common than tokens of the same type, *A* and *B*. Contrary to what is usually thought of functionalism in general, it seems that the functionalist analysis of belief is open-ended (Heil, 2000: 63), i.e., that a range of characterizations we can ascribe to a person having a certain belief can vary *ad libitum*. For this reason, we can generate as many counterexamples to analytic functionalism as we want

by ‘carving nature by its joints’, viz. by choosing precisely those causal links so that functionalism conflicts with either the classification principle or a definition of token-similarity.

2. 2. 3 The epistemic argument

Zhu’s second objection consists of two parts. First, he insists that in order to have a belief it is necessary to have knowledge of at least some other mental states. Second, he implies that for all the tokens of a given belief, a certain epistemic condition will be common. Had this not been the case, it would follow that the functionalist description could classify two tokens of the same type of belief under two different types, since they would differ in terms of their first causal link which includes epistemic condition.

In contrast with his previous objection, Zhu’s second objection is more specific since it cannot be applied to all mental states. For instance, it cannot be applied to pain because one does not need to have *any* epistemic condition in order to be in pain. Notwithstanding, the epistemic condition is necessary in case of the cognitive mental states.

The first part of Zhu’s objection is highly intuitive. In order to have a belief that it is raining, Zhu claims, we need to have a certain epistemic condition:

> If the agent had never known what raining is like […] The stimuli must work with certain other mental states, which embody the agent’s knowledge about raining, to cause the belief that it is raining. (Zhu, 2006: 96)

Therefore, Zhu implies that we cannot have the belief that it is raining unless we have the knowledge of ‘certain other mental states’ from the first link in the causal chain (e.g., sight

of rain, sound of rain, memory of the rain, information about the rain etc.) Zhu's remark certainly follows the functionalist claim that our beliefs must rest on *some* concepts (Armstrong, 1993: 339) and it is without any doubt plausible.

However, the second part of Zhu's objection is not so readily acceptable. Zhu contends that the epistemic condition we acquire need to be common to all the tokens of that belief. But it is quite possible to conceive of two tokens of the same belief which would be utterly different with regard to such knowledge. In this case, analytic functionalism would violate the corollary (2).

To show this, let us propose an example in analogy with Frank Jackson's thought experiment about Mary's room (Jackson, 1982). Let us imagine Mary who has been confined to her room from her birth, having the information about rain only from books but never having seen or heard rain. Accordingly, Mary will lack a memory of rain, a capability to compare the current rainfall with the previous ones and everything which Zhu might list within the first causal link. Suppose now that each hour Mary receives an extensive report about her town, whereby one of the reports for the *first* time in her life reveals the information about rain, e.g., "It is raining at 10:40 AM". We can agree that Mary does have a *belief* that it is raining. Now, let us introduce Mary's counterpart $Mary_2$ who is identical with her in everything except for the fact that $Mary_2$ never received such a report because she left her room at the time she was supposed to receive it. If we teach her that what she sees and hears at 10:40 AM is rain, it would follow that the epistemic condition of $Mary_2$ is utterly different from the epistemic condition of Mary. Since Zhu argued that the epistemic condition will be common to all the tokens of beliefs, i.e., both the beliefs of Mary and $Mary_2$, his suggestion was flawed.

2. 2. 4 The inflation argument

In the previous two sections, we have examined Ben-Yami's cognitive argument, the completeness argument, the epistemic argument and Zhu's two objections. Each of these arguments and objections was restricted to *actual* causal links of mental states. However, since analytic functionalism characterizes mental states in terms of both actual and *counterfactual* causal links, one could argue that we did not undermine analytic functionalism as long as our criticism was not proved to be valid for *both* actual and counterfactual causal links.

This reply was anticipated by Ben-Yami who offered a counterfactual argument later to be criticized by Zhu. In this section, I will first analyze Ben-Yami's counterfactual argument, present Zhu's objection and then propose the inflation argument which threatens Zhu's defense of analytic functionalism.

Having shown that *A* and *B* are not the same tokens in terms of actual causal links, what was left to show for Ben-Yami was that *A* and *B* are not the same tokens in terms of counterfactual causal links either. His counterfactual argument goes as follows:

> Suppose A and B are two tokens of the same type of belief. According to [the functionalist hypothesis], A and B have the same set of actual plus counterfactual causal links. To illustrate the problem with the view, let us consider the particular counterfactual situation in which A would change into a state C that has the same actual causal links that B actually has. According to our hypothesis, C's actual causal links would be different from those of A. Why, then, should we say that C would be the same token mental state as A, and not that A would change into a different mental state? […] It is surely possible to change one mental state into another by changing its causal links: why shouldn't that be the case with A and C? (Ben-Yami, 1999: 321)

However, Zhu's third objection was intended to reject Ben-Yami's counterfactual argument as being only apparently persuasive. In formulating his third objection, Zhu relied on the following counter-example:

> Let us suppose that A is the belief that it is raining which has an actual causal link to the behaviour of closing the window. In a counterfactual situation that the agent had a desire to smell the rain, A would change into C, the belief that it is raining which has an actual causal link to the behaviour of opening the window. Likewise, for C, in a counterfactual situation that the agent had a desire to keep the water out, it would change into A, the belief that it is raining which has an actual causal link to the behaviour of closing the window. So A and C share the same set of actual plus counterfactual causal links. Therefore, according to the hypothesis, A and C are two tokens of the same type of belief (Zhu, 2006: 98).

In order to show that A and C are two tokens of the same type of belief, Zhu employed actual *plus* counterfactual causal links. If the *actual* causal links of A were, for instance, a desire to keep the water out and close the window, its *counterfactual* causal links might be a desire to smell the rain and open the window. Now, suppose that these two counterfactual causal links of A were the actual causal links of C. In this case, the counterfactual causal links of C might be a desire to keep the water out and close the window. Zhu arrives at the possibility that the counterfactual causal links of C would be the actual causal links of A and counterfactual causal links of A would be actual causal links of C. In other words, A would become C and *vice versa*, because A and C would 'share the same set of actual plus counterfactual causal links' (Zhu, 2006: 98).

Zhu implicitly presupposes the false claim that the relation between actual and counterfactual causal links should be restricted so that counterfactual causal link is necessarily either *contrary* or *contradictory* to actual causal link. I conclude this on the basis that *all* four causal links mentioned in his example stand in these relations: the first pair, opening of the window and closing of the window stand in the relation of contradictoriness (they cannot be both true); the second pair, i.e., keeping the water out (by having the window closed) and smelling the rain (by having it open) stand in the relation of contrariness (they can be both false)[11]. Moreover, if causal links would not stand in these relations, I will argue, A and C would not 'share the same set of actual plus counterfactual causal links' (Zhu, 2006: 98).

Before proceeding to that, it should be noticed that Zhu's argument can be strengthened. It is enough to maintain that the set of all causal links of A is coherent and that each of these causal links has its contrary or contradictory link in the causal links of C. Once this is satisfied, A will counterfactually become C and *vice versa*, but instead of having only two causal links in common, as in Zhu's example, A and C would share as many causal links as one can consistently list. However, Zhu's implicit hypothesis is untenable – there is no reason to restrict the relation between actual and counterfactual causal link to that of contrariness and contradictoriness. Instead, they can be merely different. Once we allow for the relation of mere difference, A and C will not share 'their actual plus counterfactual causal links' since A will not become C and C will not become A.

In order to show this, let us first propose an example in which the relation between actual and counterfactual causal links will not be either contrary or contradictory but merely different. Let A and C be the tokens of the belief that it is raining. Now suppose that A's *actual* causal links contain a desire to go for a walk in the rain which leads to a behavior of walking in the rain. According to Zhu's implicit proposal, A's *counterfactual* causal links will

[11] Since the notions of contrariness and contradictoriness are usually applied to propositions, they would need to be extended so as to include actions.

be either *contrary* of „walking in the rain', for instance 'lying in a bed' (they can be both false) or *contradictory* of 'walking in the rain', namely 'not-walking in the rain' (they cannot be both true). But does the relation between actual and counterfactual links *need* to be restricted to these? What prevents us from introducing indefinitely many counterfactuals which are neither contrary nor contradictory, but merely different: e.g., 'playing a piano', 'watching a movie', 'reading a book' etc.? Once we introduce one of merely different counterfactual links, A will not become C and C will not become A.

As we have said, C is the token of the same belief that it is raining. According to Zhu's implicit proposal, C's actual causal links will be A's counterfactual links, whereas C's counterfactual causal links will be A's actual links. Thus, C's actual causal link will be *one* of indefinitely many links, e.g., 'playing a piano', 'watching a movie', 'reading a book' etc. Once we choose one of them, for instance, 'playing a piano', it would follow that its counterfactual causal link will *have to be* going for a walk, but this does not follow and is clearly *ad hoc*. Therefore, A and C will be easily distinguished, not sharing „their actual plus counterfactual causal links'.

In conclusion, I have argued that all three of Zhu's objections were insufficient to rebut Ben-Yami's argument and to defend analytic functionalism against the criticism, by proposing an argument for each of his claims: the completeness argument, the epistemic argument and the inflation argument. These arguments have shown that analytic functionalism violates uncontroversial classification principle, its corollary and a definition of token-similarity.

2. 3 Triviality Arguments

Functionalists contend that the complex systems, such as humans or animals, exhibit functional organization responsible for their mental properties. Triviality arguments threaten this claim regarding it as being either trivial or having much less content than we are ready to admit. The first argument of this kind, 'The Hinckfuss pail', suggested that by way of suitable categorization of states a bucket of water sitting in the sun can realize the functional organization of humans (Godfrey-Smith, 2009: 274).

Hinckfuss suggested that since the Sun would cause the convection currents and a variety of complex movements of water molecules, it would make a pail of water a sufficiently complex system to 'realize a human program for a brief period' (Lycan, 1981: 39). The way to implement this realization was imagined to proceed by way of 'correlations between certain micro-events and the requisite input-, output-, and state-symbols of the program'. In the long run, it would lead functionalism into panpsychism since what was true of a pail of water can be shown to be true of any other complex physical system (Lycan, 1981: 39).

Triviality arguments need not be restricted to functionalism, and as Searle proposed, they can be addressed to theories logically independent from it, such as computationalism (Piccinini, 2009). On Searle's view, the computationalism allowed that anything can be described as a digital computer under a certain description. That is to say, computationalism undervalued the problem we started with – how the brain works. Indeed, the answer that brains are digital computers is a trivial one if we claim that a can of beer, the city of Budapest or the constellation of Cassiopeia are all digital computers. For *this* reason, Searle maintained that any program and any sufficiently complex object are such that under a convenient description the latter will implement the former. For instance, a wall can implement the

Microsoft Word program since its molecules might exhibit the pattern isomorphic with the structure of the program (Searle: 1990)[12].

Be that as it may, we will focus on the trivialization of functionalist account and examine Godfrey-Smith's recent argument, which can be outlined as follows:

1. For any sufficiently complex system B, there is a possible system that differs internally from B only in its transducer layer, and that has the input-output properties of a human agent with non-marginal mental properties.
2. Any sufficiently complex system with the input-output properties of a human agent is a functional duplicate of that agent.
3. Functional duplicates share all their mental properties.
4. Two systems that differ only in their transducer layers must either both have, or both lack, non-marginal mental properties (Godfrey-Smith, 2009: 288).

Before engaging in discussion, several notions need to be explained first. A transducer layer (TL) is intended to be a simple input-output device, receiving physical impacts which a system can use, thus being an 'interface between the system and its environment' (e.g., a retina or hair cells in the inner ear are such TL's). A control system represents everything functionally important which is not a TL[13]. Although a sufficiently complex physical system is not given any criteria (the term was inherited from Searle), a Hinckfuss' pail was taken to be such a system (for this purpose, we can regard it as being the least complex of the sufficiently complex). Finally, to be a non-marginal property is to be a property of the system whose change of TL would not lead to a change of its functional properties.

[12] Similarly, an ocean which adds another layer of send in two weeks can carry out the 'program' of adding 1 to itself fourteen times (Sober, 1999: 64).

[13] A control system consists of memory, intelligence etc. In short, whereas our transducer layers overlap significantly with that of animals and even plants, a control system distinguishes our mind from the minds of animals.

While premises 2-4 seem rather acceptable, the first premise is highly controversial. If Hinckfuss' pail is a complex system, the argument goes, there is a possible TL which can be assigned to it, so that it yields a system with the input-output profile associated with human functional organization. In what follows, I will first discuss a TL intended to bring about the similarity between a bucket of sea water and a human being, and then address a difficulty with Godfrey-Smith's argument.

It should be noted that Godfrey-Smith is reluctant to provide a description of a TL which would make a pail of water sitting in the sun a behavioral duplicate of human. For the sake of simplicity, he outlines a TL for a coke machine, promising that if what is at stake can be shown with such a system, it can be shown with any other system as well.

According to Godfrey-Smith, the bucket's input device will receive (identical) coins, while its output device will be 'the effects of ripples in the water on air molecules at the surface'. Admittedly, much is left to a designer in order to construe a coke-machine behavior, but this is the idea in a nutshell. One may still be suspicious about the construction of behavioral output resembling the one of a human being. Namely, the reason why we succeeded in making a water bucket being a behavioral duplicate of a coke machine might be related to the fact that the input and output device did not add anything to a water bucket. In other words, a bucket was sufficiently complex for *that* purpose. But, if we are to build a behavioral duplicate of a human, we can be distrustful that the change of TL of a water bucket would not add *too much* to the bucket itself. Put differently, we *might* produce a behavioral duplicate of a human out of a bucket, but it will not be a bucket any longer. Rather, it will be a robot.

Godfrey-Smith does not seem to acknowledge this. Although he is right in saying that one could add a TL to a bucket of water thereby making it a behavioral duplicate of human,

he is not right when implying that this will continue to be a bucket of water. In other words, I suggest, changing a TL into a more complex TL could *add* too much to a system which would cease to be what it initially was, becoming more complex. Interestingly enough, Godfrey-Smith never mentions that a change of TL might do that, but only emphasizes that a change of TL will not *change* or *reduce* the functional features a system already has. For example, if we change a human TL (e.g., hair cells in the inner ear) Godfrey-Smith writes, we would not lose a functional property of hearing. But that is immaterial for his purpose since he is concerned with the change of TL which would make less complex system to be a duplicate of a more complex system, but not *vice versa*.

For all these reasons, we might think that the first premise is false and that no sufficiently complex system can be made into a behavioral duplicate of a human agent. But the third premise (a part of the functionalist program) is inconsistent with falsity of the first premise, since it contends that there could be a system which is functionally identical with human even though it is not a human. Thus, the problem was not the first premise but the example which supports it – since a water bucket is *not* a 'sufficiently complex physical system'[14]. Once we establish this, the difficulty disappears and Godfrey-Smith's argument still holds. The only difference is that we will say that it is not a water bucket which could be made into a behavioral duplicate of human, but a more complex physical system such as the homunculi-headed system[15]. This system would be 'sufficiently complex' to be 'made into a behavioral duplicate of an intelligent agent' (Godfrey-Smith, 2009: 284).

[14] The similar argument appears in Block (2000: 324-5), although his treatment is less ambitious and consequently more persuasive since he treats a bucket of water as "a parity-detecting automaton", an action admittedly capable of being performed by bucket's TL.
[15] Cf. Block 2006.

2. 4 Multiple Realizability Arguments

The multiple realizability thesis (MRT) was outlined in the first section. We have seen that it served to refute central state identity theory. In this section, I will try to show that MRT consists of two claims, the first of which is an empirical observation which refutes the identity theory while the second is an inference from the first. By and large, the second claim was meant to be understood as a non-trivial conclusion. Thus, by taking the second claim to be a trivial statement, I aim to show that MRT understood as non-trivial inference is false.

2. 4. 1 *Shapiro's dilemma*

Let us first present Shapiro's insightful dilemma. The first horn of the dilemma states that the realizing kinds might differ in their causally relevant properties, in which case they turn out to be different kinds, so the multiple realizability does not come in. The second horn of the dilemma asserts that if realizing kinds do not differ in their causally relevant properties, the case of multiple realizability does not arise either (Shapiro, 2000: 647). Shapiro's "hard question" offers two horns of the dilemma, neither of which admits of multiple realizability. To support his thesis, he gives the following example:

> Similarly, what do camera eyes and compound eyes have in common other than the fact that they have the function to see? If they share many causally relevant properties, then they are not distinct realizations of an eye. If they have no or only few causally relevant properties in common then there are no or just a few laws that are true of both of them (Shapiro, 2000: 649).

2. 4. 2 *MRT as a trivial inference*

Shapiro's example denies the validity of MRT with regard to natural kinds. Before proceeding to the example with a mental state, let us discuss the claim that MRT consists of two theses, in order to argue that non-trivial interpretation of its second claim is false.

(1) There are many physical realizations of the same mental state.

(2) There is a multiple realized mental state.

Obviously, (1) is an empirical observation, whereas (2) is the *inference* from (1). Two things should be noticed. First, a sentence (1) denies the validity of the identity theory. Insofar as this is the case, the premise of MRT is by no means a trivial one. Nonetheless, it should be noticed that (2) is meant to be a *non-trivial* claim as well, namely a non-trivial *inference* from (1). If I am right, the truth of (1) does not imply non-trivial truth of (2) and in showing that (2) is a trivial inference, it will be shown that MRT understood as non-trivial *inference* is false.

According to the multiple realizability thesis, (2) is a non-trivial conclusion. I want to suggest that the non-triviality of (2) implies that the significance of physical or materialistic realization has been overrated[16]. Consider the role of a color in the conclusion (2*).

(1*) There are many color realizations of the same knife.

(2*) There is a multiple realized knife.

Namely, the role given to color in (2*) is analogous to the role given to physical realizer in (2) since neither physical medium nor colors make mental states and knives what they are.

[16] By the same token, it can overrate the significance of immaterial medium in which mental states would be realized as well.

The *conclusion* (2*) is a trivial one, since the fact that there are knives in different colors does not imply that knife is multiple realized, except in a *trivial* sense. In the *same* sense, it would be true that two *identical* knives would be multiple realized as well[17]!

What makes this conclusion trivial is the scope of the syntagm 'multiple realized'. It is too narrow since the property in question (color) is irrelevant for a knife to be a representative of its kind. Similarly, a physical realizer is irrelevant for a mental state to be a representative of its kind.

Although the first claim of MRT was not trivial since it refuted the dominating central state identity theory, we can conclude that the multiple realizability inference is a trivial one. Thus, being a trivial claim, the multiple realizability thesis cannot be falsified since it lacks an informative content.

[17] I reject trope-theoretic account on which there are no two identical things. None of the properties which two knives have would be the same property (Daly, 1994)

3. Conceptual Behaviorism: The Return

3. 1 An Overview

Behaviorism is a term used for several views in the philosophy of mind in the XX century. Three most important strands of behaviorism in the previous century were:

(i) Metaphysical behaviorism

(ii) Analytic behaviorism

(iii) Psychological behaviorism

It should be noted that these versions of behaviorism do not entail each other. Namely, as Kim showed, metaphysical behaviorism does not imply either analytic behaviorism or psychological behaviorism. At the same time, these two strands of behaviorism do not entail metaphysical behaviorism either (Kim, 1998: 38–9). This is not always recognized or is taken to be irrelevant as some authors suggest that analytic or logical behaviorism is a 'variety' of metaphysical behaviorism (Kukla & Walmsley, 2006:114).

(i) Metaphysical, eliminative or ontological behaviorism denies the existence of mental phenomena and contends that there is nothing more to a mental state and its 'nature' than behavior associated with it (Kim, 1998: 38). As such, metaphysical behaviorism is a precursor of eliminative materialism. One of its most famous proponents was Quine who adopted this view based on his ideas on the indeterminacy of translation and the claim that belief and desire cannot be logically represented (Byrne, 1994).

Metaphysical behaviorism, it should be noticed, depends on the premise that dualism is false (Kukla & Walmsley, 2006: 114). But that being so, analytic behaviorism is not a

variety of metaphysical behaviorism, as Kukla and Walmsley opined, since, being a semantic theory, analytic behaviorism is not incompatible with dualism like metaphysical version of it.

(ii) Analytic or logical behaviorism is a theory of meaning (Putnam, 1997b: 425) or more precisely a theory of meaning of psychological expressions (Kim, 1998: 38). Mental notions and propositions about the mental are semantically equivalent to propositions about behavior and behavioral dispositions. Arguably, this view has been defended by Gilbert Ryle in his *The Concept of Mind* and Ludwig Wittgenstein in his *Philosophical Investigations*[18].

Analytic behaviorism was threatened by number of objections which attempted to show that its translations cannot proceed for various reasons. Perhaps the most resilient argument is the one which insists that in final instance behaviorism must embrace some mental terms. For instance, suppose that Amy has a belief that there is a perfect odd number. Since she would be inclined to manifest such a mental state by verbal behavior[19], analytic behaviorism would propose that the sentence

(1) 'Amy believes that there is a perfect odd number'

should be translated as:

(1') 'If asked, Amy will be disposed to say that there is a perfect odd number'

But at the same time, Amy will be disposed to utter this sentence only if she *wants* other people to hear it, if she *thinks* that they understand what she means by these words etc. Since the italicized predicates are mental, the behaviorist translation fails.

[18] According to Armstrong (1993), although they did not admit they were behaviorists, their seminal works propound these ideas. Luckhardt (1983) argues against the presence of behavioral doctrine in Wittgenstein's *Investigations* whereas Byrne (1994) refers to *The Concept of Mind* as being 'quasi behaviorist polemic'.

[19] The same mental state could be manifested by other behavioral outputs as well; e.g., doing the calculations in order to prove that there is such a number etc., but according to foes of analytic behaviorism, all of these behavioral outputs would be treated as containing *some* mental predicate.

(iii) Psychological or methodological behaviorism is a philosophical thesis about psychology[20]. It famously argued that psychology should be restricted to the examination of behavior (Watson, 1913). Evidently, psychological behaviorism does not rest on dualism so metaphysical behaviorism does not imply psychological behaviorism either. Its argument goes as follows:

(1) Scientific data must refer to publicly observable events.

(2) Physical events are public; mental events are private.

(3) Reports of physical events (like behavior) can be scientific data, but reports of mental events can't (Kukla & Walmsley, 2006: 110-1).

Thus, according to the psychological behaviorist, a science is reduced to a physical science and the private states do not enter this scientific picture.

3. 2 Conceptual Behaviorism

Although tenets of conceptual behaviorism are sometimes interpreted as a part of analytic behaviorism[21], conclusions of the former do not entail the *semantic* nature of the latter. Unlike analytic behaviorism, as I will explain it shortly, conceptual behaviorism does not have an ambition to provide translations of mental statements.

In order to answer the fundamental question 'what is a mental state', conceptual behaviorism emphasizes that we need to analyze our ordinary concepts of mental states, acquired on the ground of manifestation which proceeds in terms of *behavior*.

With respect to the fundamental question, we can distinguish between three answers: empirical, introspective and conceptual, the last of which represents conceptual behaviorism.

[20] Arguably, it belongs to the philosophy of psychology.
[21] Armstrong, D. (1993:58)

3. 2. 1 *Empirical answer*

The empirical answers provide information about the bearer or the neurophysiological correlate of a given mental state. The history of the philosophy of mind largely relies on such answers. The identity theory was advancing a scientific hypothesis, on par with propositions such as 'Lightning is a motion of electric charges' (Place, 1956). Analytic functionalism defended the identity between experience and neuro-chemical state (Lewis, 1966: 17) whereas Fodor and Block provided empirical considerations in favor of the multiple realizability thesis (Block and Fodor, 1972). Interestingly, even the refutation of these theories proceeded on the empirical grounds[22].

For this reason, empirical answers are more scientific than philosophical. Far from saying that such investigation should be abandoned, not being able to 'resolve the mystery' (McGinn, 2006), I want to suggest that the real problem is that empirical answers are conceptually *irrelevant* for our mental concepts. Our test for relevancy will be that if all the answers of a given type[23] fit equally well to the conceptual scheme, the answers will be treated as irrelevant. The empirical answers to the question what is pain might be various: the C fibers firing, the A fibers firing, perhaps a process in a soul, or even a screen in our head with 'Pain!' written on it. Admittedly, all of these answers fit *equally* well our conceptual scheme, and according to our test they are all irrelevant. Of course, they would not fit physiological facts, but the usage of our mental concepts in most cases is not related to physiological discoveries[24].

Thus, it is not a scientific fact that we are looking for when asking what pain is. If it was, we might frown when told about any of these solutions. We might debate whether it was

[22] Putnam, among others, proposed such a refutation (Putnam, 1997a)
[23] This assumes that some of these answers are distinct enough so that the truth of one will imply the falsity of the other.
[24] In some instances, the physiological facts will influence the usage of certain mental concepts. E.g., we will stop using the expression 'my brain hurts' once we get to know that this is not possible.

A or D fiber firing supposed to account for pain. But we never do. Only neurophysiologists can frown and debate here.

3. 2. 2 *Introspective answer*

In general, two most commonly cited types of introspectible mental states are *attitudes* and *conscious experiences* (Schwitzgebel, 2010). They are said to have 'an introspectable interior' which is described by various qualities (Goldstein, 1994: 49). Thus, when I feel pain, I also feel some interior qualities associated with it, which are lacking when I am not in pain. Such knowledge is held to be accessible only to a person who is doing the introspection.

Unlike the empirical answers, different introspectable interior answers need not fit equally well to the conceptual scheme[25]. Rather, different 'interior' properties we ascribe to pain or other mental states render the introspective answers to be *relevant*. Still, handling such intrinsic qualities, inaccessible to anybody else to be *intrinsic* to mental states as the so-called A-causal properties (Goldstein, 1994), is highly problematic since its *criterion* is hard to find:

> [I]f there were such inner states and operations, one person would not be able to make probable inferences to their occurrence in the inner life of another (Ryle, 1949: 54).

The internal processes which Goldstein invokes are not only problematic for reasons which Ryle describes, but also because they do not have a conceptual role in psychology (Ben-Yami, 2005). For this reason, I want to suggest the move towards a conceptual answer.

[25] I don't want to suggest here that our mental concepts should include interior qualities acquired by introspection.

3. 2. 3 *Conceptual answer*

Although our fundamental question 'what is mental state' could be answered in an empirical fashion or in an introspective manner, there are better reasons for it to be answered in a *conceptual* way.

> When we characterize people by mental predicates, we are not making any untestable inferences to any ghostly processes […] which we are debarred from visiting; we are describing […] their predominantly public behavior (Ryle, 1949:51).

When we ascribe mental predicates such as 'being intelligent' or 'being in pain', we rely on certain facts about the behavior[26]. Arguably, before people started to use these mental predicates, they could have had observed that, given the same stimulus, behavior of someone who is later to be named intelligent differed from the one who is not. Thus, the new notions, mental predicates, were supposed to replace more long-winded expressions consisting of behavioristic terms, thereby introducing a 'stylistic alternative' (Kukla and Walmsley 2006:116).

Adopting the view that our mental concepts were formed (and that some mental concepts are still forming) in this way, in answering our fundamental question we should reverse this process. From this it follows that our conceptual answer is already contained in our question, at least to a certain degree. This is contained in the observation that when we ask 'what is pain' we presuppose questions such as 'what is *x* which is caused by injury/infection/pinched nerve etc.?' and 'what is *x* which makes us wince, writhe, yell, etc.?' Thus, the concepts of cause, effect and objects are already presupposed.

[26] Even among non-behaviorists it is widely believed that the behavior and tendencies to behave enter our concepts of mind (Armstrong, 1993: 68). It follows that behaviorism is erroneous not for thinking that mental concepts contain behaviorist elements, but for holding the view that behavior and tendencies to behave *exhaust* our mental concepts.

As opposed to an empirical answer, conceptual answers in most cases do not provide the new information. At most, this answer *adds* something to our concept (we can learn, for instance, that one cannot have pain in a brain) which further enables us to use this mental notion in other contexts.

A conceptual answer, it should be noticed, does not need to conflict with an empirical answer. Following Wittgenstein, we can allow that it could be a *chaos* in our heads [Z, § 608] thereby denying that *whatever* is inside our heads is not related to our concepts of mental states, which is precisely what philosophers strived to find out. Still, Wittgenstein sometimes seem to propound a different view when saying that we might be on a wrong track trying to associate mental state with a physiological process [Z, § 610]. From the conceptual point of view, one should not either defend or argue against it, but only acknowledge that the whole issue goes beyond the scope of what our answer can provide.

Finally, the test for relevancy is 'positive' for conceptual answers, and unlike with the introspective answer, the criterion can be given.

3. 2. 4 *Conceptual Behaviorism and Translation*

It was said that analytic behaviorism does not intend to provide translations of mental statements. This has got to do with two things. First, such translations, to be possible, would require ideal conditions, viz. acquaintance with the whole input history and output data. However, in order to tell what is a concept of mental state we do not need to provide an exhaustive description of the input history and output data. In fact, when having such concepts we can still make mistake as far as translations are concerned (indeed, we do this in our everyday life). Secondly, sentences such as "He stopped *behaving* in pain-way as soon as paracetamol decreased prostaglandin production" is not meaningless. 'Meaning' of pain, in a

certain sense *can* be translated by its neural correlate. The only problem is that this meaning is not conceptual, but empirical and thus irrelevant.

Motives for conceptual behaviorism are that it cannot be *falsified* by the science, since it does not offer any empirical hypothesis. In this respect, it bears resemblance to anomalous monism since neither of these theories requires any empirical evidence for their claims (Yalowitz, 2005). Thus, conceptual behaviorism does not offer any hypothesis about the neurophysiological nature of physical event and cannot be rejected by science (as phlogiston theory or caloric theory were rejected); moreover, it does not purport to explain which *relations* hold with respect to mental state either, which philosophers eagerly postulate (token-token, type-type, multiple realizability etc.) Because of this trait, conceptual behaviorism will stay compatible with the correct scientific solution since it offers the framework to which any scientific solution would fit. In this respect, conceptual behaviorism is topic-neutral.

3. 3 Criteria Behaviorism

The viewpoint of *criteria* behaviorism is that for two systems which bear resemblance in their behavior, it is fair to say that their internal processes would scarcely matter at all (Ben-Yami, 2005). A challenging view is that of psychologism, which emphasizes the role of internal processes and states that 'whether behavior is intelligent behavior depends on the character of the internal information processing that produces it' (Block, 1981).

In his paper 'Psychologism and Behaviorism' (1981), Ned Block proposes an interesting argument against behaviorism[27]. To reconstruct his argument, let us first define some essential notions. 'A typable string of sentences' will be understood as a string of

[27] However, Block's argument attacks functionalism as well. As we will see, Block's argument rejects the possibility of the homunculi-heads' intelligence, which is by and large adopted by functionalists.

sentences whose members can be typed by a human typist in relatively short time. The set of all these strings will include both 'sensible strings' and those which are not sensible, whereas the former will be 'naturally interpretable in conversations in which at least one party's contribution is sensible'. In more detail:

> [I]f we allot each party to a conversation one sentence per 'turn' […] and if each even-numbered sentence in the string is a reasonable conversational contribution, then the string is a sensible one (Block, 1981).

The argument depends on the processing of a sophisticated machine which emits sensible strings of sentences to verbal stimuli resembling an intelligent agent. Such a system would be able to 'exercise imagination and judgment' and its working is supposed to proceed in the following way:

> The interrogator types in sentence A. The machine searches its list of sensible strings, picking out those that begin with A. it then picks one of these A-initial strings at random, and types out its second sentence, call it 'B'. The interrogator types in sentence C. The machine searches its list, isolating the strings that start with A followed by B followed by C. It picks one of these ABC-initial strings and types out its fourth sentence, and so on (Block, 1981).

All that is required to build such a machine is to store a large amount of sensible sentences in its memory and to write a program which enables the machine to pick a sensible sentence from its memory depending on a similarity with a question asked.

Once this is accomplished, Block's 'mindless machine' will presumably threaten the behaviorists' account of intelligence according to which an agent is intelligent if it passes *neo-Turing Test*. Since Block's machine passed this test, it is fair to say that it exhibits intelligence. Its behavioral criteria, viz. the output we read, strongly suggest this conclusion.

But for all we know about its internal structure, it is not the machine which is intelligent, but its programmers. Block's conclusion is that behaviorism is false since the machine's capacity to emit sensible responses is not sufficient for intelligence as suggested by behaviorism.

Interestingly, Block's argument seems to rely on a restrictive interpretative viewpoint about behaviorism which robs it of one of its most important theoretical feature: the *input history*[28]. By the input history I understand the set of all physical data relevant for a given mental state or a quasi-mental state, such as the purported intelligence of Block's machine. In order to delineate the input history of a machine performing the neo-Turing test from the input history of a machine performing some other test, it would be enough to list the physical data which are responsible for possible machine's outputs on the Turing test. For instance, the input history of Block's machine includes building of a memory which contains sensible sentences, writing of a program which 'picks out' sensible statements according to given rules etc.

Now, it is precisely because of this negligence that Block's argument seems plausible. Indeed, looking *only* to the machine's output does suggest that a machine is intelligent. However, everything in the input history of the machine alludes to the quite different conclusion. The input history of the machine contains the memory and the program built by intelligent human agents which are responsible for machine's putative intelligence[29]. I will try to show that there is nothing over and above *these* features which constitutes the machine's intelligence.

It is important to notice that the relation between the input history and behavioral outputs is utterly different with respect to Block's machine and the intelligent systems.

[28] Cf. Teichmann (2001) and Kukla & Walmsley (2006).
[29] Interestingly, already Descartes expounded similar views by saying that machines could never use the language in the way people do (Descartes, 2006: 46).

Namely, the intelligent agents are able to produce sensible responses without having them in their input history (for instance, without having read them or heard them). In this sense, the behavioral output of an intelligent agent cannot be mapped in biunivocal manner to its input history, since its input history lacks this sensible answer. At the same time, the comparison of behavioral input and output of Block's machine will reveal different fact, namely that the behavioral output can be mapped in biunivocal manner to its input. Thus, a sensible response is only a *reproduction* of a sentence already stored in Block's machine which precludes Block's machine to be an intelligent system (Ben-Yami, 2005). Since Block's argument rested on the premise that behaviorism will admit that this machine is intelligent, his conclusion is invalid.

3. 4 Concluding remarks

3. 4. 1 *Topic-neutrality*

I have argued that the empirical considerations of neurophysiology are irrelevant for our mental concepts. Although topic-neutrality could have been a motive for conceptual behaviorism to treat empirical notions in this way, I have tried to show that conceptual behaviorism was topic-neutral to begin with. This being the case, conceptual behaviorism is coherent with scientific progress and cannot be falsified by it, since it does not put forward any scientific hypothesis with respect to mental states.

3. 4. 2 *Typical cause and effect*

Admittedly, there is no typical input or output of many mental states. But once behaviorism allowed for an indefinitely many behavioral dispositions in accounting for the mental states,

the difficulty emerged with respect to such open-ended lists of causes and effects. Namely, it was proven that these lists did not have any explanatory power. Accordingly, functionalists proposed two lines of thought: the characterization of mental states as being *de facto* typical, and grouping these open-ended lists into many coherent systems of typical causes and effects. Appealing to the notion of the typical causes and effects, functionalism tried to eschew the threat to behaviorism but both of these approaches turned out to be problematic. With regard to the former, we have argued against the coherency of functionalism in section 2.1 by showing that hallucinations of which we are not aware do not have typical causes and effect. With regard to the latter, we argued that typical profiles of mental states are indefinitely multitudinous. Correspondingly, the functionalist classification model, which allowed for the possibility that two tokens of the same type have less in common that two tokens of the different type, was dismissed as invalid.

Finally, since conceptual behaviorism admits that we *know* the mental concept before we try to explain it, open-endedness does not come up as a problem.

3. 4. 3 *Inner mental state*

Although behaviorism is not a unified theory, what was programmatic for all of its versions was its refrain from the concept of inner states. Since our mental concepts do not refer to such 'occult episodes' of our mental life, inner mental states are avoided in conceptual behaviorism. The way we use our mental concepts is restricted to its manifestation, namely behavior.

3. 4. 4 *Multiple realizability*

In section 2.4 we attempted to refute the multiple realizability thesis by showing that although its premise was not trivial being sufficient to refute the central state identity theory, its conclusion was trivial. Nonetheless, the multiple realizability thesis was significant for removing the anthropocentric claims that mental states should be defined with reference to solely human physiological features. It should be noted that conceptual behaviorism is not chauvinistic either. It is perfectly consistent with a possibility that the use of mental concepts might proceed in the way different from ours.

References

Armstrong, D. M., (1993), *A Materialist Theory of the Mind*. London and New York: Routledge.

_____. (1999), 'The Causal Theory of Mind' in: Lycan, W. G. (ed.), *Mind and Cognition. An Anthology*. Blackwell Publishers Inc, pp. 20–27.

Ben-Yami, H. (1999), 'An Argument Against Functionalism', *Australasian Journal of Philosophy* 77, pp. 320–24.

_____. (2005), 'The "Hercules" in the Machine. Why Block's Argument Against Behaviorism is Unsound', *Philosophical Psychology* Vol. 18, No. 2, pp. 179–86.

Bickle, J. (2006), 'Multiple Realizability' in: *Stanford Encyclopedia of Philosophy*.

Block, N. (1981), 'Psychologism and Behaviorism', *Philosophical Review* Vol. 90, No. 1, pp. 5–43.

_____.(1990), 'Inverted Earth', *Philosophical Perspectives*, Vol. 4, Action Theory and Philosophy of Mind, pp. 53–79.

_____. (2000), 'Functionalism' in: Guttenplan, S. *A Companion to the Philosophy of Mind*, pp. 323–332.

_____. (2006), 'Troubles with Functionalism (Revised)', in: Beakley, B. and Ludlow, P. (eds.), *The Philosophy of Mind*, Classical Problems/Contemporary Issues, pp. 107–131

Block, N. & Fodor, J. (1972), 'What Psychological States are Not', *The Philosophical Review*, Vol. 81, No. 2, pp. 159–181.

Byrne, A. (1994), 'Behaviorism', in: *A Companion to the Philosophy of Mind*, Guttenplan, S. D. (ed.), Blackwell.

Chalmers, D. (2002), 'Does Conceivability Entail Possibility?', in Gendler, T. & Hawthorne, J. (eds), *Conceivability and Possibility*, Oxford: Oxford University Press, pp.145–200.

Daly, C. (1994), 'Tropes', *Proceedings of the Aristotelian Society*, New Series, Vol. 94, pp. 253–261.

Descartes, (2006), *A Discourse on the Method*. Translated by Ian Maclean. Oxford, New York: Oxford University Press.

Geach, P. T. (1967), 'Identity', *The Review of Metaphysics*, pp. 3–12.

Godfrey-Smith, P. (2009), 'Triviality arguments against functionalism', *Philosophical Studies*, Vol. 145, No. 2, pp. 273–295.

Goldstein, I. (1994), 'Identifying Mental States: A Celebrated Hypothesis Refuted', *Australasian Journal of Philosophy*, Vol. 72, No. 1, pp. 46–61.

Graham, G. (2007), 'Behaviorism' in: *Stanford Encyclopedia of Philosophy*.

Heil, J. (2000), *Philosophy of Mind. Contemporary Introduction*, Cambridge: Cambridge University Press.

Hempel, Karl. (1980), 'The Logical Analysis of Psychology', in: *Readings in the Philosophy of Psychology*, Block, N. (ed.), Cambridge: Harvard University Press, pp. 15-23.

Kim, J. (1989), "Mechanism, Purpose, and Explanatory Exclusion", in J. Kim, *Supervenience and Mind*. Cambridge, Cambridge University Press.

_____. (1992), "Multiple Realization and the Metaphysics of Reduction." *Philosophy and Phenomenological Research*, 52, pp. 1–26.

_____. (1998), *Philosophy of Mind*. Oxford: Westview Press.

Kukla, A. and Walmsley, J. (2006), *Mind. A Historical & Philosophical Introduction to the Major Theories*. Indianapolis/Cambridge: Hackett Publishing Company Inc.

Levin, J. (2009), 'Functionalism' in: *Stanford Encyclopedia of Philosophy*.

Lewis, D. (1966), 'An Argument for the Identity Theory', *The Journal of Philosophy*, Vol. 63, No. 1, pp. 17–25.

_____. (1972), 'Psychophysical and Theoretical Identifications', *Australasian Journal of Philosophy*, Vol. 50, No. 3, pp. 249–258.

_____. (1991), 'Mad Pain and Martian Pain', in D. Rosenthal, ed., *The Nature of Mind*. New York: Oxford University Press, pp. 229–235.

Lowe, E. J. (2000), *An Introduction to the Philosophy of Mind*, Cambridge: Cambridge University Press.

Luckhardt, C. (1983), 'Wittgenstein and Behaviorism', *Synthese*, Vol. 56, No. 3, pp. 319–338.

Lycan, W. (1981), 'Form, Function, and Feel', *The Journal of Philosophy*, Vol. 78, No. 1, pp. 24–50.

_____. (2000), 'Functionalism' in: Guttenplan, S. *A Companion to the Philosophy of Mind*, pp. 317–323.

McGinn, C. (2006), 'Can We Solve the Mind-Body Problem?' in: Beakley, B. and Ludlow, P. (ed.), *The Philosophy of Mind*, Classical Problems/Contemporary Issues, pp. 107–131

Nagel, T. (1974), 'What Is It Like to Be a Bat?', reprinted in: Beakley, B. and Ludlow, P. (2006), (ed.), *The Philosophy of Mind*, Classical Problems/Contemporary Issues, pp. 255–265.

Piccinini, G. (2009), 'Computationalism in the Philosophy of Mind' *Philosophy Compass*, Vol. 4, No. 3, pp. 515-532

Place, U.T. (1956), 'Is Consciousness a Brain Process?', *British Journal of Psychology*, 47, pp. 40–50.

Putnam, H. (1997a), 'Philosophy and Our Mental Life', in: *Mind, Language, and Reality*. Cambridge: Cambridge University Press, pp. 291–303.

_____. (1997b), 'The Mental Life of Some Machines', in: *Mind, Language, and Reality*. Cambridge: Cambridge University Press, pp. 408–429.

_____. (1999), 'The Nature of Mental States', in: Lycan, W. G. (ed.), *Mind and Cognition. An Anthology*. Blackwell Publishers Inc, pp. 27–34.

_____. (2001), *Representation and Reality*. MIT Press.

Pylyshyn, Z. (1984), *Computation and Cognition: Toward a Foundation for Cognitive Science*. MIT Press.

Robinson, H. (1982), *Matter and Sense*: *A Critique of Contemporary Materialism*. Cambridge: Cambridge University Press.

Ryle, G. (1949), *The Concept of Mind*, London: Hutchinson's University Library.

Schwitzgebel, E. (2010), 'Introspection' in: *Stanford Encyclopedia of Philosophy*.

Shapiro, L. (2000), 'Multiple Realizations', *The Journal of Philosophy*, Vol. 97, No. 12, pp. 635–654.

Smart, J. J. C. (1959), 'Sensations and Brain Processes', *The Philosophical Review*, Vol. 68, No. 2, pp. 141–156.

Searle, J. (1990), 'Is the Brain a Digital Computer?' *Proceedings and Addresses of the American Philosophical Association* 64, pp. 21–37.

Sober, E. (1999), 'Putting the Function Back into Functionalism' in: Lycan, W. G. (ed.), *Mind and Cognition. An Anthology.* Blackwell Publishers Inc, pp. 27–34.

Teichmann, R. (2001), 'The Functionalist's Inner State', in: Schroeder, S. (ed.), *Wittgenstein and Contemporary Philosophy of Mind*, pp. 24–35.

Watson, J. (1913), 'Psychology as the Behaviorist Views It', *Psychological Review*, Vol. 20, pp. 158–177.

Wittgenstein, L. (1998), *Zettel*. Oxford: Basil Blackwell.

Wittgenstein, L. (2009), *Philosophical Investigations*. Translated by: G. E. M. Anscombe, P. M. S. Hacker, and J. Schulte. Blackwell Publishing Ltd.

Yalowitz, S. (2005), 'Anomalous Monism' in: *Stanford Encyclopedia of Philosophy*.

Zhu, J. (2006), 'In Defense of Functionalism', *Philosophia* 34, pp. 95–99.