AN ARGUMENT IN FAVOR OF MENTAL REALISM

Abstract

This paper proposes a philosophical defense of the reality of mental states. In the first section it will be tested two very important contemporary accounts regarding mental realism, respectively, the theory of anomalous monism and the non-reductive physicalism. The test has to show us whether their upholding of irreducibility of intentionalist explanation (called also mentalist or psychological explanation) bears any significant consequence on the reality of mental states. The conclusion will be, unfortunately, that from their accounts as such it does not follow that the mental states are real. In section 2 we try to develop a realist argument which, accepting that the argument given by anomalous monism and non-reductive physicalism in favor of irreducibility is valid, asserts that this irreducibility really needs real items in order to work. Such an argument takes the shape of a transcendental argument. In section 3 it will be shown that mental causation, though real, doesn't imply Cartesianism.

Keywords: mental realism, non-reductive physicalism, anomalous monism, mental causation

Introduction

This paper proposes a philosophical defense of the reality of mental states. First, it will be tested two very important contemporary accounts regarding mental realism, respectively, the theory of anomalous monism and the non-reductive physicalism. The test has to show us whether their upholding of irreducibility of intentionalist explanation (called also mentalist or psychological explanation) bears any significant consequence on the reality of mental states. The conclusion will be, unfortunately, that from their accounts as such it does not follow that the mental states are real.

1 By “mental states” we understand all the intentional facts such as thinking, desiring, planning, intending, suffering, etc. We use both “mental state” and “mental event” with the same meaning, though they may be different, but this does not matter in this text.

2 Other forms of non-reductive theories, as emergentism, are not treated directly, but the results reached by testing non-reductive physicalism apply also to emergentism.
This kind of failure may be understood as a consequence of a direct confrontation of mental realism with the metaphysical principle of the causal closure of the physical world. Faced with the proper task of a realist account, namely, causal efficiency, an adept of non-reductive explanation has serious difficulties to avoid the closeness entailed by that principle. This will be discussed in section 1. In section 2 we try to develop a realist argument which starts right from this irreducibility. The point will be that this irreducibility really needs real items in order to work. Such an argument takes the shape of a transcendental argument. In section 3 it will be shown that mental causation, though real, doesn't imply Cartesianism.

Now it is necessary to show what does a realist account of mental properties mean for us. A good criterion may be to assert that something exists, if this thing can produce something else, can influence the occurrence of an effect. To put it differently, something is real, that is, is individualized as a distinct thing, if its occurrence does influence the occurrence of other things, if its presence does matter in its environment. I will call this criterion the causal test (CT). This test is quite different from reduction. Reduction, as it is understood in the analytic philosophy, says that something is real just when it is possible to explain it through the underlying physical basis, that is, to explain its causation by means of the physical processes. Our test judges the causation not by reducing it, but by establishing a counterfactual connection between cause and effect. In the following we shortly explain this counterfactual expression of causation. Given the factual statements:

1) “The streets are wet because it rained last night”, or
2) “John is going to the bus station because he wants to visit his mother“.

Here we have expressed the causes whose presence determined the effect: the rain, John's desire to visit his mother. The modal character of 1) and 2) can be proved counterfactually as follows:

---

3 This criterion appears also in Fodor, 1989, p.61, Shoemaker, 1999, p. 110. Applied to properties, this test helps us to distinguish a property whose presence instantiated in a fact changes the behaviour of that fact from properties that make no such change. Therefore, the property will not be defined by its physical substrate or by linguistic means, but by the causal difference made by that property.

4 Not to be confounded with the control test of Lynne Rudder Baker.
1') “If it had not rained, then the streets wouldn't have been wet”, and

2') "If John had not wanted to visit his mother, then he wouldn't have gone to the bus station”.

In both cases, we observe, the causes are sufficient conditions for the occurrence of the effect, but they are not necessary since there are also other facts that might have produced the wetness of the streets, or might have determined John to visit his mother (he could be a sleep walker). Hence, the relation expressed in 1) and 2) is contingent, context dependent, 1') and 2') are true, if the other conditions remain the same. In Davidsonian terms, 1) and 2) are non strict laws. Now, I ask myself if this is a good reason to claim that this relation is not a causal one. Does a causal relation really need to hold necessary? The answer is, I think, negative. There are few cases instantiating a causal relation in all possible worlds, most of all require a clause or other of ceteris paribus type. The modality does not hinder a relation to hold, it just describes in how many contexts it holds. For example, a causal relation between c and e holds just in one certain context, nevertheless we are justified to accept that c causes e because in this context the occurrence of c suffices to produce e. The metaphysical assumption that a causal relation always instantiates a necessary connection is based on the confusion between necessity and causality itself. Using an old distinction, we may say that the basic condition for the causal relation is sufficiency, not necessity. Sufficiency gives the intensity of how c influences the occurrence of e, and necessity gives us the amplitude of this relation, namely, the number of worlds in which it holds. The number of worlds can vary depending on the possibility of reproducing the same context in those worlds. The important thing is sufficiency because just from a certain degree of influence c begins to determine e. The c-e relation turns out into a causal one just in the moment when c is a sufficient cause to produce e. In our examples, 1. and 2. are not without exception, but in a

---

5 From Aristotle, Spinoza, Kant, and nowadays Davidson and Kim it was argued in favor of the nomological character of causality. One of the principles of Davidson’s monism is that “events related as cause and effect fall under strict deterministic laws”. See Davidson, 1970, p. 208. Kim claims also that non strict laws (like those of psychology) must be backed on strict relations in order to express efficacious facts. See Kim, 1992, p.144.

6 David Lewis holds the view that regularity is an "almost but not quite sufficient" condition to explain causal dependence. Lewis, 1999, p. 187.
certain context, under certain circumstances, John's desire really caused him to go to the bus station.

Criticizing the Lewisian account of causality, Kim formulated, among other things, a condition for causal statements, namely, to have predictive power (Kim, 1999, p. 192). And this is, I think, a symptom of using the same assumption of the nomological character of causality as a condition for the existence of causality. Put it shortly, this assumption says that the causal relation is just the one that can instantiate a regularity so as to permit predictions. It follows that such predictions are available only between strict interactions, that is, just in the domain of the physical sciences. This is the reason why Davidson insisted that the causal relations hold only among extensional facts.

It is not my purpose to discuss here in full extent this assumption. I will give some arguments against it and these could be considered also arguments in favor of the counterfactualist approach to causality. First, the prediction is enabled by the necessity feature of the causal claim since we can predict something on the basis of regularity, that is, on the capacity to repeat the production of the effect in similar contexts. If we agree that the degree of modality does not impede a cause to produce the effect, then prediction feature ceases to be an argument for limiting causality just at the physical domain.7 Second, we have to make some distinctions in order to clarify the nomological character of causality:

- There are two distinct domains: nomological domain (belonging to the order of knowledge) and that of empirical instances (order of being);

- one event causes another event even if we are not able to pick out that causal relation (I would like to call this assumption the principle of independent existence of causality);

- Regarding the order of being, it is possible for the causal relation to exist in just one context. An adept of strict determination could, in turn, argue that, being given all conditions in that context, we have to obtain always the same effect when we reproduce the context with all conditions and hence, one might conclude that causal relations must always instantiate general relations. I think that he will be right, but empirically we face sometimes a) irreversible conditions or b)

---

7 Causal relations depending on ceteris paribus clauses could be made stricter if we admit that those laws work only in worlds where the ceteris paribus clause are satisfied. See Fodor, 1989, p. 74.
infinite conditions. For example, Peter could lift 100 kg when he was 20 years old, but the same person is unable to do such thing after 20 years. To repeat exactly the same conditions with John who is now 20 years old might imply an inventory of qualities with infinite terms. Neither a), nor b) is an argument for saying that Peter could not lift 100 kg 20 years ago.

Given this arguments, we can use now a counterfactualist analysis of causality. I am aware of the fact that this account is not perfect, but I think it would be far more helpful in avoiding philosophical wrong roads in comparison to other accounts.

1. Non-reductive physicalism and epiphenomenalist objection

There are aproximative two strategies to prove the irreducibility of mentalist explanation simultaneously with the asserting of physicalism:

1.1. To emphasize the specific nature of mental states, their categorial differences that make impossible the attempt to reduce it to the underlying physical basis (anomalous monism).

1.2. To point out the too heterogeneous features of the physical subvenient basis that make it unable to support a univocal deduction of the mental states (multiple realization argument).

What do these strategies consist in and do they really succeed to demonstrate the reality of mental states? We want to see, relative to realist requirements, how the postulation of irreducibility of intentionalist explanation works in these two important theories in the philosophy of mind.

1.1. Anomalous monism

According to Donald Davidson, the assignment of mental states is ruled by other criteria than those used in natural sciences. A mental state has intentional content only in connection to other mental states and on the basis of rational principles. As Davidson said, “it is a feature of the mental that the attributions of mental phenomena must be responsible to the

---

8 Kim argued that the counterfactual analysis is too large since it includes also relations that are not causal, like, for example, in the following sentence: "If my sister had not given birth at t, I would not have become an uncle at t". It is determination but not causality. See Kim, 1999, p. 191.
background of beliefs and intentions of the individual. An expression of a belief is mediated by the presence of other beliefs, intentions and so on. The interpretation of mental states has to take into account this mental holism regulated by the principles of rational coherence while the physical counterpart is described by means of the extensional strict laws. Given this categorial difference, the reduction of the mental domain to the physical domain is not possible since the necessary bridge laws are not possible. There is no possibility to translate the psychological explanation into the physical explanation because the latter does not work in accordance with the principle of rational coherence. Hence, mental events will be described only in mental terms. This is the anomaly of the Davidsonian monism: there is only one substance, but one part of its events is not describable in terms of the physical sciences.

On this account, mentalist explanation has a lot of independence, but how can we reconcile this with the first principle of monism which asserts that "singular causal relations are backed by strict laws" (Davidson, 1993, p. 3)? If this is true, and being given the second principle of monism, the fact that strict laws hold only between physical events, then the causality is bound to the physical domain. Mental events have no causal role as such. The conclusion is, though Davidson would not accept it, that the mental events are epiphenomenal. For example, the behavior is a physical fact and mental facts play no role in producing it because any physical fact has a strict physical determination which will render futile a new mental causation.10

But Davidson is an adept of mental causation. He insists that an event can cause apart from its description: "if causation is a relation between events, it holds between them, no matter how they are described. So there can be descriptions of two events (physical description) which allow us to deduce from a law that if the first event occurred, the second would occur, and other descriptions (mental descriptions) of the same events which invite no such inference" (Davidson, 1993, p. 6). In our case, John's behavior can be mentalistically explained in terms of desires and beliefs, whereas just the naturalist explanation in terms of neural events will pick out the real causes. According to the above mentioned CT, an event is real if it is causally potent. But we observe that, according to anomalous monism, mental events are simple names, they have qua mental no causal

9 Davidson, 1970, p. 222.
10 This is one of Kim's reproaches to anomalous monism. See Kim, 1993, p. 270.
role. This consequence results even from the alleged situation in which mental events cause as physical events and from the explanatory completion that this fact does not imply the impotence of the mental because an event cause apart from its description. Continuing this idea, Le Pore and Loewer affirm that Davidson's account of mental causality means the possibility of mental events of supporting counterfactuals independently from their basic properties, though mental properties *qua* mental cannot cause (see Le Pore, Loewer, 1987, p. 636). 11 Thereby it is possible, following this account, to speak about mental causation without to endanger the monism (expressed through the causal closure principles). Even with this commentary, I do not understand how mental causation is possible. Moreover, it is inappropriate to speak about properties that are causally relevant but they are not instantiated *as such* in the causing event.

If mind is to be distinguished from physical events, we have to accept that it causes in a specific way, apart from a physical causation. How can this anomaly – physical causing events, but intentionalist description – express a distinct property of mental, if it means just a specific mode of description? Davidson's theory amounts to the following representation: we have two events c1 and c2 picked out in two corresponding explanations T1 and T2. c2 represents the subvenient base of c1, but c1 is such an event that just in T1's way can be described. c1 cannot cause by itself, but through its subvenient base c2. To understand c1 as causally potent we have to conceive it as c2 because in this respect c1 is another name for c2, though we know that c1 behaves anomalously since T1 is irreducible to T2. As a consequence, c1 has a specific feature picked out by T1, but this makes no difference in the world. Hence, I conclude that, though Davidson’s irreducibility argument is true, it is not, however, an argument for mental realism.

1.2. Multiple realization argument

In order to demonstrate the irreducibility of mental explanation, some philosophers argue as follows:

1. A mental state M is a function which is physically realized.
2. There are many physical states P1, P2 ..., Pn that can realize M.

---

11 This means that, according to their conception, the counterfactual analysis is just a way of description which points to no ontological difference.
3. There is no necessary connection (or identity relation) between M and Pi (from 1. and 2.).

4. There are no bridge laws to express the features of M in terms of the physical features of the realizing basis Pi (from 3.).

5. Reduction of M to Pi fails (from 4.).

A mental state can be functionalized inasmuch as we find a causal role (a kind of operation input-output) played within the whole system of mental states. This function can be realized by neural, mineral or other physical instances. For example, intelligent behavior can be realized with silicium processors, as we know from our computers. The first two steps are accepted by all adepts of functionalism; the ardent discussions begin with the transition from 1 and 2 to 3, 4 and 5. The above reasoning is proposed by non-reductive functionalists like Fodor (1974) and Putnam (1975). For Fodor, a bridge law must have a biconditional form such as

\[ P_x \leftrightarrow M_x \]

expressing an identity relation that means that all x satisfying P must also satisfy M, otherwise we could have the situation that not all M are realized by P and this would be the failure of reduction (Fodor, 1974, pp. 52 ff.). Therefore, for the purpose of the unity of science to be realized, it is necessary that these bridge laws are "true empirical generalizations" (ibidem, p. 59), that is, for example, a psychological explanation M is completely translatable in Pi laws. But Pi is a disjunction with an infinite set of terms, impossible to be the object of a lawlike sentence. Hence, it follows the steps 4. and 5. If M and P might be correlated in a sentence, this would express a contingent fact. From the ontological point of view, Fodor endorses the token physicalism while he denies the type physicalism (the claim that all properties are physical properties): "Token physicalism does not entail type physicalism because the contingent identity of a pair of events presumably does not guarantee the identity of the properties whose instantiation constitutes the events (..." (ibidem, 53). This kind of physicalism does permit him to assert that a mental event has a mental property, but falls also under a physical event (that realizes the mental event) by having also physical properties. For example, an economical event, as the money exchange, has physical properties (the money are made from paper or metal), but these properties are not enough to explain that event. So, all events have physical properties, but not all events are explained through physical properties.

Judging the transition from 1 and 2 to 3-5 Kim criticizes this dualism of properties which, in his opinion, meets the same difficulties as
emergentism, that is, it must make comprehensible how new properties are causally potent, being given the causal closure of the physical (Kim, 1992, p. 133). His objections against this non-reductive physicalism sound as follows:

a. If P is heterogeneous, then its realization effect, M, is heterogeneous too. Consequently, the argument that should have backed the independence (e.g. irreducibility) of the special sciences, appear to make them impossible because M also cannot be a scientific kind (Kim, 1993, p. 324). At first sight, this objection, I think, can be refuted since M is a function which remains the same for any physical substrate. But this independence could lead to an unwelcome dualism which Kim takes into account further.

b. If M has a physical realization basis P, then P must be at least sufficient for the occurrence of M (ibidem, pp. 273-275).

c. In addition to b, it is not necessary that M depends universally upon P in order to accomplish the reductive task. It is true that finding an universal correlate of, for instance, pain may be a difficult task because pain is differently realized in humans and other animal species, but a local dependence S → (P ↔ M) (where S represents species' constraints) it is enough to realize the purpose of reduction (ibidem).

d. The most important objection considers the commitment of non-reductive physicalism to downward causation. A realistic and irreducible theory of mind, like Fodor's theory, asserts that mental properties are causally potent. The irreducibility feature requires that this potency is new in comparison to the potency of the underlying physical basis. The core of the critique is that a mental state M₁ is able to cause another mental state M₂, if it simultaneously causes the underlying basis of M₂, respectively P₂ (Kim, 1992, pp. 135-136). But P₂ is a physical fact, and in accordance with the causal closure principle, P₂ is sufficiently caused by another physical fact, say, P₁. The dilemma that we encounter is: either we give up the causal closure assumption and accept that the mental events have new causal powers, or we hold further this assumption and give up the causal autonomy of mental events. In our materialist era the first horn of dilemma is implicitly rejected and thereby it is argued in favor of the other option, a

---

12 M₁ must cause first P₂ in order to cause M₂ because it is supposed (and Fodor and other adepts of functionalism agree with that) that mental properties supervene on physical properties and this means that a mental property does not make any difference where the subvenient basis makes no difference.
reductive account of mental causation.\textsuperscript{13} It is clear that, in a direct confrontation between any non-reductive attempt and the causal closure principle, this principle will always win the game.

The claim that the mental causation can be postulated only through the reduction to its physical basis is a bad news for the adepts of mental realism. Our question is whether this is all what it can be said about the reality of mental properties. I think the answer is no. Moreover, we can see that Kim's approach is also committed to epiphenomenalism. It is uncertain whether Kim wanted this result, but it is sure that this is a normal consequence of his reductive approach.\textsuperscript{14} He uses the same strategy to functionalize the mental properties, but instead of looking to the function itself, he focuses his attention on the physical realizers. Actually his strategy of identifying a mental property $M_i$ with a physical property $P_i$ resumes to these steps (Kim, 2005, p. 111):

"$x$ has $P_i$ at $t$.
$P_i$ satisfies causal role $C$ (in systems like $x$).
Having $M = \text{def} \text{ having some property satisfying causal role } C$.
Therefore, $x$ has $M$ at $t$"

On the same page, as he tries to defend a non-problematical concept of mental causation, he adds that in the third line, where he defines $M$, he does not describe a reality; he just intends to offer a conceptual definition. His emphasis means that $M$ is another name used to express the causal role played by $P$. In this case, it is normal to suppose that the intentional or psychological explanation can be rewritten with physical terms because there is nothing special concerning mental events that could preclude this translation. For instance, given the property of being mammal and a functionalization of it may sound so: "vertebrate animals characterized by milk-producing mammary glands..." The property of being a mammal is defined through the causal role played by a physical item, that of being

\textsuperscript{13} Kim's point is synthesized with the following words: "If mentality is to have a causal influence in the physical domain – in fact if it is to have any causal efficacy at all – it must be physically reducible". Kim 2005, p.161. We observe that this is the same conclusion as in the case of anomalous monism.

\textsuperscript{14} He distinguishes his reductive explanation from reduction. "Explaining a phenomenon seems in no way to damage or diminish its ontological status, whereas reducing it strikes us as undermining its standing as a genuine existent" (Kim, 2005, p. 95). But I have seen nothing in his books so as to make me affirm that for him mental properties stay as a "genuine existent". Ontologically speaking, there is no sounding difference to be attributed to mental phenomena.
vertebrate with mammary glands. In an explanation about some animals, we may use the definition "vertebrate animals characterized by milk-producing mammary glands..." instead of mammals without any change to the meaning of the new explanation in comparison to the initial explanation because both "mammals" and "vertebrate animals characterized by milk-producing mammary glands..." refer to the same property. If the analogy is correct, then Kim's definition of Mi must be replaceable with Pi in the same way.

But we do not see any sense to fulfill this replacement. I am not going to mention the failure of the extensional replacement and hence the intensional feature of mental representations. I prefer a test with more ontological significance. Given the definition of "M" as "having property P satisfying the causal role C at t", and replacing M with this definition in mentalist explanations, we obtain something like that: "John's firing neurons P playing the role of processing some input x to obtain an output y (causal role C) at t, caused him to go to the bus station." This translation deprives mental events of any causal potency; his intention qua intention makes no difference in determining the behavior. I will discuss this translation in more details in section 2.

We have seen that a confrontation with the causal closure principle leads to the unwelcome result that the mental properties have no causal influence which is equivalent with saying that they are epiphenomenal. Davidson wanted to defend the autonomy of the mental vocabulary without giving a corresponding ontological autonomy of its reference. Fodor, Putnam and other functionalists and emergentists wanted to defend the irreducibility of the mentalist explanation grounding it on the difference between properties. But this dualism of properties violates the constraints imposed by causal closure principle. Must we accept this? Or is this claim: "if it isn’t literally true that my wanting is causally responsible for my reaching, and my itching is causally responsible for my scratching (...), then practically everything I believe about anything is false and it’s the end of the world" our single defense? From the non-reductive physicalism we retain, however, this premise that intentionalist explanation is irreducible. But we have the feeling that this premise alone is incomplete as an argument in favor of mental realism and that we need something more to make it intelligible. In order to avoid the epiphenomenalist objection,

---

irreducibility argument must be backed by a realist argument regarding mental properties. But how can we get this argument?

We have to look for another strategy in order to find this argument. We may use even this irreducibility as an argument to assert the reality of mental states. Hitherto the strategy was to start from the property itself (anomaly, functional character) and hence to try to explain how it is causally potent and, after looking for the substantial basis of this substance, to assert that the single form of mental causation is physical. We start right from the presence of mental causation itself in order to assert the reality of the mental properties. In this sense we may find a way out of Kim’s dilemma by breaking in a specific sense the causal closure principle. This could be made by proving that mental causation exists as a distinct kind.

2. A transcendental argument in favor of mental realism
   2.1. Some general considerations

There are some observations to be made before we deal with the specific function of this argument in the domain of mental realism. In general, a transcendental argument has a form like this: "For Y to be possible, X must be the case" (Stern, 2000, p. 6). Y may be a fact about us and our capacities and X is the transcendental condition that enables the working of Y. A well known example about such an argument is the argument for the existence of an external world based on our knowledge: in order for knowledge to exist, there must be an external world.

This type of argument for the reality of something has been strongly criticized, especially by stressing the point that it develops itself, however, on an idealist foundation. The assertion of the reality of an external world is subject-based affirmation because it is derived from the system of our knowledge. This is a sufficient condition for a skeptic to emphasize that the gap between what we know and what exists still remains. To this challenge, the best answer is to conceive the transcendental condition as a basic assumption, though we are aware that this is not a strong argument in favor of the transcendental argument that could determine a skeptic to change his position. But compared to the skeptical attitude, this is the best what we can hope to obtain. This kind of idealism is inoffensive as long as it makes the existence of the world a condition for the intelligibility of knowledge and does not try to limit this external world to what we know. This argument does not contain an epistemological reference to how the world is, if it is what we know, if it is subject independent or constructed, and so on. It
simply asserts that, if there were not an external world, there would not be knowledge.

In addition to this, it may be stressed that, being a transcendental condition, it does not make any sense to look for further arguments to back it. Its validity will result from what we take as granted. For example, in the case of Moore's argument for the external world, we have: 1. "I have two hands" and from 1. we derive 2. "The external world exists". But in order to assert 1. and to prove its truth we have to find an external object as its truth condition and that means that this object exists. But this represents even the conclusion, 2, which is to be proved. Therefore, 2 is an intelligibility condition for 1, not its consequence. The existence of the external world is not a truth condition like any other because the very idea of correspondence presupposes the external world. This case expresses in a clear way how a transcendental argument works, its necessity emerges from what we need to accept.

This is also the case with the Cartesian argument. The existence of the knowing subject is derived from the fact that the subject thinks, and thinking cannot be denied because denying thinking makes impossible the action of proving itself. We have to remark, however, that in both cases the existence of the external world and that of the knowing subject are derived from the existence of knowledge and that of doubting, though they are understood as existence conditions for knowledge and knowing subject. There is no circularity here because the arguments for the existence of knowledge and that of thinking are not derived from the existence of external world, or that of a subject. We accept, first, that knowledge and thinking exists and afterwards we are constrained to accept that knowledge and thinking cannot exist, if the external world or the subject does not exist. The arguments for the existence of knowledge, for example, are independent from the conviction concerning the existence of the external world. If a skeptic insists that there is no knowledge, like Plato's Gorgias, we cannot defend our realist position by pointing out the existence of the

---

16 We find these arguments in John Searle, 1995, pp. 178 ff. To this point he said (p. 178): "One can show that this or that claim corresponds or fails to correspond to how things really are in the 'external world', but one cannot in the way show that the claim that there is an external world corresponds to how things are in external world, because any question of corresponding or failing to correspond to the external world already presupposes the existence of an external world to which the claim corresponds or fails to correspond. External realism is thus not a thesis nor a hypothesis but the condition of having certain sorts of theses or hypotheses."
external world because the skeptic might further deny the existence of the external world (and he really did so). In other words, knowledge and thinking are the means by which we assert the existence of the external world and that of the knowing subject (in the order of knowledge), but the external world and the knowing subject represent the conditions for the existence of knowledge and thinking (in the order of being).

2.2. A transcendental argument for the existence of mental states

Let us see how a similar argument could be constructed in the domain of mental properties. We need, first, to identify that form of discourse whose existence, as in the case of knowledge or thinking, is evident by itself and, second, to prove that this existence cannot be conceived without the existence of mental properties. Returning to our generic definition, Y is represented by mentalist explanation and X is represented by the existence of mental states. We will try to demonstrate that X is a necessary condition for Y. We start with the following reasoning:

TA 1. There are irreducible intentionalist explanations.
TA 2. If the things picked out in these explanations were not real, then these explanations would not be irreducible.
Therefore, TA 3. Mental properties are real.

My aim is to prove the validity of TA 2 which is the core of this transcendental argument. The first premise is frequently discussed and I think that there are serious reasons to accept it, so that I will not insist too much upon it. As we have seen in section 1, this premise is backed by different arguments emphasizing the categorial difference between two kinds of discourse. This premise remains valid even if the critique from the causal closure principle has denied the conclusions inferred from it. Why do we think that the mentalist explanation is irreducible? It is, ultimately, a matter of bone sense, to accept the truth of the following assertion: "But what is interesting about monetary exchanges is surely not their commonalities under physical description. A natural kind, like monetary exchange, could turn out to be co-extensive with a physical natural kind, but if it did, that would be an accident on a cosmic scale." (Fodor, 1974, p. 56). Intuitively, we cannot hope that, controlling the physical laws, we will be able to predict how the European currency will evolve in comparison to the American dollar.
A materialist may resort to the notion of supervenience and argue, however, that the economical processes cannot make any distinction where the physical foundation doesn't make any distinction. This is an accepted idea of supervenience. Davidson argued that this notion of supervenience does not impede the irreducibility of the supervening concept. For example, he said, the semantic concept of truth cannot distinguish any sentence, not distinguishable in syntactical terms, though the truth is not defined in syntactical terms. (Davidson, 1993, p. 5). But a conceptual autonomy in the case of mental properties amounts to nothing if it is not accompanied by an ontological independence because it was not so difficult for a reductivist like Kim to reject Davidson's argument for mental realism.

Actually, we may ask ourselves to what extent this notion of supervenience is justified to assert the reducibility of the higher-levels to the physical realizers. If the economical or the mental processes are unpredictable from the physical point of view, and we have to agree to this, then there is no necessary connection between a physical event and an economical or psychological event. Relative to Fodor's example, Lynne Rudder Baker adds that we can imagine a world where the physical realizers (papers, metals, electronic chips) of an investment exist without being accompanied by the occurrence of the investment itself (Rudder Baker, 1995, p.131).17 To this, Kim would claim that it is enough to assert a local determination because a physical realizer of the investment is at least sufficient for its occurrence. In my opinion, the physical realizer is a condition for the occurrence of the investment, but not the most important. An investment, to use Searle's terminology, is a social fact and as such it depends on social rules, governments, ownership and – the most important condition – a person with the intention to invest. These are necessary conditions for the occurrence of the investment because we cannot imagine a world with investment, but without rules and persons who make investments. And our predictiveness derives from the knowledge of those facts, not from the underlying physical basis. Therefore, either the strong notion of supervenience – which, on the basis of the axiom “there is no economical or psychological difference without a physical difference”, asserts the reducibility thesis – is false because there is no necessary connection between the physical realizers and their higher-order properties, or the weak notion – based on the same axiom, but with even the

17 Rudder Baker, 1995, p. 131. She develops a convincing argument to show that supervenience is false. See also pp. 126-144.
postulation of irreducibility (ex. Davidson) – is inoffensive for our argumentation because this axiom, we will see below, does not affect the postulation of mental realism.

Considering TA 2, it is normal to ask: why does not the mentalist explanation work irreducibly, if its picked out items are not real? It can be answered by rejecting the possible objections to this premise. But first, there are still some clarifications to be made about the fact that an explanation requires specific facts as truth condition according to its vocabulary. Let us take the claims $R(r)$ and $R'(r')$. They refer to the same reality, but $R(r)$ is a description of the macromolecular aspects, while $R'(r')$ of its micromolecular structure. We may express linguistically these claims, for example, as follows:

\begin{align*}
R(r): & \text{ John killed Bill with a knife} \\
& \text{and} \\
R'(r'): & \text{ A system of neurons n and a set of other cells c and a bulk of metal atoms} \\
& \text{provoked a mortal damage to another system of neurons and cells.}
\end{align*}

We want to know what happened with that reality. We use different lexical choices according to our interests. If we are fond of neurobiological explanations, we could choose $R'(r')$ to describe that reality, but we will find out information about neurons and other molecules and not about persons like John and Bill. If we want to find out the cause of Bill's death, we have to choose $R(r)$. We cannot describe Bill's death by means of $R'(r')$ because, as we have seen also in the case of investment, Bill's death is a different event than a damage to a system of neurons and cells and as such it belongs to another type of discourse. According to Quinean philosophy, $R(r)$ does not alone confront the reality, it comes together with many other similar facts such as persons and knifes. The choice between $R(r)$ and $R'(r')$ is not just a choice between two propositions, it is between two theories, between two types of facts. Hence, $R(r)$ represents a better answer to our question than $R'(r')$. The claim R: "Bill's death was caused by John" need as truth condition the fact $r$ – John killed Bill with a knife - and not $r'$ - a chunk of neurons and other cells. It could be true that these neurons are implied in Bill's death, but they do not satisfy the sentence R because R's terms do not pick out such a fact as $r'$. It could be false that John killed Bill, the murderer might have been Paul, but this does not change the result because the intelligibility condition is not damaged. Therefore $r$ is the necessary kind of
truth condition for R. In other words, in order to find out the murderer, we do not look for neurons and other cells. This seems to be a trivial fact, but nevertheless the reductive explanations try to explain behavior in neural terms.  

There are domains where the reductive explanation works. For instance, we can explain the boiling of water by using descriptions about kinematical movement of the molecules of O and H₂. The same holds if we want to explain how a T.V. functions. For this kind of properties the supervenience holds strongly. This structure H₂O has necessarily the same features (transparency, colorlessness, colorlessness) in all possible worlds. Producing this relation H₂-O we will obtain also the properties of being transparent, odorless and colorless. Such supervenience does not hold for all properties. Some properties cannot be predicted from the knowledge of their physical basis. Producing in another world the subvenient basis of an intention to invest, it is not necessary to obtain that investment. The physical explanation does not tell the whole story. There is something left out. In order to depict this difference we use another vocabulary. But our aim is to know what makes it inevitable. I have to argue further that this constraint exerted upon us to use the intentionalist vocabulary is due to the real existence of mental states.

2.3. Objections

Consider now some objections: the claim that this constraint is based on habit or the claim that a good functioning of an explanation may lay on pragmatic reasons or even the claim that the irreducibility does not imply logically that mind has new causal powers in comparison to the physical counterpart. To all of these, it may be replied that this constraint is too strong to be out of habit or pragmatic calculus and that mind pass the above proposed causal test.

a. Habit objection

One can raise the objection that this irreducibility is not based on real properties, but on our inherited habit to use mental terms. This kind of critique may be included in the eliminativist category of reductive approaches. Rorty gives us a fictive example of a far off planet where the

---

18 See, for example, Gerhardt Roth, 2003.
inhabitants (Antipodeans) are accustomed to use a neurobiological vocabulary corresponding exactly and in all details to our terrestrial psychological terms (Rorty, 1979, pp.70-98). As he says, the Antipodeans pick out the same facts as we do with their neurobiological terminology, though they do not know what a mental property means. The example could be interpreted in two ways: 1. the Antipodeans uses a different language, but this does not mean that their terms do not refer to the same mental states as our terms; 2. this example shows the non-existence of such things as mental properties and the possibility of dispensing with mentalist vocabulary. Rorty chooses the second eliminativist variant.

Let us say that an Antipodean asserts:

(A): "John is going to the bus station because his firing neurons in C100 area which point to his mother and neurons in C120 representing a strong impulse to see intimate persons determined him to visit her."

The expression "firing neurons in C100 area" depicts just a biological excitation. The property of pointing to John's mother is not contained in it or, at least, does not seem to be picked out by a biological vocabulary. This vocabulary will describe the neurons in C100 in a third person perspective by invoking shape, numbers, electricity, etc, and it will not be able to say that they point to John's mother because this property is not accessible from the third person perspective, it depends on a certain subject. It is, as Kripke said, like the situation when God decided to create the heat. It was not enough for Him to create just the molecular motion. He had also "to create some sentient beings to ensure that the molecular motion produces the sensation S in them" (Kripke, 1972, p. 340). Kripke's idea is that both designators (heat and molecular motion) are not identical because we can imagine one in the absence of the other. His distinction also concerns the non-identity between the heat property and that of molecular motion because the non-identity between rigid designators implies also a difference between things named by those terms. Though the final remarks in Naming and Necessity concern the identity theory, we may use some results from there because eliminativism and identity theory from a certain point go hand in hand, so that the critique against one can be used also against the other. The eliminativist argues that we can dispense with the intentionalist vocabulary using the same argument as a representative of the
Cristinel UNGUREANU

identity theory, namely, that exists a single kind of properties, e.g. the neural properties. 19

Lycan wants to loose this conclusion by showing, in turn, that "pain" and "fiber stimulation" are not rigid designators and thereby to claim a local identity (Lycan, 1974, p. 677-689). On his account, "pain" is not a rigid designator because it may pick out either the phenomenon of pain itself or just the impression of pain. It is supposed thereby that we can have the impression of a pain without having the pain itself. I think that the discussion about heat or pain is a little tricky because they depend too much on the external stimuli. In accordance to this, it can be distinguished many degrees of pain, from the very hard sensation to the merely false impression. The very hard sensation may be understood as a chemical process and in naturalistic terms it is another item than a merely impression of pain. Even if Lycan is or is not right, the concept of pain is problematic, while using a propositional attitude may clarify better Kripke's idea. A belief and the impression of having a belief are the same. To feel that I believe that today is Sunday means to have that belief. Consequently, Lycan's critique fails, when it is applied to proper mental concepts.

Relative to Lycan's denial of the rigidity of the "fiber stimulation" by means of multiple realizability argument, I think this does not affect the identity of firing neurons with fiber stimulation. In this case, supervenience works strongly: the property of being neurons in C100 area is equivalent to fiber stimulation. In other example: the property of being electric current is identical with the movement of electrons. The multiple realizability argument does not matter here because each individual physical fact is identical with itself. It is absurd to say that a disjunction (neurons, silicon, etc.) realizes the property of being fiber stimulation. Disjunctions exist only as abstracts, but this stimulation occurred in a concrete event. One may reply that that could be a contradiction with a previous claim, where we have endorsed the multiple realizability argument against the identity theory. But this is exactly Kripke's idea that a particular event like fiber stimulation is not the same as a pain. An isolated chunk of neurons similar with neurons in C100 area have in all possible worlds the property of being fiber stimulation, whereas the property of being pain does not exist in a world where there are no sensitive beings. In the first case we have a strong supervenience of property of being fiber stimulation upon that chunk of

19 See more on this topic in Ungureanu 2008.
neurons; the later case shows us a weak supervenience which cannot underlie the identity theory.

b. Pragmatic objection

We have the explanation E which picks out the cause c and we know that the real cause may be c₁, but we still refer to c because of its good consequences. For example, we may accept that God is the Creator of human justice (He gave Moses the Book of Law) instead of believing that they have the source in our need for social security because the former variant may set a stronger constraint upon us. In this sense, it may be argued that intentionalist explanation is preferred on pragmatic grounds, it gives us the tools (terms as desire, hope, will, etc.) to understand each other, but it could be said further that there are no such things as desire, hope, will, etc. Daniel Dennet is one of the upholders of this instrumentalist view. According to him, the ascription of mental states is made through the intentional stance, that is, as if the subject were a rational being (Dennet, 1987, p. 17). The mental states must be instrumentally interpreted because the ascriptions are never perfect, that is, there is no perfectly rational subject.

As a general remark, it is argued that the good consequences of a theory are usually its side effects, not its purpose. Its purpose is to establish how things stay together. A pragmatist may argue, however, that the good working of a theory is not dependent on its truth and that we can achieve good consequences even from its falsehood and, consequently, it is not necessary for an intentionalist explanation to pick out real items with its terms. Though this remark is true, it is not thereby proved that the intentionalist explanation is backed by a pragmatic perspective. Moreover,

---

20 We remember the well known James' reasoning that faith may be pragmatically justified. See, for example, James, 1907, p.107.
21 We never know exactly "what intentional systems in question really believed". Ibidem, p. 40. This epistemological point is completed with the ontological claim that only patterns are real, ascriptions are just points of view chosen to understand those patterns. In evolutionary respect, there are differences between patterns, but there is no such a pattern to support a perfect rational ascription. This is equivalent with the claim that rationality as such does not correspond to facts. Fodor finds this claim as being, at least, obscure. See Fodor, 1985, pp. 79-80
22 I adhere without other arguments to the thesis that the pragmatism may provide a good criterion to identify the truth, but in no case may constitute its definition.
as Fodor has shown\textsuperscript{23}, there is a problem with Dennett’s instrumentalist view. The intentional ascriptions are guided by the principle of rationality and the systems are thereby more or less far from this principle. To question how the systems are in some degree rational, imply, however, to look at the structure itself. These differences are supported by the structure of the system itself; otherwise the ascription would be aleatory. In order to say that a system is more rational than another so that its actions are better described through the intentional stance, than another system, which can be better described through the design stance (or even the physical stance), it is necessary that the system itself really has those properties which enable the ascriptions. It follows that rationality is not just perspective.

It is difficult to assume the thesis that the intentionalist vocabulary is based on such a pragmatic argument. I think that the constraint is stronger than a deliberately decision. The utility of intentionalist explanation is not like the case when we choose the gravitational explanation instead of the Aristotelian physics to describe the fall of a stone. Any choice between sorts of explanations entail an intentional perspective. We cannot step outside the intentional frames to judge what kind of explanation brings better consequences. As in the case of the transcendental argument for the existence of the external world, any choice presupposes already intentionality as its intelligibility condition.

c. The causal test

Someone may say: Ok, this constraint expressed in TA2 is not out of habit or pragmatic calculus, but still it does not follow that, what makes this kind of explanation irreplaceable, are real intentional properties. Where are we now? We found out that this constraint is based on an intelligibility condition. Extending Kripkean discovery to a whole language, we hope to reach the same result of denying reducibility. In the Antipodean claim (A), we remember, the property of neurons pointing to John's mother is not to be described biologically. This part of the expression "which point to his mother" picks out the intentionality of these neurons and as such it no longer belongs to a naturalist vocabulary.\textsuperscript{24} If we accept that pointing-to-

\textsuperscript{23} See Fodor, 1985, p. 80 and also Clark, 2001, p. 58
\textsuperscript{24} An eliminativist may reply here that the firing of neurons causes other firings of neurons as extensional facts, not as being about something. But we don't have to forget that we
mother is an intentional fact, then we have to accept also its connection with other intentional facts (like the belief that his mother lives in the other side of the city or that for public transport he needs tickets, and so on). To explain all this we have to admit the holism of the mental states. In order to think a proposition, one has to think many other propositions. Consequently, for each intentional fact we are constraint to use a whole vocabulary. Importing, like our Antipodeans in the statement (A), one single intentionalist expression, we are constrained to import a whole intentionalist vocabulary. But a naturalist vocabulary with another intentionalist vocabulary within its limits is no longer a naturalist vocabulary.

Second, the distinction between languages entails also a distinction between facts. Subduing the Antipodean claim to CT, we will see a change in the meaning of cause of behavior. John's walking to the bus station because of his firing neurons means a mechanical determination which is far from what we suppose that was happening in that case.\(^{25}\) John obeys an intentional fact (his desire) and as such he might have turned back or he might have chosen in the last moment to visit a friend. In a word, if we describe what caused his behavior by appeal to mechanical determination, saying that the simple sample of neurons in C100 caused his behavior, then we will transform John in an automaton. If the firing neurons in C100 caused his behavior, but as pointing to his mother, then it is due to this difference that is responsible for what happened with John. The relevant fact is not the firing of neurons as such (we can imagine a world in which these neurons cause no effect), but this difference picked out with the intentionalist vocabulary. The counterfactual expression sounds as follows:

\(\text{(A')}: \text{"If the firing neurons in C100 area did not point to his mother, John would not have gone to the bus station."}\)

It is this pointing to which is responsible for John’s behavior. Some explanations, like this one, are constrained to refer to mental facts; without this reference it would be unintelligible. The transcendental argument will claim here that, if these facts were not real, then this explanation could explain an intelligent behaviour and thus we need, at least, to know why some neurons begin suddenly to fire.

\(^{25}\) Neuronal determination is, after all, a mechanical determination, like the circulation of blood or the functioning of the liver.
dispense with those references. There is no other constraint except the realist argument which might prevent the failure of irreducibility. In a nutshell, mental states are real because they cause the behavior in a specific way.\textsuperscript{26} Or, they are real because they are the transcendental condition for the intentionalist explanation.

How can be this compatible with the causal closure principle? Does it still remain valid this closure in a world with mental causation? There are some suggestions in the next section.

Chapter 3. Mental causation and metaphysics

Consider the following reasoning:

3.1 Each physical event is sufficiently caused by other physical events (the causal closure principle).
3.2 Mental events cause distinctly from physical events (mental realism).
3.3 Therefore, mental causation cannot affect physical events (epiphenomenalism).

First premise is frequently called the principle of exclusion and is based on the intuition that no immaterial facts can affect the physical events. Second premise represents a pressing requirement adopted in order to provide a realist approach to mental properties. The conclusion concerns what was the critique against non-reductive physicalism, namely, the "epiphenomenalist fear" (Fodor): if mind is to be causally potent, it cannot affect the physical domain. We recall, Kim argued that there is a single way out of this dilemma, to throw away the second premise and to accept that mental causation exists just as physical causation. The transcendental argument constrains us to recognize the mental causation as a new kind comparatively to the physical causation in order to make sense of irreducibility. Agreeing to this, our problem is no more epistemological because we know that we cannot avoid mental causation. It is, we see, a problem of our ontology reflected in the exclusion argument. As a rational measure, we have to ask whether it is justified.

\textsuperscript{26} We recall, a property is real, if it contributes in a specific way to the causal potency of that event in which it is instantiated.
A starting point would be to raise the question: in what sense is causality limited to the physical domain?27 According to a reductivist view, an event is causally potent through its physical microstructure: psychology is enabled by neurobiology, neurobiology is enabled by biochemistry...and finally we reach, probably, the ground of quantum physics.28 The consequence will be, therefore, that only the microphysical particles are properly causally potent, an absurd situation that makes all other kinds of causation to be epiphenomenal.29

The causal closure principle is based on the metaphysical thesis concerning the monism of the physical substance: there is only a substance and only a form of causality derived from it. But monism of substance does not entail a monism of causality. It is a trivial fact that all facts are, ultimately, physical things, but it is not trivial (and I think is false) to assert that there is only a single form of causal potency. Though substance and causality are usually taken together, they must be treated separately, unless we will have to accept the absurd consequence of reductionism. What I want to emphasize is that we have to give up our substantialist understanding of causality.30 This means that it is no compelling argument to limit causality to strict and necessary determination. The cause does not need new substances in order to be efficacious. The causal relation needs, roughly speaking, to be backed by a distinct property. This could be a difference of strategy: instead of looking for the substantial features that can support the individualization of a property, we look for what that supposed property does. The former situation compels us to go deep into microphysical features and also brings along the general epiphenomenalist consequence while the later permits us to make sense from our knowledge reflected in the special sciences. Our transcendental argument showed that there are at least two ways of causing (intentional and non-intentional). I am not ready to apply it to all other types of causation, but this gap is enough to break the causal closure. If the conclusion of applying the causal test to mental properties is valid, then we have to admit a dualism of properties. In a substantialist sense, of course that all causes are, ultimately, physical, but

27 A similar question was raised by Lynne Rudder Baker, 1993, p. 79: where are the limits of causal closure?
28 Roger Penrose has tried to explain consciousness by means of quantum physics.
29 This critique is argued in many ways by Fodor, 1989, pp. 62-64; Ruder Baker, 1993, p. 79; Humphreys, 1997, pp. 3-4; Van Gulick, 1993, pp. 249-250.
30 I use the term substance in a Spinozist sense in which the cause necessarily determines the effect.
in a world with multiple properties each cause acts in conformity with that specific property.

The assumption behind Kim’s downward causation argument is exactly the extension of the monism of substance to causality. This problem does not exist anymore in a world with pluralism of causes. However, the question raised by Kim regarding how mental and physical causation stay together is still interesting. First, the question must be solved empirically. A cognitive psychology can research how an intentional fact is instantiated, how it determines other states with its occurrence, etc. Second, if state $S_1$ causes another state $S_2$ and both instantiate the same kind of property, there is no constraint to appeal to downward causation in order to make intelligible how $S_1$ caused $S_2$. Third, a mental state or event is physically realized and its causal influence implies neural processes, but these processes cause because they have this feature of being intentional states. I could say that their firings are guided by the intentional property, but this may lead to substantial dualism. An intentional property does not exist separately from neurons, it consists of their specific structure, but causal difference is made by this property not by the neurons as such.

REFERENCES


31 See an argument in this sense in Humphreys, 1997, pp. 12-13

Cristinel UNGUREANU, PhD Philosophy, Lecturer at the Apollonia University of Iasi.

Address:
Cristinel Ungureanu
Apollonia University of Iasi
Stefan cel Mare si Sfant, nr.2, Iasi
Email: c_ungureanu2003@yahoo.com