

## DUALISM WITHOUT ZOMBIES

Adriana-Nicoleta SORA\*

**Abstract:** The prevalent view in contemporary ontology is a materialist one, in line with the continuous success of natural sciences. Given this context, it might seem impossible to further support a dualist ontology, so long as this kind of position seems not only old fashioned but also in contradiction with the latest achievements of science. In the following essay I will try to dispel this impression as far as the mind-body relation is concerned by exploring the possibility of constructing a mild form of dualism, a dualism of properties, about some mental properties, namely those of conscious experiences. Even though we do not have grounds to further assert the existence of Descartes' ghost in the machine, the immaterial soul, we still have good grounds to support the idea that conscious properties are different from the properties with which natural sciences deal.

**Keywords:** the conceivability argument, mind-body dualism, zombies, property dualism, supervenience

The conceivability argument<sup>1</sup> is one of the most powerful arguments against mind-body identity. It has roughly the following structure: 1. it is conceivable to have a creature physically identical with a conscious creature but without consciousness; we name this creature a zombie. 2. if it is conceivable, it is also metaphysically possible; 3. and if it is possible, materialism is false. The purpose of this essay is to examine the viability of this argument when faced with materialist counter-arguments and the viability of a property dualist point of view.

We generally notice several main ways of attacking the conceivability argument. First, a materialist can challenge the first premise of the argument, showing that a zombie, a creature identical with us from a physical and psychological point of view but lacking consciousness, is inconceivable. Or, he can challenge the second premise of the argument, pretending that the

---

\* Adriana-Nicoleta SORA, University of Bucharest. Email: [adriana.nicoleta.s@gmail.com](mailto:adriana.nicoleta.s@gmail.com)

<sup>1</sup> I will be concerned here with the conceivability argument constructed by David Chalmers in his book *The Conscious Mind* and in a series of subsequent papers.

inference from the conceivability of a zombie to its metaphysical possibility is unsound.

I will examine in turn these counterarguments to the conceivability argument, beginning with the particular one which aims at attacking the inference from conceivability to metaphysical possibility; I will show that none of them is effective; at most, even if we give up the zombie hypothesis, the dualist argument in its shape of property dualism can survive, given an interesting way of defining a physical entity and a physical property. In the final section I will try to give some reasons why it is undesirable to have zombies in our theory.

The essay has the following structure: I. in the first section I present Chalmers' conceivability argument; II. in the second section I defend the argument against the attack on the inference from conceivability to possibility; III. in the third section I explore the idea of conceivability from the perspective of the future of science; IV. in the fourth section I explore a problem with the definition of materialism; V. in the fifth section I will show that the essence of dualism is a special conception of properties; VI. in the sixth section I reconsider the notion of supervenience so that we can give up the idea of a zombie, without giving up dualism; VII. in the seventh section I show why it is undesirable to have zombies in the theory of consciousness.

## I. CHALMERS' CONCEIVABILITY ARGUMENT

Chalmers defines materialism<sup>2</sup> as a supervenience thesis: materialism is true if all the facts logically supervene on the physical facts, or alternatively, if they are implied by the physical facts. So, trying to defeat materialism will involve trying to defeat the thesis of logical supervenience of all facts on the physical facts and properties.

In order to obtain a better understanding of the thesis mentioned before and of the entire construction, we should better clarify some of the concepts involved, especially those of a physical property and of supervenience. The physical properties are represented by the properties included in the theories of basic physics:

„For our purposes, the relevant A-properties are usually the physical properties: more precisely, the fundamental properties that are invoked by a completed theory of physics. Perhaps these will include mass, charge, spatiotemporal position; properties characterizing the distribution of various spatiotemporal fields, the exertion of various forces, and the form of various waves; possibly properties corresponding to the nature of various fundamental particles; and so on. The precise nature of these properties is not important.

<sup>2</sup> I will present the argument such as it appears in Chalmers (1996), in the book „The Conscious Mind“; the other constructions of the same argument in his papers are approximately identical.

If physics changes radically, the relevant class of properties may be quite different from those I mention, but the arguments will go through all the same." (Chalmers 1995, 30)

It is not important either to give a list of the fundamental physical properties or to characterize them further, the only important thing is to give a frame of their way of individuation: in a causal- structural manner. A physical property is one that is characterized in a functional way, by the interaction of the particle that is its bearer with other particles, by what is capable of bringing about in interaction with another entity. This view appears in several places of *The Conscious Mind*:

„For example, it might be that for something to qualify as an electron in a counterfactual world, it is not sufficient that it be causally related to other physical entities in the way that an electron is. Some hidden essence of electronhood might also be required" (Chalmers 1995, 120)

„The strategy to which I am most drawn stems from the observation that physical theory only characterizes its basic entities *relationally*, in terms of their causal and other relations to other entities. Basic particles, for instance, are largely characterized in terms of their propensity to interact with other particles. Their mass and charge is specified, to be sure, but all that a specification of mass ultimately comes to is a propensity to be accelerated in certain ways by forces, and so on. Each entity is characterized by its relation to other entities, and these entities are characterized by their relations to other entities, and so on forever (except, perhaps, for some entities that are characterized by their relation to an observer). The picture of the physical world that this yields is that of a world as giant causal flux, but it tells us nothing at all about what all this causation *relates*. Reference to the proton is fixed as the thing that causes interactions of a certain kind, that combines in certain ways with other entities, and so on; but what is the thing that is doing the causing and combining? As Russell (1927) notes, that is a matter about which physical theory is silent". (Chalmers 1995, 137)

On this view, everything is a composition of some particles having properties that result from their causal interaction. A physical theory is silent about the intrinsic nature of those fundamental entities, about the properties they might have in themselves, independent of their interactions; more than that, it does not even speculate about having such an intrinsic nature.

Another very important notion in Chalmers' conception of materialism is the notion of supervenience, which is understood as a dependence in the variation of two sets of properties, the physical and the supervenient ones; for our case, the physical properties of the body and the mental properties,

or phenomenal properties. Depending on the nature of the relation between the two sets of properties we can outline two kinds of supervenience, logical and natural.

Logical supervenience holds if in every possible world the instantiation of the physical properties is followed by the instantiation of those that supervene on them. For our case, the instantiation of the same physical<sup>3</sup> properties of the brain will give rise to the same mental properties in every possible world. In other words, it is impossible and contradictory to have the physical set of properties instantiated without the mental set.

The aim of a dualist theory concerning the mind-body relation is to show that in this case the relation of logical supervenience does not obtain and one of the arguments against logical supervenience takes the shape of the conceivability argument. So it is sufficient to show that we can conceive a possible world identical from a physical point of view with our world, but where the brain does not give birth to phenomenal properties, to experiences.

The conceivability argument involves three steps: 1. it is conceivable to have a creature physically identical with a conscious creature but without consciousness; we name this creature a zombie. 2. if it is conceivable, it is also metaphysically possible; 3. if it is metaphysically possible, materialism is false.

The construction of the conceivability argument depends also on a particular understanding of the modal notions involved. First of all, possibility and conceivability are understood in almost the same way, as non-contradiction: a statement is logically possible if it is not a priori that its contradictory obtains and a statement is conceivable if it does not involve contradiction. The argument is a complicated one, due to the introduction of a two dimensional semantics for terms involving a primary and a secondary intension, which results in a duality of possibilities and necessities. The clarification of these notions is very important for a better understanding of the conceivability argument.

The primary intension of a term in a world is obtained considering his extension in that possible world taken as actual; determining the primary intension is always an a priori matter and is obtained by reflecting about various scenarios concerning the facts present in the world under scrutiny: what would be the extension of a word if the world were so and so, for example what would be the extension of the term „water“ if the liquid with the same phenomenal properties had another chemical formula? Unlike the primary intension, the secondary one is the extension of a term in a possible world considered as counterfactual, given that the actual world is one way or another.

---

<sup>3</sup> As we have previously seen, Chalmers distinguishes between two aspects of matter: its causal-structural properties which he considers to be the physical properties of matter and the categorical, intrinsic properties which result from the hidden nature of matter. This frame is also applied to the brain, where we can distinguish between physical properties and nonphysical/phenomenal properties, the properties of our subjective experiences.

A sentence is necessary if it is true in all possible worlds, or alternatively, if it is a conceptual truth, considering both the primary and the secondary intensions. The difference appears only from an epistemic point of view, the primary necessary sentence being *a priori*, while the secondary necessary sentence can be known only *a posteriori*. Similarly, a statement is possible if there is a possible world that makes it true, where possible world can be read either as actual or as counterfactual. The remaining notion to be defined is conceivability, so that a statement is conceivable if it is true in a conceivable world; conceivability involves also two species according to primary and secondary intension.

The implications from conceivability to possibility hold only for the same range: primary conceivability implies only primary possibility and secondary conceivability implies only secondary possibility. According to the distinctions analyzed before we can outline also two notions of supervenience, supervenience according to primary intension and supervenience according to secondary intension, the first one being *a priori* and the second one only *a posteriori*.

Chalmers' conceivability argument centers on the first notion of supervenience mentioned earlier, the author considering that it is sufficient to overthrow the materialist thesis. The key point of the argument is that we can observe a sharp distinction between the supervenience of all macroscopic properties on microphysical properties and the supervenience of conscious properties on the properties of the brain. Having in mind the logical supervenience according to primary intension, we can say that it is inconceivable to have a complete microphysical description of the world from which it does not follow the macro-physical description; in other words, a complete microphysical description of the world *a priori* implies a sentence about an event cast in a phenomenal vocabulary. For example, a description of a liquid in terms of molecular chemical structure and properties, say of a sample of H<sub>2</sub>O, *a priori* implies the description of the same liquid in terms of phenomenal properties and common sense vocabulary. Given the condition of truth of the material conditional (implication as used in the previous lines) it is impossible to have a microphysical description of the world without the macro-physical one.

Nevertheless, the conceivability argument can be constructed for secondary intension too, showing the fact that even secondary supervenience does not hold.

## II. THE ATTACK ON THE SECOND PREMISE

We generally notice two main ways of attacking the conceivability argument. First of all a materialist can challenge the first premise of the argument, showing that a zombie, a creature identical with us from a physical and psychological point of view but lacking consciousness, is inconceivable. Or he can

challenge the second premise of the argument, pretending that the reasoning from the conceivability of a zombie to its metaphysical possibility is unsound.

Therefore, some materialist philosophers will try to block the argument from conceivability to possibility, showing that zombie is metaphysically impossible, in spite of its epistemic conceivability. The relationship between consciousness and brain is put on a par with other theoretical identities from science like „Water is H<sub>2</sub>O“. Given the fact that these cases represent cases of a posteriori discoveries, from an a priori point of view they open the possibility of having the same deep physical-chemical substrate without its actual superficial properties or to have the same word paired with another conceptual content (for example H<sub>2</sub>O will not be watery stuff or „water“ will no longer refer to „the clear, colorless, odorless liquid...“ but to something different so that water will not be any longer the clear, colorless, odorless liquid etc, but something different).

I think that the most promising two branches of this kind of materialist attack are the following two<sup>4</sup>: A. one that tries to devoid our concepts of semantic content so that we can construct conceivability arguments for other respectable identities from science and show that in spite of this conceivability, the metaphysical identities are not hindered; B. a second one, which tries to move the conceivability at the level of properties and to show that we can conceive the same physical substrate without its usual superficial properties, but this fact does not lead to dualism.

A. What I want to show in this section is that all these conceptions which try to make the cases of consciousness and of other theoretical identities from science akin are wrong since they rely on a false semantical background for theoretical terms and terms referring to consciousness; this identification of the two cases is impossible since the way of characterizing the semantics of the terms involved is different in the two cases: of ordinary identities from science and of consciousness.<sup>5</sup> Therefore, the strategy which tends to eliminate conceptual content and to show that we can construct the conceivability argument for all theoretical identities is not viable for consciousness.

In these conceivability to possibility arguments we can notice multiple levels of conceptualization: 1. there is a semantic relation between the words and their referents, which for some authors is unmediated by conceptual content<sup>6</sup> and for others is mediated by conceptual content, this conceptual

<sup>4</sup> I think that these are the two main options for a materialist who wants to challenge the second premise of the conceivability argument. The first one was discussed in the philosophical literature by Joseph Levine (especially in his book *Purple Haze*). The second option does not seem to be discussed in the literature.

<sup>5</sup> This idea was already nicely developed by Saul Kripke in his final section of the book *Naming and Necessity*.

<sup>6</sup> Philosophers who adopt the theory of direct reference give up the idea that the reference of a word to its object is mediated by conceptual content; at most, conceptual content (definite descriptions) help us identify the referent when we initiate the process of naming but descriptions are no longer necessary to the application of the name to its referent.



content being given by the superficial properties of the referent; 2. there is the metaphysical relation between the superficial properties of an object and the deep, micro-physical properties of the same object involved in a scientific explanation.

Therefore, let us analyze the case of the identity of heat with molecular motion: first, we have the word „heat“, second, we have the manner of presentation of the referent, which is given by the manner in which the motion of the molecules affects our mind<sup>7</sup> (the subjective sensation of heat), and third we have the thing itself, the motion of molecules. The relations between the first and the second level and the second level and the third are only contingent ones: the motion of molecules could have presented another effect, could have caused another experiential properties at the level of our minds so that the word „heat“ could have been associated with another mode of presentation; more than that, in the case in which we do not associate the word with mode of presentation, the word „heat“ could have named another entity, different from the agitation of molecules or with a different phenomenal appearance or both with different structure and appearance (the word would have nothing in common with the way we apply the word „heat“). Given this case, it is easy to construct scenarios in which the agitation of molecules is not heat, because „heat“ can name a different substance or the agitation of molecules could have other superficial properties. The following schematic representation will perhaps be helpful to imagine the different scenarios:

- i. the word „heat“ refers to the agitation of molecules and this agitation produces the current sensation of heat
- ii. the word „heat“ refers to the agitation of molecules and this agitation does not produce the current sensation, but a different one
- iii. the word „heat“ refers to something different from the agitation of molecules that produce the current sensation of heat
- iv. the word „heat“ refers to something different from the agitation of molecules that does not produce the current sensation associated with the agitation of molecules

On the other hand, for consciousness, the situation is different, so long as the properties of consciousness are essential for its particular states; therefore, the interesting challenge in the case of consciousness would be to show that it is conceivable to have the same neuronal structure with different phenomenal properties or with no phenomenal properties at all and to show that this case does not lead to metaphysical possibility and to denial of identity. Philosophers that attempt to construe the conceivability argument in other manner simply change the subject and in the following lines I will try to analyze such a point of view which changes the subject, Levine's point of view from his book *Purple Haze*.

<sup>7</sup> This example uses the same terms as the Kripkean one in *Naming and Necessity*.

Levine tends to assimilate the cases of zombie and of water, outlining that in both cases we have conceptual possibility without metaphysical possibility. H<sub>2</sub>O without being water is in the same measure conceivable as a creature with brain but lacking consciousness. Meanwhile it is obvious that both cases are metaphysically impossible.

A complete physical or microphysical description of the world will not a priori imply a phenomenal description, contrary to the argument proposed by Chalmers. The key point of this materialist argument is the concept of conceptual possibility: something conceptually possible is something that does not involve contradiction, and so we can have an interpretation which makes true the statement „H<sub>2</sub>O is not water“; it is true so long as we have a possible world where the reference of the two concepts involved is different.

Such an argument involves a semantic presupposition about how the reference of the terms is constituted: the denial of the idea of conceptual content. Between the symbol in the language and the referent, there is no conceptual content associated. This looks like a theory of direct reference in its strongest shape; we give up the idea that the reference of a word is determined via a definite description or a bundle of definite descriptions which best characterize the referent, in favor of a theory of direct reference, where we don't associate the word with any definite description:

„...there has to be some minimal amount of semantic knowledge associated with a concept's mode of presentation, knowledge that will ground a priori judgments... what divides the NE theorists from advocates of the conceivability argument is a general question in the theory of meaning: for most terms, do we have a priori access to sufficient information to determine their referent given a context (a possible world considered as actual)?“ (Levine 2001, 53)

This kind of non-ascriptivism leaves room for a notion of a priori truth understood as formal coherence. We can a priori assess only the truth of the sentences that are formally contradictory or identities, all other sentences being possible to evaluate only a posteriori. So all other sentences are open to conceivable scenarios, possible worlds where the terms involved are not co-referential. For example it is neither a priori that the term „water“ is associated with some qualitative description involving the superficial properties of H<sub>2</sub>O nor a priori that „water“ refers to H<sub>2</sub>O, so that we can easily construct conceivability arguments for water, but without any metaphysical effect: the structure of H<sub>2</sub>O will present the same superficial properties whether or not they are named water and our term „water“ will refer to H<sub>2</sub>O, even though a posteriori.

This kind of argument is vulnerable from two points of view. First, it misrepresents the semantics of phenomenal terms, which refer essentially to some phenomenal properties and not only contingently. Therefore, in a case where we have the same physical structure without the phenomenal



properties we will not have a state of consciousness and it seems very counter-intuitive to pretend otherwise.

For a state of a subject to qualify as a conscious state or as a conscious state of a certain type it is necessary to have certain phenomenal feel. A state of the brain that realizes the experience of seeing red must have the feeling of red. This aspect indicates a great difference between the case of common natural kind concepts and of the phenomenal ones. For some substance to be water it is necessary and sufficient to have the chemical structure of H<sub>2</sub>O, no matter how it feels, whether it is liquid or not, whether it is colorless or not, whether it is tasteless or not. For a state of mind to be a conscious state of pain it is necessary to have the feeling of pain; it does not matter if it has the same physical structure of pain as in the actual world.

The case of water involves a semantically open situation, as we can choose between some alternatives: either to rigidify our concept of water and to say that water in other worlds is only a liquid with the molecular structure of H<sub>2</sub>O, or to consider that we have many kinds of water that have in common only the phenomenal appearance. The necessity of a statement like „Water is H<sub>2</sub>O“ arises from the decision to take the first semantic way. In the case of consciousness we don't have an open space for different semantic decisions: we have the intuition that a conscious state will be identical with our state if it feels like our state, although it will be instantiated in another chemical configuration. This argument shows that the secondary intension of mental terms could not be given by the associated states of the brain.

The second point where this argument can be challenged is that it misrepresents the idea of conceivability involved in the dualist argument. It is not a priori conceivability, it is conceivability supplemented with the a posteriori discovered information and with semantic relations (if we want to make these relations a posteriori). In science, so long as we have the micro-physical description of a phenomenon, its macro-physical and phenomenal description follows as an a priori matter. It is inconceivable to have a liquid with the structure of H<sub>2</sub>O or XYZ (or whatever chemical compound which has the same superficial properties as water: „the clear, odorless, tasteless liquid...“) without being colorless, odorless, tasteless and so on. But when we construct the conceivable scenario, we already know that water is realized in some physical substrate and eventually know the substrate.

Applying the same way of reasoning to consciousness does not give the same results. From the description of the neural correlates of consciousness its phenomenal description does not follow in an a priori manner, so that it is conceivable to have the same physical description without any phenomenal property or the same physical structure with other phenomenal properties. In the case of consciousness, the conceivability argument works both for primary and secondary intensions (conceivability supplemented with a posteriori information about the actual world). When the dualist constructs the arguments, the conceivability included in the first premise can be read

also as conceivability according to the secondary intension of the terms. The possibility of constructing the conceivability argument considering the secondary intension is due to a special feature of the terms referring to the states of consciousness: their primary and secondary intension is the same, unlike the case of most other terms from natural sciences.

It is important to note that in the conceivability argument for consciousness, the central relation is the one between a phenomenal property and set of properties and their physical constitution. In the analogue case of water the relation between H<sub>2</sub>O/XYZ / whatever realizer of water and the superficial / phenomenal properties of this liquid becomes a necessary one once we discover the chemical substrate. Unlike this case, for consciousness, the relation between phenomenal state and realizer does not become a necessary one upon a posteriori discovery.

We can say that the materialist has wasted his time trying to show that even in the case of scientific identities we can conceive without contradiction cases of the base properties without the supervenient ones due to the fact that the identity is discovered only a posteriori. This helps us no longer, because the key of the Kripke-Chalmers conceivability arguments is not here.

B. As we saw earlier in Levine's argument, we can conceive a zombie and H<sub>2</sub>O that is not water because there is no a priori contradiction in the statement describing these situations, although this conceivability does not put into trouble the explanation of watery stuff by the chemical structure associated. Therefore, if we unload concepts of any theoretical content associated and consider a priori non-contradiction sufficient for conceivability, we will have many conceivable situations in science without nevertheless violating a materialist ontology. This is still a case where the same bunch of superficial properties is associated with the same chemical structure, but only under different names.

The discussion in the previous paragraphs reveals something interesting about the conceivability argument: this argument may take two forms, namely a semantic and a stronger, metaphysical one. The first form is the particular one discussed in the previous paragraph in connection to Levine's point of view. The second form is represented by the possibility of constructing scenarios where the same chemical compound leads to different appearances (like H<sub>2</sub>O having the same appearance as in our world in some possible worlds and at the same having other different appearance in other possible worlds: being red instead of being colorless or being solid instead of being liquid). The analogue case for the brain-consciousness relationship will be: the same physical structure of the brain with different experiences from the actual ones.

Let us see the effect of this more radical conceivability argument. On a more careful reading we have to distinguish between different cases: one in which the difference in superficial properties is due to a different setting of the perceivers's mind or to a difference in environment and another, in which the difference is a genuine difference, the other conditions being identical (ways of perceiving, environment).

The first possibility does not affect the idea of logical supervenience in the case of ordinary natural kinds, since we have other possibilities of explaining where the difference in appearance comes from: we can consider that the difference in appearance is due to a different setting of our mind or we can consider that it is due to a different total constitution of the world. Also the case of H<sub>2</sub>O without any phenomenal properties associated could be a product of the interaction between mind and reality: we can consider that there are worlds where the structure of H<sub>2</sub>O is not in the visible spectrum.

It remains only one option for the materialist who wants to argue against the conceivability argument by attacking its second premise: to argue that the same chemical formula can be associated with different appearances (and this fact is not due to a different setting of the mind or to a different environment) and that, in spite of the fact that the connection between appearances and deep structure is a contingent one and logical supervenience no longer holds, materialism is not threatened. From a scientific point of view, the cases outlined at the beginning of the paragraph seem absurd because we have some strong intuition that a chemical substance cannot be associated with different appearances. So, the relation between the chemical and superficial properties of a substance is a necessary one (presupposing that the environmental conditions are the same). If we open the possibility to pair one and the same chemical formula with different appearances we seem to open the way to dualism for cases which we treat as identities in science: if we allow the possibility of H<sub>2</sub>O with another appearance, where this appearance is not due to a different environment or to a different setting of the perceiver's mind, we are forced to acknowledge that the chemical structure does not account for its phenomenal appearance.

Therefore, given the fact that the last way of constructing an attack on the conceivability argument is absurd, and that the resources of attacking the second premise are exhausted, the only remaining way is to show that the first premise is false, that zombie is inconceivable.

### III. THE INCONCEIVABILITY OF ZOMBIES

The terminology used in constructing the conceivability argument is a little bit misleading at first sight. We saw that we can distinguish between the primary intension and the secondary intension of words; consequently, we have two notions of supervenience: supervenience according to primary intension and according to secondary intension (supervenience in the actual world). I think that the second form of the argument, according to secondary intension is the interesting one for the case at hand: given all the empirical discoveries about the brain, it is conceivable to have the same physical structure of the brain without consciousness.

A conceivable scenario from the perspective of the secondary intension is one that could be constructed a priori without contradiction, but at the same

time in the space left open by all a posteriori constraints given by the knowledge of our world. . The majority of our concepts are derived from a posteriori knowledge; therefore, upon discovery, a sentence like „Wate is not H<sub>2</sub>O“ becomes a contradiction since both concepts involved in this sentence refer to the same object. On the other hand, if we look at the same sentence from the perspective of primary intension, it is not contradictory since we can have a possible situation in which the two concepts do not refer to the same thing (even our world could have been such a scenario, but we discovered that it is not). I think that the relevant notion of possibility involved in the conceivability argument is possibility according to all scientific discoveries: once we discover an identity in science, it is inconceivable and impossible to contradict this identity since it is a necessary truth.

The strange thing about the conceivability argument applied to a zombie is the fact that it holds even when we apply it to the actual world. The separability of consciousness and body is something that we are prepared to accept in the actual world, unlike the separability of water and H<sub>2</sub>O. It seems that we have no reason to assert the mind-body identity, mainly due to their great dissimilarity. How can something so particular as consciousness result from the composite physical structure of the brain? Maybe someday, our scientific experiments will give an answer to that wonder; this future possibility leads us to reexamine the idea of conceivability, because mere apparent lack of contradiction is not sufficient to prove that a zombie is really possible or in principle possible<sup>8</sup>.

We notice that there is still another way to challenge the conceivability argument and his power to settle some ontological facts. We have succeeded in showing that the materialists' arguments that try to challenge the second premise of the conceivability argument are not successful: the conceivability of a zombie still implies its possibility. All we still have to do is challenge also the first premise of the argument, the conceivability of a zombie.

First, I will reconsider the inconceivability of a zombie. In Chalmers 2002a, we can identify some different ways of defining conceivability, among them some invoking the idea of negative and positive conceivability: a statement negatively conceivable is one that could be a priori true or one whose negation is not a priori false, and a positively conceivable statement is one that describes a situation that can exist. At the same time, we have the notion of an ideally conceivable situation, a situation that proves coherent upon careful reflection and maybe, we can add, coherent from the perspective of a final, complete scientific theory<sup>9</sup>. Taking together these two notions, we can say that

<sup>8</sup> And the conceivability argument needs this idea of in principle conceivability.

<sup>9</sup> This idea of conceivability from the perspective of a complete scientific theory does not appear in Chalmers (2002), but I think that it is easy to add this clause to an ideally conceivable situation.

a zombie is an ideally positively conceivable creature and so metaphysically possible.

However, is really a zombie an ideally positively conceivable creature? If we further explore the idea of ideal conceivability I think that we could arrive at the conclusion that a zombie could be inconceivable and impossible in light of future scientific developments which will fill the explanatory and epistemic gap which we still face regarding the relationship between the physical properties of the brain and consciousness. This possibility lies at the ground of materialist conceptions of consciousness and seems to be the only available theoretical position for a materialist: zombies are still conceivable, due to the fact that we do not have properly discovered the relationship between experiences and their physical substrate, but zombies are nevertheless metaphysically impossible because consciousness is identical with its physical basis.

Theoretical identities from science are the best example of necessary a posteriori sentences, in which two sets of properties (superficial and chemical-physical properties) prove to be identical. More than that, the identity is a necessary one due to the fact that the terms involved in the identity sentence refer rigidly to the same entity. At the same time, the case of consciousness and the brain is different. Therefore, materialists have to give a different account of this case and to challenge the intuitions that we can have the same kind of physical constitution without consciousness, with a different kind of consciousness or even disembodied consciousness.

The most hard to challenge assumption which gives support to the conceivability argument is the assumption concerning the phenomenal properties and phenomenal vocabulary: phenomenal properties are essential to the states that instantiate them and every physical state that can realize a phenomenal property must have this property. If the materialists want their conception to be a viable one, the mind body- identity should be a necessary one. At the same time some of our intuitions take us in an opposite direction, pointing at the contingency of this relation (the relation between phenomenal states and states of the brain) because we can have cases of occurrence of the physical properties of the brain that in our world accompany consciousness without the occurrence of any phenomenal properties. Further cases are of same physical structures with different qualia or qualia without any physical support.

But why not consider these intuitions misleading? It may be the case that the physical properties of the brain and phenomenal properties are identical, are one and the same property. At the same time, our actual scientific theories are not capable to adequately explain this fact.

#### IV. A PROBLEM WITH THE DEFINITION OF MATERIALISM

Given the fact that Chalmers defines materialism as a supervenience thesis and not as an identity thesis, there is a new interesting model for the brain-consciousness relationship. Another possibility for the relationship

between brain and consciousness might be the following one: it may be a necessary relation between the phenomenal property and its actual physical correlate, where the necessity does not arise from rigid designation, but from a connection at the metaphysical level. Whenever we refer to a phenomenal property we implicitly refer to its neuronal substrate.

We have seen that we can define materialism as a supervenience thesis, where supervenient properties cannot occur in any possible world without the occurrence of the physical properties on which they supervene. But is this way of defining materialism as a supervenience thesis sufficient to ground the reduction of consciousness to the physical properties of the brain? I think it is not and this fact can be seen from the fact that it has an interesting consequence for the mind-body relationship:

1. We can have a case in which conscious properties are accompanied in every possible world by the same physical properties just because they are one and the same physical property; in other words, consciousness can be analyzed in terms of physical-structural properties or in functional terms which can be reduced to physical-structural ones even if we can have only a posteriori such a description. From this point of view, the conceivability of zombies is caused by the fact that in our present stage of scientific development, the scientific account does not properly explain consciousness. This middle position, where the temporary conceivability position is due to an epistemic gap, will amount in the end to the inconceivability of zombies, therefore to rejection of the first premise of the conceivability argument.
2. On the other hand, we have the case where phenomenal properties occur only in the presence of physical properties because they are connected to the physical by a necessary relation of metaphysical constitution, despite the fact that they are different. Such points of view are put forward by Thomas Nagel (2002) and by Galen Strawson (2008); both of them consider themselves materialists, but without reducing the properties of consciousness to the strictly physical properties of the brain. According to their views, matter has two aspects, a physical-structural one and an intrinsic one, which are nevertheless closely tied to each other.

I think that the first kind of materialism is hard to support due to the fact that such materialists have really to say that phenomenal properties, which are not functional-structural ones, arise from such structural properties or that it is an illusion that they are not structural properties.

It is debatable whether the second view alluded to here is a genuine materialist point of view. On the one hand, it is advertised as a materialist point of view, but on the other hand, in view of the fact that it proposes two series of properties, it seems not to be a materialist conception but a dualist one or



a monist neutral one. Does it allow for the conceivability of a zombie? Both authors previously mentioned deny the conceivability of zombies as a serious hypothesis, and they do not take this metaphysical possibility as a viable metaphysical alternative: Strawson makes this point explicit, by considering that between the intrinsic nature of an entity and its physical-structural properties, there is a necessary metaphysical connection. On such views, we do not need to appeal any more to the conceivability argument to ground a dualist point of view, but to the conception of a dual set of properties, neither of which is reducible to the other. It is in this respect that we can speak of dualism without zombies.

## V. DUALISM WITHOUT ZOMBIES

The key of the Chalmers-style dualist argument is the conception about physical entities and properties which appears in the definition of materialism or physicalism. From this point of view, a material entity (and matter generally) presents two kinds of properties: on the one hand it has structural properties, of the kind taken into account by physics, and on the other hand it has some intrinsic properties, about which physics is silent. Given this state of the things, we can consider that consciousness arises from these intrinsic properties of matter given the fact that its specific properties, phenomenal properties are also intrinsic. Whether matter presents directly phenomenal properties or only some proto-phenomenal properties from the combination of which phenomenal properties arise at certain level of complexity (of some intrinsic proto-phenomenal properties) is only a matter of choice.

We can also see what can potentially differentiate between zombie worlds and worlds with conscious minds, in the case in which we still support the metaphysical possibility of such creatures: a different deep structure of matter. In zombie worlds, matter either has completely different intrinsic properties that are incapable of giving rise to consciousness, or its intrinsic properties didn't evolve toward the complexity from our world. There is also another possibility, namely that matter in such worlds does not have intrinsic nature, but this seems to me absurd, is like saying that there can be a world where entities have only form (structural/dispositional properties), without intrinsic ones.

The materialist who wants to further support his point of view must attack this particular characterization of materialism. There are some options available here: 1. deny that matter has some intrinsic properties besides the structural ones which constitute the subject matter of physics and at the same time pretend that phenomenal properties will someday be reduced by scientific explanation to structural ones or<sup>10</sup> 2. consider that the intrinsic properties of matter are also part of a materialist view of the world and that their existence show only that it is something wrong with our way of conceiving science as

<sup>10</sup> We have already seen in the first section what is a structural property.

dealing only with structural properties or 3. consider that from physical-structural properties could arise some intrinsic properties at some level of complexity; intrinsic phenomenal properties emerge from structural properties.

## VI. SUPERVENIENCE RECONSIDERED

What I wanted to show is that the anti-materialist argument can survive the materialist objections, even without invoking the idea of a zombie. Therefore, we can concede that there are logically coherent worlds which cannot be metaphysically realized and even that zombie is inconceivable (inconceivable in the sense in which the same structural-causal aspect of the brain is always accompanied by the same phenomenal properties, the causal and structural properties being necessarily tied to their intrinsic counterpart so that we cannot imagine a possible world where the first does not occur without the second; despite this co-occurrence, they are not identical), that the idea of the same brain without consciousness is an illusion. The possibility of still being a dualist is granted by the fact that the real force of the argument is given by the definition of materialism and by the way of conceiving the physical, namely as an array of structural properties of some entities with a hidden nature.

This view takes us to a reconsideration of the idea of supervenience, which seems incapable of giving a reductive view of consciousness. Suppose that in every possible world the same physical properties of the brain are co-present with the same phenomenal properties and vice versa, the same phenomenal properties are co-present with the same physical properties of the brain. Will this case lead us to say that they are identical, and that the conscious properties are reducible to the neuronal properties, instead of saying that they are merely correlated? I think that we will not have any ground to choose between identity and mere co-instantiation of still different properties.

In settling these matters, I think that it is useful to analyze the way in which we establish the usual identities in science, of the kind „Water is H<sub>2</sub>O“ or „Heat is mean molecular motion“. As we can notice from the literature there are two ways of assessing these identities, one which proceeds via conceptual analysis and the other without such an analysis. In the first case we have an a priori definition of say water given in functional terms and we later discover some substance whose structure instantiates the relations of dependence as they appear in the functional definitions: for example in the case of water we discover that the structure of H<sub>2</sub>O is capable of explaining all the superficial properties associated with the term „water“. The other way to assess the identity proceeds without conceptual analysis, considering that we cannot usually find such analysis for our theoretical terms. With or without functional analysis the two points of view finally lead to the same point, as far as the second step of the first procedure is identical for them; they differ only from a semantic point of view, so long as the second option leaves open the semantic content of the concepts involved.

I think that a comment should be made here: the possibility of having such analysis is the core of one of Chalmers' anti-materialist epistemic arguments, namely the argument from the absence of analysis. The argument involves the following steps: reduction proceeds via functional analysis, we cannot give a functional analysis to our phenomenal concepts; therefore, phenomenal properties are not reducible. I want to make two observations regarding this argument: 1. first, it can only affect a form of materialism which supports functionalism in the case of consciousness and not a stronger form of biological or physical materialism concerning the mind-body relation so long as the materialist can grant that we cannot give a functionalist analysis of phenomenal properties but could nevertheless give a biological analysis; 2. second, I do not consider that it is the key of the dualist position and despite the fact that we can reject the idea of conceptual reduction via functional analysis, we are not compelled to reject dualism; the burden is on the statement asserting that from structure and dynamics we can derive only structure and dynamics.

The assessment of the identity for natural kinds is a pragmatic one: from the fact that we observe that the bundle of macro-physical properties is always associated with the same micro-physical structure we infer that they are identical. The essential condition which makes possible the reduction is the following one: what are the macro-physical properties other than a way of translating at the level of our minds the micro-physical ones? Nothing, the macro-physical properties are the expression of the profound structure of matter at the level of our minds. For example in the case of water, behind our phenomenal concept of liquidity lies the concept of a substance with a particular molecular structure that can be easily broken; we can easily see why H<sub>2</sub>O is liquid, so long as the forces that keep the molecules of this chemical compound together are weak ones and could be easily broken. The same considerations apply to other properties, like colorless, odorless and so on. At the level of the world there are only particles and waves that interact in many ways, but the superficial properties are introduced by our minds and in the case of assessing identities we give such a translation of the two ways of talking about the world. Therefore, one of the components of the identity is the appearance created by our mind itself which in turn awaits an explanation.

Not all identities involve the phenomenal properties of our mind on the side of the reducing properties. For example in the explanation of life or in the explanation of other biological identities no phenomenal concept is involved. Even though there are not phenomenal concepts involved there, we can nevertheless deal with such cases in the same manner previously envisaged: the manual of translation involves this time two levels of individuation, a finer grained one on the one hand and a coarser grained one on the other hand. „Life“ for example is a bundle concept which subsumes a host of phenomena characterized in functional terms, in terms of structures of relations which can be instantiated by different arrays of material complexes of

molecules; at the end, we have at the metaphysical level only physical components and their way of interacting, which we can further individuate in different ways.

Now in light of the previous analysis we can easily see how the relation of supervenience is guaranteed in the familiar cases: this relation holds in virtue of the constitution of the objects in some cases and constitution of the objects plus facts about how they affect our minds in other cases. So long as the setting of our mind or the setting of other minds which encounter H<sub>2</sub>O is the same and also the background conditions are the same, the structure of H<sub>2</sub>O will have the same superficial properties and the relation of logical supervenience will hold.

There is not anything mysterious about the logical supervenience of material objects, but when it comes to the mind itself, things get complicated. There is not anything between the structure of the brain and the experience that could mediate between these two entities (or aspects of the same entity). Supervenience in the mind body case is a special relation, different from the one present in the case of other natural kinds because the relation of the mind with the brain is not mediated like the relation between the world and the mind, which is mediated by the particular setting of the mind. So, co-presence of the same phenomenal properties and neuronal properties in every possible world will not result in identity at the ontological level. For the relation of co-instantiation or correlation to become identity, we need a stronger ontological ground and the materialists have to present an ontological model which is capable of challenging the dualist distinction between the two aspects of matter one.

The supervenience in the two cases, that of the ordinary identities from science and that of consciousness and the brain are different in some respects. In the case of scientific identities, once we have discovered the deep physical structure of the entity (for example the formula of water: H<sub>2</sub>O) we want to reduce, this identity becomes a necessary one: it is inconceivable to have a physical identical world where H<sub>2</sub>O does not have the superficial properties associated with water. Here, natural necessity is sufficient for metaphysical necessity: if H<sub>2</sub>O had other superficial properties and this change of appearance could not be accounted in terms of different perceptions or different environmental conditions, the physical structure would not explain its superficial properties and we would be faced with the question „where do they come from if not from the physical level?“ More than that, it is worth emphasizing the fact that the superficial appearance is a product of the interaction between the perceiver and the physical reality.

On the other hand, in case of the supervenience of consciousness on the brain, their correlation in every case in our world is not sufficient to ground their identity. To show that it is identity, we have to overcome the difference between a structural set of properties, those of the brain and a non-structural one, those of intrinsic, non-structural properties of consciousness. It is not

obvious at all that we can really achieve something like that. And the logical conceivability of zombies is the direct consequence of this problem.

More than that, the case of consciousness involves an observable and an unobservable side: the brain is publicly observable, whereas consciousness is accessible only from the first person point of view. Therefore, we know for our own case that we have experiences, we have phenomenal consciousness, but what about the other brains, do they feel anything, do they have the same experiences like us? They most probably have, but this fact cannot be established via an empirical observation; it is highly reasonable to suppose the fact that a brain with the same configuration will give rise to the same experiences, but only if we suppose that nature has a uniform deep nature: nothing prevents us from *a priori* believing that the others are zombies, not even all empirical observations; it is logically consistent with them to have only zombies around us. We can fight against this skepticism with a principle of natural order, which is a metaphysical and not a physical component in a theory.

The same hypothesis can be extended afterwards to all worlds, supposing that we can have the same physical properties of the brain without any experience. We have again a metaphysical possibility which cannot be fought on empirical grounds. By opening these metaphysical possibilities we have the following scenarios: same physical properties (in the sense used by Chalmers) without any phenomenal properties in some worlds but not in others, same physical properties with different phenomenal properties in some worlds or same physical properties with same experiences in all possible worlds. It seems impossible *a priori* to choose between these possibilities; what is important for our case is that all of them present a difference of properties and irreducibility of one kind to the other. We can choose between these possible scenarios only if we support the idea that zombies are inconceivable or if we introduce a notion of strong metaphysical necessity so that the same physical structures to be connected with the same intrinsic ones in every possible world.

## VII. WHY ZOMBIES ARE NOT DESIRABLE IN ONE'S THEORY

In this last section I will give two reasons for which I think that it is not reasonable to have zombies in one's theory, even though I acknowledge that we cannot definitively rule out the conceivability of such creatures.

When constructing the zombie case, philosophers generally do not give many details about the mentality of such a creature. Apart from the fact that it is psychologically like us<sup>11</sup>, it has no experiences and lacks phenomenal consciousness. There we can develop the scenario a little bit and fill some

---

<sup>11</sup> This is the way in which Chalmers (1995) defines a zombie: psychology amounts to the functional organization of a conscious being, his behavior.

informational gaps. Therefore, a zombie can process information about his environment and himself, but seems to have entirely different means of representation compared with us; certainly, he lacks our picturesque and vivid manner of representing reality (given in phenomenal consciousness). What is surprising is that he acts in the same manner as us, is capable of making the same judgments and same behavioral acts. Given this fact, we are tempted to put the following question: how can such different representations play the same causal roles in constituting on the one hand the mentality of a zombie and on the other hand the mentality of a conscious being? And more than that why does the same structural composition of matter, the same kind of brain viewed from its physical-structural point of view, supports two different kinds of representations?

The image of mentality just outlined in the zombie argument makes phenomenal properties inefficient in bringing about some psychological effect, in other words makes them epiphenomenal. Therefore, phenomenal properties do not have any causal role whatsoever. My intuition is that this is a problematic point, because the evidence seems to pull us in a different direction: consciousness seems to have a causal role. If we succeed in showing that phenomenal consciousness has a big role in constituting mentality and has behavioral effects, the zombie argument no longer works: a zombie cannot have the same mentality as a conscious being because he lacks experience and experience is directly involved in the causation of behavior.

How can we account for the fact that phenomenal properties are causally efficacious for behavior and are not epiphenomenal? Let's analyze the following scenario: a zombie cuts his finger and at the same time a human person cuts his finger. Of course, zombie does not feel phenomenal pain, but at the same time he must have the same psychological mechanism<sup>12</sup> which prevents him for doing this kind of self-destroying acts in the future; he must have some mechanism of pain avoidance, even though his pain is different from our pain. On the other hand, in the case of the human person, the pain itself seems to be a sufficient reason for avoiding such future acts. We can argue that avoiding pain is not an end in itself in this case, that it serves a broader evolutionary purpose, that by avoiding pain we learn to avoid these situations that can lead to our self-destruction. Nevertheless, acknowledging this fact does not make us to change our mind concerning the causal effect of the experience of pain on behavior: this unpleasant feeling of pain is the shortest way to implement a mechanism of avoidance, which further implement a long term evolutionary goal.

A more complicated mechanism of avoidance would be one which would involve rationally following some rules. So, we can learn that a particular representation at the level of our mental system is associated with harm and whenever we come upon the situation that this particular representation

---

<sup>12</sup> The neural counterpart of our brain when we feel pain and the specific pain behavior.



gives us information about, we should develop a behavior of avoidance. It seems that the mental structure of zombie has to support such a psychological mechanism, given the fact that he does not have experiences. I don't want to say that such a system is not possible, what I want to underlie is that it must be much more complicated than our system which relies on phenomenal properties in order to bring about some behavioral reactions.

The contingent link between the phenomenal properties and behavior is used in the inverted spectra kind of arguments and the classical example are colors. The argument proceeds roughly as follows: we can conceive two creatures with inverted experiences when stimulated with the same kind of stimuli, but who are indistinguishable as far as behavior is concerned despite their different experiences. We can conclude that experiences do not play any causal in producing behavior, all our reactions connected to the experience of a particular color are learned and therefore contingent. I think that in spite of the fact that we can give such an account in the case of colors (although I doubt that colors do not have immediate effects on our behavior which are not learned in the process of learning how to apply the word), such a model does not work in the case of much elementary feelings, like pains or pleasures. A creature whose spectrum is inverted as far as pain and pleasure is concerned, will not have the same behavior as a normal creature.

The second reason for rejecting zombies is an epistemic one, which has profound implications for the theory of consciousness in a dualist form. From a dualist perspective, phenomenal properties reveal themselves only in experience, so that the access is a limited one, restricted to the subjective point of view of the first person perspective. On the other hand, science requires objectivity and inter-subjective testability. Therefore, a natural way of scientifically investigating phenomenal consciousness would be through a series of correlations between phenomenal properties and physical- functional properties.

As far as natural sciences are concerned, we presuppose a principle of uniformity of nature, stating roughly that there are no ontological discontinuities in the material world and that the laws of nature apply in all possible cases of some kind. In the same manner, in a science of consciousness we have to presuppose uniformity in pairing experiences with physical structures in the brain (if we allow the case of physical and phenomenal properties being two distinct kinds of properties): the same structures of the brain have to be accompanied by roughly the same experiences. The zombie case seems to break this relation of symmetry. Another case which can break the symmetry is one of inverted spectrum: let's say Mary and inverted Mary are perfect duplicates as far as the structure of the brain and functional organization are concerned and they are inverted with respect to their experiences. The cases can be multiplied by other imaginary examples of permutations between sensorial modalities: cases of creatures which respond which auditory experiences to the stimuli to which we respond with visual representations (and

so on for the other possible combinations between modalities and stimuli) or which respond with totally different representations to the same stimuli.

If we suppose that we can have such cases of zombies and of different experiences, the science of consciousness becomes impossible. The link between the structure of the brain to which we have objective access and the experience associated is a too loose one for grounding a science of consciousness. Therefore, we better suppose that the same physical structure is tied to a particular phenomenal structure in every possible case, so that to avoid the cases previously outlined.

Suppose that a zombie comes to visit our world and happens to go to a neuroscience laboratory where scientists investigate phenomenal consciousness. Do these scientists have any means to realize that the creature in front of them is a zombie and he does not have phenomenal consciousness in spite of the fact that he has the same neuronal structure as a conscious being? The obvious answer is no, they don't have any clue. Zombie's utterances<sup>13</sup> as far as phenomenal consciousness are false but we have no means to detect that they are false. I think this fact will render the dualist science of consciousness<sup>14</sup> useless. Therefore, we better give up the idea of a case where the same structures of the brain are not accompanied by experience at all or are accompanied by another type of experience, unless we want to fall back on a functionalist theory of consciousness.

## REFERENCES:

- Brown, Richard. 2010. „Deprioritizing the a priori arguments against physicalism“. *Journal of Consciousness Studies*, 17 (3-4).
- Chalmers, David J. 1995. *The Conscious Mind*. Oxford: Oxford University Press.
- Chalmers, David J. 1999. „Materialism and the Metaphysics of Modality“. *Philosophy and Phenomenological Research*, 59 (2): 473-96.
- Chalmers, David J. 2002a. „Does Conceivability Entail Possibility?“. In Tamar S. Gendler & John Hawthorne (eds.). *Conceivability and Possibility*, Oxford: Oxford University Press.
- Hill, Christopher S. 1997. „Imaginability, Conceivability, Possibility, and the Mind-body Problem“. *Philosophical Studies*, 87 (1): 61-85.

<sup>13</sup> Since we supposed that zombie is identical from a physical and psychological point of view with a conscious being, all their behavioral acts, including the claims about consciousness, will be identical, despite the fact that zombie's utterances about its phenomenal consciousness are false.

<sup>14</sup> The dualist science of consciousness presupposes some bridging principles between the physical-structural and phenomenal properties (Chalmers 1995).

- Hill, Christopher S. & McLaughlin, Brian P. (1999). „There are fewer Things in Reality than are Dreamt of in Chalmers’s Philosophy“. *Philosophy and Phenomenological Research*, 59 (2):445-454.
- Kripke, Saul. 1980. *Naming and Necessity*. Harvard University Press.
- Levine, Joseph. 2002. *Purple Haze*. Oxford: Oxford University Press.
- Loar, Brian. 1990. „Phenomenal States“. In Chalmers, David (2001), *Classical and Contemporary Readings in the Philosophy of Mind*, Oxford: Oxford University Press .
- Loar, Brian. 1999. „David Chalmers’s The Conscious Mind“. In *Philosophy and Phenomenological Research*, 59(2): 465-472.
- Puccinini, Gualtiero. 2008. „Access denied to zombies“, web draft.
- Nagel, Thomas. 2002. *Concealment and Exposure and other Essays*. New York: Oxford University Press.
- Strawson, Galen. 2008. *Real Materialism and other Essays*. Oxford: Oxford University Press.