Zombies cannot be there

Marco Giunti
University of Cagliari
email: giunti@unica.it
homepage: http://giuntihome.dadacasa.supereva.it

Keywords
reductive explanation; supervenience; hard problem; explanatory gap; zombie; logical possibility; conceivability; ontology; physicalism; materialism; dualism; qualia; concept of consciousness

Short abstract
On the basis of the distinction between phenomenal and psychological consciousness, I propose a formal framework where we can express and analyze a strong form of Chalmers' zombie argument. By employing such formal framework, I make clear the kind of problem that this argument poses to anyone who is willing to (i) construct a theory of phenomenal consciousness and (ii) maintain the reductive explainability of phenomenal consciousness by physics. I then extend such formal framework so as to provide a theory of consciousness in axiomatized form. The explanation of phenomenal consciousness provided by this theory is by no means inconsistent with a physicalist perspective. In fact, once the theory is supplemented with a minimal physicalist assumption, we can prove the

reductive explainability of phenomenal consciousness by physics and, as a consequence, the logical impossibility of zombie worlds. I then conclude that Chalmers' intuition, according to which phenomenal consciousness is beyond the scope of any physical theory, is not tenable. Finally, I hint at a possible source of such erroneous perception, that is, not realizing that the hard problem of consciousness is not a problem of scientific explanation but, rather, a philosophical problem of conceptual explication.

Long abstract

I propose a simple formal possible world ontology where we can naturally express and analyze a strong form of Chalmers' zombie argument. By employing such formal framework, I make clear the kind of problem that this argument poses to anyone who is willing to (i) construct a theory of phenomenal consciousness and (ii) maintain the reductive explainability of phenomenal consciousness by physics.

I then extend such formal ontology so as to provide a theory of consciousness in axiomatized form. The key idea of my theory is to show how, within an appropriate extension of the proposed possible world ontology, we can make sense of the distinction between first person or internal facts of consciousness (*i.e. phenomenal consciousness*) and third person or external ones (*i.e. psychological consciousness*), in a natural manner. The formal machinery which allows such an extension is a straightforward adaptation of the standard axioms of the lambda calculus. According to my theory, being internal (phenomenal) or external (psychological or functional) is not an intrinsic feature of a fact of consciousness but, rather, depends on whether or not a special relationship between the fact and a reference system holds. Thus, the very same fact of consciousness turns out to be internal (or phenomenal) with respect to a reference system, and external with respect to other systems. It is a theorem that, for any fact of consciousness, there is exactly one system for which the fact is internal. I call such a privileged system *the subject* of the fact of consciousness; the fact is external for any other system.

My theory of phenomenal consciousness is by no means inconsistent with the claim that phenomenal consciousness is reductively explainable by physics. In fact, once the theory is supplemented with a minimal physicalist assumption, according to which the *external* (or third person) facts of consciousness are logical consequences of the physical facts, we can *prove* that the internal (or first person) facts are logical consequences of the physical facts as well. A corollary of this theorem is the logical impossibility of a zombie world (*i.e.* a world where all the physical facts of our world hold, but no internal fact of consciousness holds). These results pose a dilemma to Chalmers and all zombies' friends: if a zombie world is logically possible, then, contrary to Chalmers' position, the *external* facts of consciousness are not entailed by the physical facts or, equivalently, *psychological* consciousness is not reductively explainable by physics. Conversely, if psychological consciousness can be reductively explained by physics, then zombie worlds are logically impossible.

I conclude that Chalmers' intuition, according to which phenomenal consciousness is beyond the scope of any physical theory, is not tenable. Finally, I hint at a possible source of such erroneous perception, *i.e.*, not realizing that the hard problem of consciousness is not a problem of *scientific explanation* but, rather, a *philosophical* problem of conceptual *explication*.

1 THE PROBLEM OF PHENOMENAL CONSCIOUSNESS IN LATE 20TH CENTURY

In the second half of the 20th century, the analytical tradition in the philosophy of mind did not consider the phenomenal, experiential, or subjective dimension of consciousness as its primary concern. According to either philosophical behaviorism (Ryle 1949), identity theory (Place 1956; Smart 1959; Lewis 1966, 1972; Armstrong 1968, 1981b) or functionalism (Putnam 1960, 1964, 1965, 1967a, 1967b), the mind was to be studied from a strictly objective, external, or third person point of view. For behaviorism, the problem was to describe the direct nomological relationships between externally observable behaviors; identity theory reduced mental states to physical states; and, finally, functionalism was concerned with the relationships between externally ascertainable inputs/outputs and internal, notwithstanding objective, mental states.[1]

This situation, however, slowly began to change around the mid seventies. On the one hand, a number of now classic arguments[2] were devised, which purported to show how the phenomenal, experiential or subjective dimension of consciousness is ultimately irreducible to the psychological, functional or objective aspect. On the other one, the problem of consciousness attracted an increasing number of scholars, psychologists and philosophers, as well as neuroscientists, physicists, mathematicians, and many other researchers active in the diverse fields of cognitive science.

This growing interest in the problem of consciousness resulted in a first conference, which took place in Tucson in 1994. The Tucson conferences have then become regular biennial meetings.[3] Retrospectively, the first Tucson conference can be considered the founding act of a new interdisciplinary field, called *Consciousness Studies*. One of its main centers is the *University of Arizona* at Tucson, where in 1998 the *Center for Consciousness Studies* was founded, whose directors are David Chalmers, Stuart Hameroff and Alfred Kaszniak. Ned Block is the president of the *ASSC* (*Association for the Scientific Study of Consciousness*), which organizes regular annual meetings.[4] In addition, new journals expressly dedicated to consciousness studies have appeared; among these, the *Journal of Consciousness Studies*,[5] *Psyche*,[6] *Consciousness and Cognition*,[7] and *Consciousness and Emotion*.[8]

2 THE DISTINCTION BETWEEN PHENOMENAL AND PSYCHOLOGICAL CONSCIOUSNESS

The problem of consciousness, in the special form that took shape at the end of the 20th century, consists, in the first place, in recognizing that there

is at least a conceptual distinction between two kinds of facts of consciousness: (i) the *internal* (or *first person*) *facts of consciousness*, or the *conscious experiences*, and (ii) the *external* (or *third person*) *facts of consciousness*, or the *conscious functions*. The set of all the internal facts of consciousness is the *phenomenal consciousness*; the set of all the external facts of consciousness is the *psychological consciousness*.

The distinction between phenomenal and psychological consciousness is the focus of a series of classic arguments: the *What is it like to be* argument (Nagel 1974), the *Absent qualia* (Block 1978) and the *Inverted spectrum* (Block 1978; already in Locke 1690), *Mary's* or the *Knowledge* argument (Jackson 1982), the *Zombie* argument (Chalmers 1996, also see note 2). Each of these arguments, besides making clear the *conceptual* distinction between conscious experiences and conscious functions, purports to show, in a more or less explicit fashion, that such a distinction is *real* too.

To get a first understanding of this distinction it is useful to recall the paradigmatic instance of an internal fact of consciousness – the experience of *qualia* or, equivalently, the experience of *immediately felt qualities*.

The experience of qualia can be conveniently described by means of the following general scheme:

(Q) **somebody**'s conscious experience, who is [seeing / hearing / tasting / smelling / feeling] the **quality** of **something**

By substituting in scheme (Q) a determinate value for the terms in bold, and by choosing the most appropriate among the five verbs of perception, we get, for example, the following descriptions of particular qualia experiences:

1. *Mary's conscious experience, who is seeing the green of a mountain meadow*;
2. *Mary's conscious experience, who is tasting the sweet of a lump of sugar*;
3. *John's conscious experience, who is hearing the pitch of a violin sound*;
4. *John's conscious experience, who is feeling the roughness of a tree trunk*;
5. *Mary's conscious experience, who is smelling the perfume of a rose*;
6. *Mary's conscious experience, who is feeling the pain of a hammer blow in her toe*;
7. *John's conscious experience, who is feeling the pleasure of smoking a cigar*;

8. *John's conscious experience, who is feeling the warmth of an August day*;
9. *Mary's conscious experience, who is feeling the coldness of a block of ice*;

A little reflection on these examples suffices to show that each qualia experience they describe corresponds[9] to an external fact of consciousness. Let us consider example 6. The pain lively experienced from the inside by Mary corresponds to a specific activation state of the relevant parts of her peripheral and/or central nervous system and, typically, to a series of externally ascertainable behaviors (verbal, motor, etc.). It is exactly this complex (neuro-physiological state + behaviors, if any) that, given its functional organization, constitutes the external fact, or the conscious function, which corresponds to Mary's internal consciousness of a pain.

Furthermore, the preceding observation concerning a correspondence between qualia experiences and conscious functions can be generalized to all facts of consciousness: for any fact of consciousness, if it is internal, there is an external counterpart[10] and, conversely, if it is external, there is an internal one.[11]

Having thus made clear that internal and external facts of consciousness are at least conceptually distinct, the question whether a real distinction holds too is nevertheless wide open. In other words, do the two concepts of internal and external fact of consciousness indeed individuate two different sets of facts or, rather, these concepts are just two different ways of describing the same set? I will take up again this issue later on (see sec. 9.1).

3 CHALMERS' PROBLEMS OF CONSCIOUSNESS

Chalmers' (1996) most important contribution to the consciousness debate probably consists in his new formulation of the problem, which shifts the focus just on the distinction between psychological and phenomenal consciousness. With respect to this distinction, Chalmers speaks of an *easy* and a *hard problem* of consciousness. For the sake of the present inquiry, however, it is convenient to make a sharp division between two different versions of Chalmers' hard problem, which I will call, respectively, *hard problem* and *super-hard problem*.

3.1 *THE EASY PROBLEM*

In general, the easy problem consists in *finding and developing an adequate theory of neuro-psychological consciousness*. Specific formulations of the easy problem are dealt with and solved with increasing success by scientific psychology, neuroscience, and cognitive science. For Chalmers, notwithstanding the different levels of description proper of these disciplines, their particular solutions are all based on the *functionalist paradigm*, according to which the external facts of consciousness can always be explained by pointing out the specific causal relationships between external inputs and outputs and mental (or neuro-physiological) internal states.

For Chalmers, psychological consciousness does not involve any special difficulty, even from the viewpoint of the mind-body problem. It is in fact clear, in principle, how any theory of the functionalist kind can be reductively explained by a physical theory. Therefore, if the external facts of consciousness are *conscious functions*, they are *reductively explainable*, at least in principle, by means of *physical facts*.

3.2 *THE HARD PROBLEM*

The hard problem consists in *finding and developing an adequate theory of phenomenal consciousness*. This problem, according to Chalmers, has been widely overlooked by scientific psychology, neuroscience, and cognitive science. In addition, that is even worse, at the moment we do not have any slight hint as to how employing the functionalist paradigm to construct an explanation of phenomenal consciousness. In fact, any functionalist explanation that has been put forth so far does not seem to shed light on its intended *explanandum* but, rather, just on its external correlates, *i.e.*, ultimately, just on psychological consciousness.

As a consequence, it is by no means clear how a reductive explanation of the internal facts of consciousness in terms of physical facts could be possible. For such explanation would first require a functionalist explanation of phenomenal consciousness; but, since the latter is a mystery, a reductive explanation of conscious experience in terms of physical facts is a more deep-rooted mystery.

3.3 *THE SUPER-HARD PROBLEM*

According to Chalmers, however, the internal facts of consciousness *are not reductively explainable* by means of physical facts and thus, *a fortiori*, by conscious functions either. This explains why, so far, there has been

no progress in the construction of a scientific theory of conscious experience. Such a theory, in fact, *cannot* be based on the functionalist paradigm, for we know that any theory of this kind is reductively explainable, in principle, by physical theory.

It thus follows that, in order to be able to deal with the hard problem of consciousness, it is necessary to construct first a new paradigm, which should turn out to be as appropriate for the study of the internal facts of consciousness as the functionalist one is for the external facts. In more detail, we need:

1. to construct a radically new scientific paradigm, which should then allow us to elaborate an adequate theory of phenomenal consciousness;
2. such a paradigm should be developed within a strongly antireductionist,[12] and thus dualist, metaphysical framework; hence, there are two types of fundamental facts: *conscious experiences* and *physical facts*;
3. we should therefore deal with the general problem of the relationships between these two types of fundamental facts;
4. in particular, as regards the relation of cause, we should deal with the problem of the causal efficacy of conscious experiences. As it is well known, Chalmers' position on this issue is a form of epiphenomenalism;
5. we should also find new empirical methods, adequate to the study of phenomenal consciousness. Thus, besides the intersubjective, or third person, methods proper of both the natural sciences and scientific psychology, we also need to consider *subjective* or *first person methods*. A renewed interest for the historicist, hermeneutic, phenomenological, existentialist tradition of '900 European philosophy ensues.

If we consider all points 1-5, however, it seems that the problem set forth by Chalmers[13] is not just *extremely* difficult, but almost intractable. Perhaps, then, the following approach is both more convenient and interesting:

- deal with, not deny, the problem of constructing a theory of phenomenal consciousness;
- nevertheless, carry out such a construction within a rigorous physicalist framework.

Yet, if Chalmers is right in maintaining that phenomenal consciousness is not reductively explainable in terms of physical facts, this line of research cannot be pursued. It is then necessary to analyze first the key argument

on which Chalmers grounds his tenet. As it is well known, this is the *phenomenal zombie* argument. To gain a sufficiently deep understanding of the argument, however, we need to introduce a minimal formal apparatus.

4 THE ZOMBIE ARGUMENT AND CHALMERS' CHALLENGE

The mathematical structure we need to formulate and understand the zombie argument is a quintuple (**D**, **P**, **Q**, **W**, **w***) such that:

1. **D** is a non empty set, called *domain*;
2. **P** is a non empty set; each element of **P** is an ordered pair (P, n) such that n is a natural number and $1 \leq n$; any such pair will be indicated with the notation P^n; every element of **P** is called a *property* and the number n is called its *arity* (or its *number of places*). Note that the precise nature of the first element of each property is not specified, and that it is by no means relevant in what follows. In addition, if the context makes clear the arity n of property P^n, such property may be indicated with the shorter notation P ;
3. **Q** is a subset of **P**; **Q** is called *the set of physical properties*;
4. **W** is a set of functions; each of them, to any property P^n member of **P**, associates a set of n-tuples $\{(x_1, x_2, \ldots, x_n)\}$ such that, for any n-tuple in such set, any of its elements, x_i , is a member of **D**; every function w member of **W** is called a *possible world*; the set of n-tuples assigned by w to P^n is indicated by $P^n{}_w$ and it is called the *extension of* P^n *in world* w;
5. **w*** $\in$ **W**; **w*** is called *our world* or *the actual world*.

This structure allows us to give a series of definitions:

[1] *f is a possible fact* (or, simply, *a fact*) iff:
f is an ordered pair of the type (P^n, (x_1, x_2, … , x_n)), where $P^n \in$ **P** and (x_1, x_2, … , x_n) is an n-tuple ($n \geq 1$) of elements of **D**;
the property P^n is called the *predicate of f* and the n-tuple (x_1, x_2, … , x_n) is called the *subject of f*; keep in mind that, if $n = 1$, the subject reduces to an n-tuple of just one element, (x_1), which, by definition, is identical to the element itself x_1; besides, we write for brevity $P^n(x_1, x_2, \ldots, x_n)$ instead of (P^n, (x_1, x_2, … , x_n)).

[2] *f is a physical fact* iff:
f is a fact and the predicate of *f* is a member of **Q**.

[3] *f holds in* (*world*) *w* iff:
$f = P^n(x_1, x_2, \ldots, x_n)$ is a fact, $w \in$ **W**, and (x_1, x_2, … , x_n) $\in P^n{}_w$.

We can now define the relation of logical consequence, indicated by the symbol |=, which may hold between any two sets of facts F_1 and F_2. The notation $F_1 \models F_2$ should be read: F_1 logically entails F_2 (or: F_2 is a logical consequence of F_1).

[4] $F_1 \models F_2$ iff:
for any possible world *w*, if for any f_1 member of F_1, f_1 holds in *w*, then for any f_2 member of F_2, f_2 holds in *w*;
in addition, it is intended that the relation of logical consequence is analogously defined for the following cases: $F \models f$ (a set of facts logically entails a fact); $f \models F$ (a fact logically entails a set of facts); $f_1 \models f_2$ (a fact logically entails a fact).

We have thus set up the formal framework that enables us to formulate and discuss the zombie argument. Let us indicate with *F** the set of all physical facts that hold in our world, with *CF** the phenomenal consciousness of our world (*i.e.*, the set of all internal facts of consciousness that hold in our world) and with *CP** the psychological consciousness of our world (*i.e.*, the set of all external facts of consciousness that hold in our world).[14]

The thesis of the zombie argument is thus

[*Z*] $\neg(F^* \models CF^*)$;

moreover, Chalmers also maintains

[*R*] $F^* \models CP^*$,

therefore, [*Z*] and assumption [*R*] also entail the corollary

[Z_1] $\neg(CP^* \models CF^*)$.

Before going on, let me make clear that assumption [*R*] expresses a *minimal* physicalist hypothesis, for it just affirms that psychological consciousness is reductively explainable in terms of physical facts,[15] while it does not say anything about the question of the reductive explainability of phenomenal consciousness.

Let us now state the argument itself:

1. let us consider a world (*zombie world*) in which exactly the physical facts that hold in our world hold, and where, nonetheless, no internal fact of consciousness holds;

2. since such world is well defined, it is obviously conceivable; however, in addition, is it one of the possible worlds in **W**?
3. Sure it is, because the claim of the existence of such possible world undoubtedly seems non-contradictory;
4. but then, if the zombie world is one of the possible worlds, by def. [4], it follows that $\neg(F^* \models CF^*)$ iff CF^* is non-empty;[16]
5. in our world **w*** there are internal facts of consciousness;
6. therefore, by 5 and 4, $\neg(F^* \models CF^*)$ follows.

Q.E.D.

The strength of this form of the zombie argument is that it shifts the burden of the proof on whoever would claim $F^* \models CF^*$. Let us see, exactly, why.

Let me first make clear that this formulation of the argument allows us to sidestep any discussion about the *conceivability* of the zombie world.[17] That such world is conceivable is granted by the fact that it can be adequately defined within the assumed formal framework. To make this point completely clear it is perhaps convenient to explicitly state such definition:

[5] w^Z *is a zombie copy of* w iff:
both w and w^Z are possible worlds and, for any f_1, for any f_2, if f_1 is a physical fact and f_2 is an internal fact of consciousness, then f_2 does not hold in w^Z and (f_1 holds in w^Z iff f_1 holds in w).

The crucial question, thus, is not whether or not a zombie copy of our world is conceivable but, rather, whether such copy does exist.[18] However, since we are concerned here with the purely logical or mathematical sense of existence, such question reduces to verify that the statement of the existence of a zombie world be consistent with the assumed formal principles.[19] And in fact, at least from an intuitive point of view, it is by no means clear how the conjunction of

1. such a statement;
2. the five assumed conditions on the ontology (**D**, **P**, **Q**, **W**, **w***);
3. possibly, further conditions, implicitly or explicitly assumed on such ontology;[20]
4. the usual principles of set theory;

could entail a contradiction.

Be that as it may, even if the conjunction of 1, 2, 3, and 4 should turn out to be contradictory, that is in no way immediately evident. Therefore, whoever would claim the reductive explainability of phenomenal consciousness in terms of physical facts (*i.e.*, $F^* \models CF^*$), should first show

that the appearance of consistency of such conjunction is, in fact, just an appearance.

The zombie argument thus challenges whoever intends to

- construct a theory of phenomenal consciousness,
- which allows us to claim the reductive explainability of phenomenal consciousness in terms of physical facts (*i.e.*, $F^* \models CF^*$).

Chalmers' challenge is

- to show that the statement of the existence of a zombie copy of our world is inconsistent with the conjunction of the principles of the ontology and set theory.

It is my intention to fully accept Chalmers' challenge. Let us then start with the first point, thus laying the foundations of a possible theory of phenomenal consciousness.

5 WHAT IS AN EXPERIENCE?

If our goal is explaining what an *internal fact of consciousness*, or a *conscious experience*, is, it is convenient to deal first with the simpler problem of what an *internal fact*, or an *experience* in general, is.[21]

Let us start with the analysis of a typical example, and ask:

> under what conditions the fact that
> *George is running* = $G(r)$
> can be considered an experience for someone?

It seems quite clear that this question does not have a univocal answer, but that it rather depends on the point of view, or the reference system, which we consider: this fact is an experience for George, but it is not such for me, or for anybody else that is not George.

But what is it that George has of exclusive with respect to his running? Isn't it perhaps, this something proper of George and of nobody else, *his being located internally to such a fact as its subject*? It seems that this is indeed the case. Let us then decide to define:

[6] *f is an experience* (or *an internal fact*) *for x* iff:
f is a fact and *x* is the subject of *f*.

[7] *f is an external fact for x* iff:
f is a fact and *x* is not the subject of *f* and, for some *n* such that $1 \leq n$, *x* is an *n*-tuple of elements of **D**.[22]

Let us notice that, by def. [6], it immediately follows:

Theorem 1 [*of experience privacy*]
For any *x*, for any *y*, for any *f*, if *f* is an experience for *x* and *f* is an experience for *y*, then $x = y$.
Proof
The thesis follows by def. [6] and [1].
Q.E.D.

Let us also define:

[8] *f is an internal fact* (or *an experience*) iff:
there is *x* such that *f* is an internal fact for *x*.

[9] *f is an external fact* iff:
there is *x* such that *f* is an external fact for *x*.

Let us now ask whether facts that are *intrinsically* either internal or external exist, *i.e.*, facts that are internal but not external or external but not internal. The answer is *no*, for the following theorem holds:

Theorem 2 [*of experience relativity*]
Every fact is both internal and external.
Proof
Let *f* be an arbitrary fact and *x* be its subject. Then, by def. [6], *f* is an internal fact for *x*; hence, by def. [8], *f* is an internal fact. Moreover, by def. [7], *f* is an external fact for the pair (*x*, *x*); therefore, by def. [9], *f* is an external fact.
Q.E.D.

6 TWO DIFFICULTIES WITH THE DEFINITION OF EXPERIENCE

In the previous section, I proposed to analyze the general concept of experience by means of the idea that a fact turns out to be an experience for a given entity *x* depending on whether or not *x* is located internally to such a fact as its subject. (def. [6]).

This definition seems to work quite well for the case of monadic facts, *i.e.*, facts whose predicate is a one place property. However, if we also consider relational facts (*i.e.*, facts whose predicate is an *n*-ary property,

where $2 \leq n$) we immediately face a difficulty, which I call of *absent experiences*.

In addition, by applying def. [6] to a few more examples, we also face some cases of experiences that, *prima facie*, don't seem to agree with our intuitions. This is the *difficulty of strange experiences*.

This section is dedicated to the exposition and solution of both difficulties.

6.1 *ABSENT EXPERIENCES*

Let us consider the relational fact:

George loves Mary = $L(g, m)$

According to def. [6], this fact is an experience *exclusively* for the pair (g, m), which is its subject. Intuitively, however, it seems that both George and Mary, as *individuals*, should in some sense experience this fact, for they are both located internally to it.

On the other hand, it is clear that George and Mary cannot both experience the very same fact, for this is excluded by the theorem of experience privacy. But then, what are Mary and George, individually, experiencing?

The intuitive solution to this difficulty is that each of them experiences *his own* fact, different from the other's, and also different from the one relative to the pair George and Mary. However, these three facts turn out to be logically equivalent.

More precisely, this means that, in the first place, the *pair George and Mary* experiences the fact that

the relation x loves y holds for <u>the pair George and Mary</u>;

second, *George* experiences the fact that

<u>George</u> loves Mary;

and, third, *Mary* experiences the fact that

<u>Mary</u> is loved by George.

Therefore, according to this analysis, there are three different experienced facts. We can represent all of them in a uniform manner by using a single line to indicate their *logical subject*:

1. *George loves Mary*
2. *George loves Mary*
3. *George loves Mary*

It is also obvious that, from an intuitive point of view, all the three facts turn out to be logically equivalent.

However, the formal framework developed so far is not sufficient to adequately represent the *specific* logical form of each of the facts 1, 2, 3, and their mutual relationships. The problem thus arises to extend such formal framework, so as to be able to translate into it the proposed intuitive solution.

Surprisingly enough, the formal tool that allows us to carry out this translation in a quite natural manner is the λ-calculus. By employing the formalism of the λ-calculus we can in fact generate facts 2 and 3 from fact 1. Let us see how.

First, let us formally represent fact 1:

1. *George loves Mary* = $L(g, m)$

This is the starting fact, from which we get the other two by means of the two operations of λ-abstraction and instantiation.

Actually, by λ-abstracting on *g*, we get the property:

$$x \text{ loves Mary} = [\lambda(g)L(g, m)]$$

and then, by instantiating this property with *g* itself, we get:

2. *George loves Mary* = $[\lambda(g)L(g, m)](g)$

Analogously, by λ-abstracting on *m*, we get the property:

$$\text{George loves } y = [\lambda(m)L(g, m)]$$

and, by instantiating this property with *m* itself, we get:

3. *George loves Mary* = $[\lambda(m)L(g, m)](m)$

The usual axioms of the λ-calculus precisely ensure that (i) the λ-abstraction operation is all right, and that (ii) 1, 2 and 3 are logically equivalent.

To adapt such axioms to the specific formalism employed here, let us extend the ontology first, by adding **λ** as a new primitive entity. From the set theoretical point of view, the **λ** operator can be thought as a two place partial function, whose codomain is **P**, whose first domain is the union of all *k*-th Cartesian products of **D** (for $1 \leq k$), and whose second domain is the set of all possible facts of the structure. Let us then assume the following two new axioms on the extended ontology (**D**, **P**, **Q**, **W**, **w***, **λ**):

6. [*closure of* **P** *with respect to* **λ**-*abstraction of properties from facts*]
λ is a two place partial function, whose codomain is **P**, whose first domain is the union of all *k*-th Cartesian products of **D** (for $1 \leq k$), and whose second domain is the set of all possible facts of the structure; $\boldsymbol{\lambda}((x_{i,1}, x_{i,2}, \ldots, x_{i,k}), P^n(x_1, x_2, \ldots, x_n))$ is defined iff $P^n(x_1, x_2, \ldots, x_n)$ is an arbitrary fact and $(x_{i,1}, x_{i,2}, \ldots, x_{i,k})$ is a *k*-tuple of exactly *k* distinct elements of **D**, each of which occurs in the *n*-tuple $(x_1, x_2, \ldots, x_n)$; if $\boldsymbol{\lambda}((x_{i,1}, x_{i,2}, \ldots, x_{i,k}), P^n(x_1, x_2, \ldots, x_n))$ is defined, it is a *k*-ary property member of **P**; in addition, let us shorten $\boldsymbol{\lambda}((x_{i,1}, x_{i,2}, \ldots, x_{i,k}), P^n(x_1, x_2, \ldots, x_n))$ as $[\boldsymbol{\lambda}(x_{i,1}, x_{i,2}, \ldots, x_{i,k})P^n(x_1, x_2, \ldots, x_n)]$;

Before stating the second axiom relative to **λ**, let us bring in the following definition:

[10] / is a partial three place function, whose codomain and whose three domains are, each, the union of all *n*-th Cartesian products of **D** (for $1 \leq n$); the values of / are defined as follows:
if $(x_1, x_2, \ldots, x_n)$ is an *n*-tuple ($1 \leq n$) of elements of **D**, $(x_{i,1}, x_{i,2}, \ldots, x_{i,k})$ is a *k*-tuple ($1 \leq k \leq n$) of exactly *k* distinct elements of **D**, each of which occurs in $(x_1, x_2, \ldots, x_n)$, and $(y_1, y_2, \ldots, y_k)$ is a *k*-tuple of elements (not necessarily distinct) of **D**, then $/((x_1, x_2, \ldots, x_n), (x_{i,1}, x_{i,2}, \ldots, x_{i,k}), (y_1, y_2, \ldots, y_k))$ is defined, and it is equal to the *n*-tuple of elements of **D** obtained by substituting the element y_j ($1 \leq j \leq k$) for all occurrences of element $x_{i,j}$ in the *n*-tuple $(x_1, x_2, \ldots, x_n)$; otherwise, the value of / is not defined; let us also stipulate that $[(x_1, x_2, \ldots, x_n) / (x_{i,1} / y_1, x_{i,2} / y_2, \ldots, x_{i,k} / y_k)]$ is an abbreviation for $/((x_1, x_2, \ldots, x_n), (x_{i,1}, x_{i,2}, \ldots, x_{i,k}), (y_1, y_2, \ldots, y_k))$.

We can now state the second axiom relative to **λ**:

7. [*conditions of validity of the instantiation of a property obtained by* **λ**-*abstraction*]
if $(x_1, x_2, \dots, x_n)$ is an n-tuple ($1 \leq n$) of elements of **D**, $(x_{i,1}, x_{i,2}, \dots, x_{i,k})$ is a k-tuple ($1 \leq k \leq n$) of exactly k distinct elements of **D**, each of which occurs in $(x_1, x_2, \dots, x_n)$, and $(y_1, y_2, \dots, y_k)$ is a k-tuple of elements (not necessarily distinct) of **D**, then, for any possible world w, $[\lambda(x_{i,1}, x_{i,2}, \dots, x_{i,k})P^n(x_1, x_2, \dots, x_n)](y_1, y_2, \dots, y_k)$ holds in w iff $P^n[(x_1, x_2, \dots, x_n) / (x_{i,1} / y_1, x_{i,2} / y_2, \dots, x_{i,k} / y_k)]$ holds in w.

Keeping in mind the line of reasoning shown above, let us finally notice that the two new axioms on the ontology (axioms **6** and **7**) allow us to solve in an elegant and altogether natural way the difficulty of absent experiences.

6.2 *STRANGE EXPERIENCES*

Let us consider the fact:

the train is running = $R(t)$

By def. [6], it turns out that such fact is an experience for the train. But how is it possible to truly say that a *train* experiences its running?

The same difficulty arises also with respect to any relational fact. For example, let us consider again the fact:

George loves Mary = $L(g, m)$

According to def. [6], such fact is an experience for the couple George and Mary. But, in this case too, it apparently seems quite strange to affirm that a *couple* experiences its own loving relationship.

My answer to this type of difficulty is that it is just apparent. That the train experiences its running seems to be paradoxical because usually we do not distinguish between an experience and the experience of being conscious of this experience.

Now, that $R(t)$ is a fact directly experienced by the train is not paradoxical at all because, according to def. [6], this just means that the train is located internally to such fact, in the role of subject. However, it would be surely paradoxical to affirm that the train experiences the fact and that, *in addition*, it is also conscious of such experience. The point is that *every thing has experiences*, but just *a few things have conscious experiences* (and, among these, neither trains nor couples are included).[23]

Hence, to begin with, we need an important distinction: experience and consciousness are not the same, in the sense that there are experiences that are not an object of consciousness for their subject (ex_1: the train is not conscious of *its running*). Analogously, there are also external facts of which one does not have consciousness (ex_2: since George is sleeping, George is not conscious that *the traffic light is red*). And, finally, there can be both experiences and external facts of which one is conscious (ex_3: George is conscious of *his running*; ex_4: George is conscious that *Mary is speaking*).

From a different point of view, however, consciousness and experience cannot be disjoint: this is the case of all the internal facts of consciousness (*i.e.*, conscious *experiences*) all of which are a particular type of experience. But, exactly, what type of experience are conscious experiences?

7 WHAT IS A CONSCIOUS EXPERIENCE?

My answer to this question is that conscious experiences, in so far as they are experiences, are first of all, by def. [6], facts. However, they are a particular type of facts, which I call *facts of consciousness*. And what is a fact of consciousness?

To answer this second question, let us consider a few typical examples. The following are all facts of consciousness:

1. *George is conscious of his running*
2. *George is conscious that Mary is speaking*
3. *Mary is conscious of her speaking*
4. *Mary is conscious that George is running*

By observing the examples, we notice:

(a) in each of them, the binary relation *x is conscious of* f_0 occurs; this is a relation between an element *x* of the domain and a fact f_0;
(b) the *subject* of each fact of consciousness is *x* (underlined in the examples);
(c) the fact f_0 is the *object* or the *content* of the fact of consciousness;
(d) facts 1 and 3 are similar, because, in both of them, the subject of the fact of consciousness and the subject of its content are identical; facts like these can thus be called *facts of internal consciousness*, or *facts of self-consciousness*;

(e) in facts 2 and 4, in contrast, the subject of the fact and the subject of its content are different; facts like these can thus be said *facts of external consciousness*.

First, on the basis of observation (a), we can further extend the ontology, by adding a special binary relation, **C**, which may only hold between an element of the domain and a fact. Intuitively, **C** is to be identified with the relationship *being conscious of*. Let us thus assume the following axiom on the ontology (**D**, **P**, **Q**, **W**, **w***, **λ**, **C**):

8. [*axiom of consciousness*]
for any fact f, $f \in \mathbf{D}$,[24] $\mathbf{C} \in \mathbf{P}$, **C** is a two place property and, for any w, for any x, for any f_0, if $\mathbf{C}(x, f_0)$ holds in w, then f_0 is a fact.

Second, on the basis of observations (b) and (c), we can define a fact of consciousness in the following way, by means of **λ**-abstraction and axiom **8**:

[11] *f is a fact of consciousness* iff:
there is x, there is f_0 such that $f = [\boldsymbol{\lambda}(x)\mathbf{C}(x, f_0)](x)$;
f_0 is called the *object* or the *content* of fact of consciousness f.

Third, on the basis of observations (d) and (e), let us also define:

[12] *f is a fact of internal consciousness* (or *f is a fact of self-consciousness*) iff:
f is a fact of consciousness and the subject of *f* is identical to the subject of its content.

[13] *f is a fact of external consciousness* iff:
f is a fact of consciousness and the subject of *f* is different from the subject of its content.

On the basis of definitions [6] and [11], we can finally define a conscious experience for *x*:

[14] *f is an internal fact of consciousness for x* (or *f is a conscious experience for x*) iff:
f is an experience for *x* and *f* is a fact of consciousness.[25]

Analogously, on the basis of definitions [7] and [11], we define a conscious function for *x*:

[15] *f is an external fact of consciousness for x* (or *f is a conscious function for x*) iff:
f is an external fact for *x* and *f* is a fact of consciousness.

Finally, by first conjoining def. [14], and then def. [15], with either def. [12] or [13], we also get definitions for the following concepts:

[16] *f is an internal fact of internal consciousness for x* (or *f is a self-conscious experience for x*) iff:
f is an internal fact of consciousness for *x* and *f* is a fact of internal consciousness;

[17] *f is an internal fact of external consciousness for x* (or *f is an external conscious experience for x*) iff:
f is an internal fact of consciousness for *x* and *f* is a fact of external consciousness;

[18] *f is an external fact of internal consciousness for x* (or *f is a self-conscious function for x*) iff:
f is an external fact of consciousness for x and *f* is a fact of internal consciousness;

[19] *f is an external fact of external consciousness for x* (or *f is an external conscious function for x*) iff:
f is an external fact of consciousness for *x* and *f* is fact of external consciousness.

Examples
(i) examples 1 and 3 above are facts of internal consciousness;
(ii) examples 2 and 4 above are facts of external consciousness;
(iii) example 1 is an internal fact of internal consciousness for George and an external fact of internal consciousness for Mary;
(iv) example 3 is an internal fact of internal consciousness for Mary and an external fact of internal consciousness for George;
(v) example 2 is an internal fact of external consciousness for George and an external fact of external consciousness for Mary;
(vi) example 4 is an internal fact of external consciousness for Mary and an external fact of external consciousness for George.

8 PROPOSAL FOR AN ANALYSIS OF THE PARADIGMATIC CASE OF CONSCIOUS EXPERIENCE – THE EXPERIENCE OF QUALIA

To test whether the definitions of the preceding section are adequate, let us now see whether or not they allow us to analyze the paradigmatic case of conscious experience, namely, the experience of *qualia* or *immediately felt qualities*. Let us then consider the following experience of an immediately felt quality:

Mary's conscious experience, who is feeling the pain of a hammer blow in her toe = ?

My analysis proposal is as follows. The hammer blow causes the fact that

Mary is in a particular neuro-physiological state d = $[\boldsymbol{\lambda}(m)S(m, d)](m) = f_d$

The subject of this fact is Mary herself; hence, by def. [6], it is an experience for Mary.

In turn, the fact that Mary is in neuro-physiological state *d* causes, under normal conditions,[26] the fact that

Mary is conscious of her being in state neuro-physiological d

The subject of this fact is Mary as well. Such fact can thus be analyzed as:

Mary is conscious of her being in state neuro-physiological d =
= $[\boldsymbol{\lambda}(m)\mathbf{C}(m, f_d)](m) = f$

Therefore, by def. [16], *f* is a self-conscious experience for Mary, whose content is her being in state *d*, which is caused by the hammer blow (*i.e.*, the content of Mary's conscious experience *f* is her experience f_d, which is caused by the hammer blow). I thus propose to identify *f* with *?* .[27]

Finally, let us also notice that the same kind of analysis can be applied to *any* qualia experience described by scheme (Q) (see sec. 2).

9 THEORY OF PHENOMENAL CONSCIOUSNESS AND PHYSICALISM

In the previous sections (sec. 5-8) I laid the foundations of a possible theory of phenomenal consciousness. The goal of sec. 9.1 is twofold: (i) to insert such a theory within a minimal physicalist framework (expressed by axiom **9**, or **9.1**, or **9.2**) and then (ii) to accept Chalmers' challenge (see sec. 4), and thus prove the physical reductive explainability of conscious experiences (theorem 4 or 4.1) and, as a consequence, the logical impossibility of zombie worlds (theorem 5 or 5.1). Finally, after introducing the concept of an *angelic copy*[28] of a possible world, I will prove the logical impossibility of angelic worlds too (theorem 6.2).

In sec. 9.2, I will show how, by assuming a stronger physicalist hypothesis (axiom **9.3**), we can obtain all the results of sec. 9.1 as particular cases.

9.1 *THE PHYSICAL REDUCTIVE EXPLAINABILITY OF CONSCIOUS EXPERIENCE AND THE LOGICAL IMPOSSIBILITY OF EITHER ZOMBIE OR ANGELIC WORLDS*

In order to deal with these questions, let us also define the absolute concepts of conscious experience and conscious function:

[20] *f is an internal fact of consciousness* (or *f is a conscious experience*) iff:
there is *x* such that *f* is an internal fact of consciousness for *x*.

[21] *f is an external fact of consciousness* (or *f is a conscious function*) iff:
there is *x* such that *f* is an external fact of consciousness for *x*.

Let us now notice that, by def. [20] and [21], the theorem of experience relativity (theorem 2) also applies to conscious experiences. In other words, the following theorem holds:

Theorem 3 [*of real identity between conscious experiences and conscious functions*]
For any *f*, *f* is a conscious experience iff *f* is a conscious function.

Proof
From def. [20], [21], [14], [15], [8], [9], and theorem 2, the thesis follows.
Q.E.D.

This theorem allows us to answer the question of whether or not the distinction between conscious experiences and conscious functions is real (see sec. 2). The answer is *no*. Conscious experiences and conscious functions are two different concepts, but they nonetheless describe the same set of facts (namely, all the facts of consciousness).

Consequently, any argument that allegedly proves the opposite is either invalid or contains some false premise. This, in particular, holds for all classic arguments mentioned above (see sec. 2).

Let us now see the consequences of this fact for the zombie argument. If Chalmers' argument is sound (valid and with all premises true), it proves that the conscious experiences of our world are not reductively explainable by the physical facts of our world (thesis [*Z*], sec. 2). Moreover, [*Z*] and the further hypothesis that the conscious functions of our world are reductively explainable by the physical facts of our world (assumption [*R*], sec. 2) entail that the conscious experiences of our world are not reductively explainable by means of the conscious functions of our world. (corollary [Z_1], sec. 2). But this contradicts theorem 3. Hence, for Chalmers, only two alternatives are available:

Chalmers' dilemma (*first formulation*)

1. if Chalmers' argument is sound, then, contrary to his claim, the conscious functions of our world are not reductively explainable by the physical facts of our world (*i.e.*, assumption [*R*] is false);
2. on the other hand, if the conscious functions of our world (and thus, by theorem 3, also the conscious experiences of our world) are reductively explainable by the physical facts of our world, then the zombie argument is not sound (*i.e.*, either it is invalid or it contains some false premise).

Chalmers' dilemma can be put into an even more cogent form. We have seen that the zombie argument has just two premises:

[E_1] there is $\mathbf{w}^{*z}$ such that $\mathbf{w}^{*z}$ is a zombie copy of our world $\mathbf{w}^*$;

[E_2] CF^* is not empty.

Thesis [*Z*] then follows from [E_1], [E_2], and the definition of logical consequence between sets of facts (def. [3]). Therefore, the zombie argument is undoubtedly valid. Also, the truth of premise [E_2] is not presumably in question. It thus follows that Chalmers' dilemma has a second formulation, equivalent to the first one:

Chalmers' dilemma (*second formulation*)

1. if a zombie copy our world exists, then, contrary to Chalmers' claim, the conscious functions of our world are not reductively explainable by the physical facts of our world (*i.e.*, assumption [*R*] is false);
2. on the other hand, if the conscious functions of our world (and thus, by theorem 3, also the conscious experiences of our world) are reductively explainable by the physical facts of our world, then a zombie copy of our world does not exist.

We have seen (sec. 2) that Chalmers maintains the reductive explainability of conscious functions by the physical facts of our world (assumption [Z]). But then, by assuming such hypothesis as a further axiom of our ontology, the second formulation of Chalmers' dilemma yields an actual *proof* of the non-existence of a zombie copy of our world. Below are the details.

Let us take assumption [*Z*] as a new axiom on the ontology (**D**, **P**, **Q**, **W**, $\mathbf{w}^*$, $\boldsymbol{\lambda}$, **C**):

9. [*minimal physicalist hypothesis relative to our world*]
$F^* \models CP^*$.

Theorem 3, in conjunction with axiom 9, allows us to prove the physical reductive explainability of *phenomenal* consciousness in our world:

Theorem 4 [*of physical reductive explainability of phenomenal consciousness in our world*]
$F^* \models CF^*$.

Proof
By axiom **9**, $F^* \models CP^*$ and , by theorem 3, $CP^* = CF^*$. Hence, $F^* \models CF^*$.
Q.E.D.

It then follows:

Theorem 5 [*of logical impossibility of a zombie copy of our world*]
There is no $\mathbf{w}^{*z}$ such that $\mathbf{w}^{*z}$ is a zombie copy of our world **w*** iff there is *f* such that $f \in CF^*$.

Proof
Let us first prove the thesis from left to right, and then from right to left.

Thesis from left to right
Let us prove the contrapositive. Thus, let us assume that there is no *f* such that $f \in CF^*$, and show that there is $\mathbf{w}^{*z}$ such that $\mathbf{w}^{*z}$ is a zombie copy of **w***. By the assumed hypothesis and def. [5], **w*** is a zombie copy of **w***. But then the thesis holds for $\mathbf{w}^{*z} = \mathbf{w}^*$.

Thesis from right to left
Let us assume that CF^* is not empty. Let us suppose, for *reductio*, that there is $\mathbf{w}^{*z}$ such that $\mathbf{w}^{*z}$ is a zombie copy of **w***. By the *reductio* hypothesis, since CF^* is not empty, and by def. [5] and [4] of zombie copy and logical consequence, $\neg(F^* \models CF^*)$. But, by theorem 4, $F^* \models CF^*$. From the *reductio* hypothesis a contradiction thus follows.
Q.E.D.

Note that we could replace axiom **9** with a similar axiom that refers to another fixed possible world *w*, not to our world **w*** ; or we could even add a corresponding axiom for all those possible worlds for which we believe that the hypothesis of the physical reductive explainability of psychological consciousness is true. In this case too, we could obviously prove the analogs of theorems 4 and 5 relative to each of these worlds.

Let us indicate with F_w the set of all physical facts that hold in a fixed world *w*, with CP_w the psychological consciousness of *w* (*i.e.*, the set of all external facts of consciousness that hold in *w*) and with CF_w the phenomenal consciousness of *w* (*i.e.*, the set of all internal facts of consciousness that hold in *w*). The version of axiom **9** for any fixed world *w* thus is:

9.1 [*local minimal physicalist hypothesis*]
$F_w \models CP_w$

from which we obtain:

Theorem 4.1 [*of physical reductive explainability of phenomenal consciousness in a fixed possible world*]
$F_w \models CF_w$.

Proof
Identical to the proof of theorem 4, after replacing "axiom **9**" with "axiom **9.1**", "F^*" with "F_w" and "CF^*" with "CF_w".
Q.E.D.

Let us then prove the analog of theorem 5 for a fixed possible world *w*.

Theorem 5.1 [*of logical impossibility of a zombie copy of a fixed possible world*]
There is no w^z such that w^z is a zombie copy of *w* iff there is *f* such that $f \in CF_w$.

Proof
Identical to the proof of theorem 5, after replacing "**w***" with "*w*", "$\mathbf{w}^{*z}$" with "w^z", "theorem 4" with "theorem 4.1", "F^*" with "F_w", and "CF^*" with "CF_w".
Q.E.D.

It has been noticed that the idea of a zombie world is in full agreement with a Cartesian vision, according to which there are two independent ontological realms – thought and matter. The zombie copy of a fixed world *w* can in fact be viewed as pure matter, separated from thought, whose existence, even if not actual, would nonetheless be a logical possibility. But we have just seen that, if we assume a rather weak physicalist hypothesis relative to such world (*i.e.*, the facts of the *psychological* consciousness of *w* are reductively explainable by its physics), this possibility must be ruled out.

What about the other side of Cartesian dualism, namely, pure thought? Just like we imagined a zombie copy, we can imagine an *angelic copy* of a fixed world *w* – a world entirely identical to *w* with respect to the internal

facts of consciousness, but where all the physical facts are lacking. Among the possible worlds of our ontology, is there an angelic copy of w? It seems that, for a negative answer to this question, the *local* minimal physicalist hypothesis (axiom **9.1**) is not sufficient. However, if we introduce a *global* version (axiom **9.2**, below) of such hypothesis, *i.e.*, generalized to all possible worlds, we can prove that there is no angelic copy of w if, in CF_w, there is at least one contingent fact, *i.e.*, a fact that does not hold in all possible worlds.

Let us define:

[22] w^a *is an angelic copy of* w iff:
both w and w^a are possible world and, for any f_1, for any f_2, if f_1 is a physical fact and f_2 is an internal fact of consciousness, then f_1 does not hold in w^a and (f_2 holds in w^a iff f_2 holds in w).

In place of axiom **9.1** (or axiom **9**), let us assume the following condition on ontology (**D**, **P**, **Q**, **W**, **w***, **λ**, **C**):

9.2 [*global minimal physicalist hypothesis*]
for any possible world w, $F_w \models CP_w$.

Let w be a fixed possible world; we prove:

Theorem 6.2 [*of logical impossibility of an angelic copy of a fixed possible world*]
If there is f such that $f \in CF_w$ and f does not hold in all possible worlds,[29] then there is no w^a such that w^a is an angelic copy of w.

Proof
Let us assume that there is $f_\#$ such that $f_\# \in CF_w$ and $f_\#$ does not hold in all possible worlds. Let us suppose, for *reductio*, that there is w^a such that w^a is an angelic copy of w. Since $f_\# \in CF_w$ and w^a is an angelic copy of w, $f_\# \in CF_w{}^a$; hence, by theorem 3, $f_\# \in CP_w{}^a$. Since w^a is an angelic copy of w, $F_w{}^a$ is empty and thus, obviously, for any w, for any $f \in F_w{}^a$, f holds in w; from this, by def. [4], it follows that $F_w{}^a \models CP_w{}^a$ iff for any w, for any $f \in CP_w{}^a$, f holds in w. But $f_\# \in CP_w{}^a$ and $f_\#$ does not hold in all possible worlds. Therefore, $\neg(F_w{}^a \models CP_w{}^a)$. But, by axiom **9.2**, $F_w{}^a \models CP_w{}^a$. From the hypothesis of *reductio*, a contradiction thus follows.
Q.E.D.

The aim of next section is to show that, by assuming a further axiom that expresses in a quite natural way a strong physicalist hypothesis,[30] we can prove the global minimal physicalist hypothesis (axiom **9.2**); all the preceding results then obviously follow from it, *i.e.*, theorems 4, 4.1, 5, 5.1 and 6.2.

9.2 *THE STRONG PHYSICALIST HYPOTHESIS AND THE THEOREM OF PHYSICAL REDUCIBILITY*

The result just mentioned will be obtained as a consequence of a more general theorem. Such theorem affirms that any fact has a corresponding physical fact that turns out to be logically equivalent to it (theorem 7.3, below). This theorem can thus be interpreted as a genuine *reducibility* theorem of an arbitrary fact to its physical equivalent.

The axiom that allows the proof of theorem 7.3 expresses the idea that, in our ontology (**D**, **P**, **Q**, **W**, **w***, **λ**, **C**), either each property is a physical property, or it is obtainable from a physical property by repeated applications of the operations of instantiation and **λ**-abstraction. This assumption (axiom **9.3**, below), according to which all properties ultimately are physical properties, turns out to be a strong physicalist hypothesis. For, by theorem 7.3, it obviously follows that any fact is reductively explainable by the set of all physical facts.

In place of axiom **9.2** (or axiom **9.1**, or **9**) let us assume the following condition on ontology (**D**, **P**, **Q**, **W**, **w***, **λ**, **C**):

9.3 [*strong physicalist hypothesis*]
for any property $P^k \in \mathbf{P}$, either $P^k \in \mathbf{Q}$, or P^k is obtained from a property $P^n \in \mathbf{Q}$ by a finite number $m \geq 1$ of successive applications of the two operations of instantiation and **λ**-abstraction (in this order).

Let us also give the following definition:

[23] f_1 *is logically equivalent to* f_2 iff:
for any possible world w, f_1 holds in w iff f_2 holds in w.

This definition and axiom **9.3** allow us to prove that any fact is logically equivalent to an appropriate physical fact (theorem 7.3 below).

Theorem 7.3 [*of physical reducibility*]
For any fact f, there is a physical fact $f_\#$ such that f is logically equivalent to $f_\#$.

Proof
By axiom **9.3**, we may indicate any property $P^k \in \mathbf{P}$ by a symbol of the kind $P^{n(p)}$, where $p \geq 0$, $n(p)$ is the arity of property P^k (thus, $n(p) = k$) and index p has the following meaning:

- if $p = 0$, $P^{n(p)} \in \mathbf{Q}$;
- if $p \geq 1$, the property $P^{n(p)}$ is obtained from a property $P^{n} \in \mathbf{Q}$ by p successive applications of the operations of instantiation and $\boldsymbol{\lambda}$-abstraction (in this order).

Let $P^{k} = P^{n(p)}$ be the predicate of f; let us prove the theorem by induction on p.

Base case
assume:
1. $p = 0$;
from the hypothesis 1 and the definition of $P^{n(p)}$:
2. $P^{n(0)} \in \mathbf{Q}$;
from 2 and def. [2]:
3. f is a physical fact;
hence, from 3, by setting $f_{\#} = f$, the thesis holds.

Recursive step
Let us suppose that:
1. the thesis holds for any fact f whose predicate can be indicated by a symbol of the kind $P^{n(p)}$;

we will prove the thesis for any fact whose predicate can be indicated by a symbol of the kind $P^{n(p+1)}$. By the definition of $P^{n(p+1)}$:
2. $P^{n(p+1)} = [\boldsymbol{\lambda}(x_{i,1}, x_{i,2}, \ldots, x_{i,n(p+1)})P^{n(p)}(x_1, x_2, \ldots, x_{n(p)})]$;
from 2 and axiom 5:
3. $P^{n(p+1)}(y_1, y_2, \ldots, y_{n(p+1)})$ holds in w iff $P^{n(p)}[(x_1, x_2, \ldots, x_{n(p)}) / (x_{i,1} / y_1, x_{i,2} / y_2, \ldots, x_{i,n(p+1)} / y_{n(p+1)})]$ holds in w;
from 3, hypothesis 1, and def. [23]:
4. there is a physical fact $f_{\#}$ such that $P^{n(p)}[(x_1, x_2, \ldots, x_{n(p)}) / (x_{i,1} / y_1, x_{i,2} / y_2, \ldots, x_{i,n(p+1)} / y_{n(p+1)})]$ holds in w iff $f_{\#}$ holds in w;
from 4 and 3:
5. $P^{n(p+1)}(y_1, y_2, \ldots, y_{n(p+1)})$ holds in w iff $f_{\#}$ holds in w;

the thesis is thus proved for any fact whose predicate can be indicated by a symbol of the kind $P^{n(p+1)}$.
Q.E.D.

From theorem 7.3, the global minimal physicalist hypothesis (*i.e.*, axiom **9.2**) easily follows:

Corollary 1.7.3 [*global minimal physicalist hypothesis*]
For any w, $F_w \models CP_w$.

Proof
If CP_w is empty, the thesis immediately follows from the definition of logical consequence (def. [4]). Let us thus assume that f is an arbitrary fact member of CP_w. Then, by theorem 7.3 and def. [23], there is a physical fact $f_{\#}$ such that, for any w, f holds in w iff $f_{\#}$ holds in w. From this, since

$f \in CP_w$, and by def. [4] of logical consequence, it follows that $f_\# \in F_w$ and $f_\# \models f$. Hence, $F_w \models f_\#$, and thus, by transitivity, $F_w \models f$.
Q.E.D.

Note that, from corollary 1.7.3, theorem 6.2 obviously follows. Furthermore, the local minimal physicalist hypothesis relative to an arbitrary fixed world *w* follows too (axiom **9.1**). Therefore, also theorems 4, 4.1, 5, and 5.1 then follow.

Let us finally observe that theorem 7.3 also allows us to prove a somewhat stronger version of theorem 6.2. In this version, we no longer require that at least one internal fact of consciousness of *w* be contingent.

Corollary 2.7.3 [*strong theorem of logical impossibility of an angelic copy of a fixed possible world*]
If there is *f* such that $f \in CF_w$,[31] then there is no w^a such that w^a is an angelic copy of *w*.

Proof
Assume that there is *f* such that $f \in CF_w$, and suppose for *reductio* that there is *w* and there is w^a such that w^a is an angelic copy of *w*. By the *reductio* hypothesis and def. [22] of angelic copy, *f* holds in w^a. By theorem 7.3, there is a physical fact $f_\#$ such that, for any w_1, *f* holds in w_1 iff $f_\#$ holds in w_1. Therefore, since *f* holds in w^a, $f_\#$ holds in w^a. But, since $f_\#$ is a physical fact, by def. [22], $f_\#$ does not hold in w^a. From the hypothesis of *reductio* a contradiction thus follows.
Q.E.D.

10 Conclusion: The hard problem is a philosophical problem, not a scientific one

Chalmers' intuition is that the phenomenal or subjective side of the problem of consciousness will necessarily be lost as soon as the question is put within a framework of a physicalist kind. Physics, the paradigmatic instance of objective science, would allow us to cope with just the external, or third person, aspects of consciousness; but its deeper and more intriguing aspect, namely, the subjective, internal, or experiential element, would definitely fall beyond the scope of any physical theory.

The present inquiry yields the conclusion that this intuition is basically ungrounded. For we have seen that (i) it is possible to develop a theory of phenomenal consciousness (sec. 5-8); (ii) the explanation of phenomenal consciousness provided by such theory is by no means incompatible with a physicalist standpoint; rather, once it is inserted within a minimal

physicalist framework (expressed by axiom **9**, or **9.1**, or **9.2**, sec. 9.2), it allows us to prove the physical reductive explainability of conscious experiences (theorem 4 or 4.1) and, consequently, the logical impossibility of zombie worlds (theorem 5 or 5.1).

However, we must concede that this conclusion does not cover all the important aspects of the problem of phenomenal consciousness. In particular, there is still a quite subtle nonetheless crucial question, which only now, when we have almost completed our inquiry, can we put in a clear form.

We have seen that, according to Chalmers, the hard problem consists in *finding and developing an adequate theory of phenomenal consciousness*. However, the crucial question that we should ask with respect to such problem is the following; what is the *kind* of the theory we are looking for? More precisely, are we looking for a theory that allows us to give a *scientific explanation* of the facts of phenomenal consciousness, or a theory that provides a philosophical explanation, or an *explication*, of the concept of phenomenal consciousness?

Even though, as far as I know, Chalmers has never explicitly dealt with this question, it is nonetheless clear that he leans towards the first alternative. My conviction, on the contrary, is that the hard problem is, in the first place, a *philosophical problem*, that is to say, a problem of conceptual clarification and delimitation, and that solving this problem is a necessary condition for then posing in the appropriate terms also the problem of a scientific explanation of the facts of phenomenal consciousness.

But then, if this is so, we can also understand why the question of phenomenal consciousness appears to Chalmers to be intractable from the standpoint of physics or, more generally, of any other discipline that shares with it the same scientific blueprint. For, if the hard problem is a philosophical problem of *conceptual explication*, and it is not a scientific problem of *explanation of facts of a given kind*, it is by no means surprising that physics, or any other science, turns out to be utterly inadequate to solve it. For the required theory is not scientific, it is rather a philosophical one.

This work has attempted to provide a first version of such philosophical theory. Should it turn out to be adequate, the limits within which then construct scientific explanations of specific facts of consciousness would be drawn.

Marco Giunti
Università di Cagliari
Dipartimento di Scienze Pedagogiche e Filosofiche
Via Is Mirrionis 1
09123 Cagliari

APPENDIX: CONSISTENCY OF THE AXIOMATIC THEORY

The first clause of axiom **8**, which requires that any fact *f* be a member of the domain, brings in some self-reference within the axiomatic theory. For, by def. [1], *f*, which is a member of the domain, turns out to be an element of the subject of other facts, which are members of the domain as well. Hence, for any *f*, it makes sense to ask whether or not *f* is an element of its own subject. We could then suspect that this situation allows the construction of an antinomy of the kind of Russell's.[32] For this reason, it becomes important to prove that the theory consisting of axioms **1** - **9.3** has a set theoretical model, and it is thus consistent with set theory.[33]

The models of the axiomatic theory are all the structures (**D**, **P**, **Q**, **W**, **w***, **λ**, **C**) that satisfy axioms **1** - **9.3**. The structure defined below turns out to be a model of the theory.

Let $\mathbf{P} = \{P^1, C^2\}$, where $P^1 = (\varnothing, 1)$ and $C^2 = (\varnothing, 2)$;
let $\mathbf{Q} = \{C^2\}$;
let $\mathbf{W} = \{\mathbf{w}^*\}$, where $\mathbf{w}^*$ is the function on **P** defined by: $\mathbf{w}^*(P^1) = \varnothing$ and $\mathbf{w}^*(C^2) = \varnothing$;
let **D** = the union of all *n*-th domains Δ_n, where, for any $n \geq 0$, domain Δ_n is inductively defined as follows:

$\Delta_o = \{\varnothing\}$;
suppose Δ_n has been defined, and let us define Δ_{n+1}:
$\Phi_n = \{f$: *f* is an ordered pair of the kind $(P^n, (x_1, x_2, \ldots, x_n))$, where $P^n \in \mathbf{P}$ and $(x_1, x_2, \ldots, x_n)$ is an *n*-tuple ($n \geq 1$) of elements of $\Delta_n\}$; let us write for brevity $P^n(x_1, x_2, \ldots, x_n)$ in place of $(P^n, (x_1, x_2, \ldots, x_n))$;
Δ_{n+1} = the union of Δ_n and Φ_n.

Let us immediately notice that, by the definitions of **D**, **P**, **Q**, **W**, and **w***, axioms **1**, **2**, **3**, **4**, and **5** are obviously satisfied.

Let us recall that, from the set theoretical point of view, the **λ** operator can be thought as a two-place partial function, whose codomain is **P**, whose first domain is the union of all *k*-th Cartesian products of **D** (for $1 \leq k$), and whose second domain is the set of all the possible facts of the structure (see def. [1]). Let us define such function as follows:

if $P^n(x_1, x_2, \dots, x_n)$ is any fact of the above defined structure, (thus, $1 \leq n \leq 2$), and $(x_{i,1}, x_{i,2}, \dots, x_{i,k})$ is a k-tuple of exactly k different elements, each of them occurring in n-tuple $(x_1, x_2, \dots, x_n)$, then:
if $k = 1$, $[\boldsymbol{\lambda}(x_{i,1}, x_{i,2}, \dots, x_{i,k})P^n(x_1, x_2, \dots, x_n)] = P^1$;
if $k = 2$, $[\boldsymbol{\lambda}(x_{i,1}, x_{i,2}, \dots, x_{i,k})P^n(x_1, x_2, \dots, x_n)] = C^2$.

By the definition of **λ**, it obviously satisfies axiom **6**. Furthermore, since only one possible world is in the structure we are considering (that is, **w***) and, by the definition of **w***, no fact holds in **w***, axiom **7** is trivially satisfied.

As regards the binary property **C**, let us set $\mathbf{C} = C^2$. By this definition, all clauses of axiom **8** trivially hold, except the first one. Let us then prove that:

For any fact f, $f \in \mathbf{D}$.

Proof
Let f be an arbitrary fact. Then, f is an ordered pair of the kind $P^m(x_1, x_2, \dots, x_m)$, where $P^m \in \mathbf{P}$ and $(x_1, x_2, \dots, x_m)$ is an m-tuple ($m \geq 1$) of elements of **D**. Since **D** is the union of all n-th domains Δ_n, for any element x_i in $(x_1, x_2, \dots, x_m)$, there is $n(i)$ such that $x_i \in \Delta_{n(i)}$. Let MAX = max$[n(1), n(2), \dots, n(m)]$. Then, by the definition of n-th domain, for any i, $\Delta_{n(i)}$ is included in Δ_{MAX}. Hence, $f \in \Phi_{MAX}$. But, by the definition of **D**, for any n, Φ_n is included in **D**. It follows that $f \in \mathbf{D}$.
Q.E.D.

Finally, axiom **9.3** is satisfied too. Since $\mathbf{Q} = \{C^2\}$, we only need to show that P^1 is obtainable from C^2 by means of a finite number of successive applications of the two operations of instantiation and **λ**-abstraction. But this is obvious: let us first instantiate C^2 with $(\varnothing, \varnothing)$, and thus obtain the fact $C^2(\varnothing, \varnothing)$. Let us then apply **λ**-abstraction to $\varnothing$ and $C^2(\varnothing, \varnothing)$; by the definition of **λ**, we get $[\boldsymbol{\lambda}(\varnothing)C^2(\varnothing, \varnothing)] = P^1$.

REFERENCES

Armstrong, David M. 1968. *A materialist theory of mind*. London: Routledge and Kegan Paul.

—— 1981a. *The nature of mind and other essays*. Ithaca, NY: Cornell University Press.

—— 1981b. "The causal theory of the mind." In Armstrong (1981a). [Reprinted in Lycan, ed. (1990), pp. 37-47.]

Block, Ned. 1978. "Troubles with functionalism." In *Perception and cognition: Issues in the foundations of psychology*, ed. by C. W. Savage. Minneapolis: University of Minnesota Press, pp. 261-325. [Reprinted without the section on Lewis in Block, ed. (1980), vol. 1, pp. 268-305. The same version, with the title, "An excerpt from *Troubles with functionalism*" also in Lycan, ed. (1990), pp. 444-468.]

Block, Ned, ed. 1980. *Readings in the philosophy of psychology*. 2 voll. Cambridge, MA: Harvard University Press.

Block, Ned and Robert Stalnaker. 1999. Conceptual analysis, dualism and the explanatory gap. *The Philosophical Review* 108:1-46.

Carnap, Rudolf. 1950. *Logical foundations of probability*. Chicago: The University of Chicago Press.

Chalmers, David J. 1995. Facing up to the problem of consciousness. *Journal of Consciousness Studies* 2:200-220. [Reprinted in Hameroff, Kaszniak and Scott, eds. (1996), pp. 5-28.]

—— 1996. *The conscious mind: In search of a fundamental theory*. New York: Oxford University Press.

—— 1997. Forward on the problem of consciousness. *Journal of Consciousness Studies* 4:3-47.

—— 2002. "Does conceivability entail possibility?" In *Conceivability and possibility*, ed. by T. Gendler and J. Hawthorne. New York: Oxford University Press, pp. 145-200.

Chalmers, David J. and Frank Jackson. 2001. Conceptual analysis and reductive explanation. *Philosophical Review* 110:315-361.

Cottrell, Allin. 1999. Sniffing the camambert: On the conceivability of zombies. *Journal of Consciousness Studies* 6:4-12.

Hameroff, Stuart R., Alfred W. Kaszniak and Alwyn C. Scott, eds. 1996. *Toward a science of consciousness: The first Tucson discussions and debates*. Cambridge, MA: The MIT Press.

—— 1998. *Toward a science of consciousness II: The second Tucson discussions and debates*. Cambridge, MA: The MIT Press.

Hameroff, Stuart R., Alfred W. Kaszniak and David J. Chalmers, eds. 1999. *Toward a science of consciousness III: The third Tucson discussions and debates*. Cambridge, MA: The MIT Press.

Hempel, Carl G. 1952. *Fundamentals of concept formation in empirical science*. Chicago: The University of Chicago Press.

Dennett, Daniel. 1991. *Consciousness explained*. Boston: Little Brown.

Descartes, René. 1637. *Discours de la méthode*.

Jackson, Frank. 1982. Epiphenomenal qualia. *Philosophical Quarterly* 32:127-136. [Reprinted in Lycan, ed. (1990), pp. 469-477.]

Kirk, Robert. 1974. Zombies versus materialists. *Aristotelian Society* 48(suppl.):135-152.

Lewis, David K. 1966. An argument for the identity theory. *Journal of Philosophy* LXIII, 1:17-25.
—— 1972. Psychophysical and theoretical identifications. *Australasian Journal of Philosophy* 50:249-258. [Reprinted in Block, ed. (1980), vol. 1, pp. 232-233.]

Locke, John. 1690. *An essay concerning human understanding*.

Lycan, William. G., ed. 1990. *Mind and Cognition: A reader*. Oxford, UK: Blackwell.

Mathieson, Chris. 2000. Reining in Chalmers: On the logical possibility of zombies. *Canadian Undergraduate Journal of Philosophy* 1:10-18.

Moody, Todd C. 1994. Conversations with zombies. *Journal of Consciousness Studies* 1:196-200.

Nagel, Thomas. 1974. What is it like to be a bat? *Philosophical review* 4:435-450.

Place, Ullin T. 1956. Is consciousness a brain process? *British Journal of Psychology* 47:44-50.

Putnam, Hilary. 1960. "Minds and machines." In *Dimensions of mind*, ed. by Sidney Hook. New York: New York University Press, pp. 148-179.
—— 1964. Robots: machines or artificially created life? *Journal of Philosophy* LXI, 21:688-691.
—— 1965. "Brains and Behavior." In *Analytical philosophy, Second Series*, ed. by R. J. Butler. Oxford: Basil Blackwell, pp. 211-235.
—— 1967a. "The mental life of some machines." In *Intentionality, minds, and perception*, ed. by Hector-Neri Castañeda. Detroit: Wayne State University Press, pp. 177-200.
—— 1967b. "Psychological Predicates." In *Art, mind, and* religion, ed. by W. H. Capitain and D. D. Merrill. Pittsburgh: University of Pittsburgh Press, pp. 37-48. [Reprinted with the title "The nature of mental states", in Rosenthal, ed. (1987), pp. 150-161. With the same title

also in Lycan, ed. (1990), pp. 47-56].

Rogers, Robert. 1971. *Mathematical logic and formalized theories*. Amsterdam: North Holland Publishing Company.

Rosenthal, David M., ed. 1987. *Materialism and the mind-body problem*. Indianapolis: Hackett Publishing Company, Inc.

Ryle, Gilbert. 1949. *The Concept of mind*. London: Hutchinson.

Seager, William. 2002. "Are zombies logically possible?" Presented at *CPA Meeting*, Toronto, May 2002. http://www.scar.utoronto.ca/~seager/zombie.html

Smart, John Jamieson C. 1959. Sensations and brain processes. *Philosophical Review* 68:141-156. [Reprinted in Rosenthal, ed. (1987), pp. 53-66].

Suppes, Patrick. 1957. *Introduction to logic*. New York: D. Van Nostrand Company.

NOTES

[1] For functionalism, the internal character of mental states basically reduces to their status of theoretical entities, thus not immediately observable ones, which are implicitly defined by their role in a network of causal input/output relationships.

[2] They include: the *What is it like to be* argument (Nagel 1974), the *Absent qualia* argument (Block 1978), the *Inverted spectrum* argument (Block 1978; already in Locke 1690), the *Knowledge* or *Mary's* argument (Jackson 1982), and the *Phenomenal zombie* argument (Chalmers 1996; different forms of the zombie argument are in Kirk 1974 and Dennett 1991. Moody's 1994 was the target article for a discussion forum of thirteen contributions, *Conversations with Zombies*, in *Journal of Consciousness Studies*, vol. 2, n. 4, 1995).

[3] Between 1994 and 2002, five meetings took place in Tucson, named *Biennial Tucson Consciousness Conferences*. MIT Press published the proceedings of the first three conferences; the first two volumes were edited by Hameroff, Kaszniak e Scott (1996, 1998), while the third one was edited by Hameroff, Kaszniak e Chalmers (1999).

[4] Seven meetings, named *Annual ASSC Conferences*, took place between 1997 and 2003. The latest one was held in Memphis, from May 30 to June 2, 2003.

[5] Edited in the UK by Keith Sutherland, it is perhaps the most influential among the consciousness studies journals. It was founded in 1994, and it is actively involved in the promotion of the *Biennial Tucson Conferences*. Most of the original discussion of the hard problem is contained in five forums, which were open and closed by two articles by Chalmers (1995, 1997). The first article had been previously presented at the *First Tucson Conference* in 1994.

[6] It was founded in 1995, and it is one of the two official ASSC journals. *Psyche* is an electronic free journal which can be found at http://psyche.cs.monash.edu.au .

[7] It is the second official *ASSC* journal, founded in 1992.

[8] It was founded in 2000, and it is mostly focused on the study of emotions.

[9] The term *corresponds* is purposely ambiguous, for the precise nature of such correspondence cannot be presupposed at this initial stage of the analysis. Thus, the possibility that internal and external facts of consciousness turn out to be identical is by no means excluded. If this were the case, the distinction between these two types of facts would be just *conceptual*, that is, relative to the way of describing, apprehending or knowing them. Two more possibilities to be considered are a relationship of logical consequence between the two types of facts (or logical supervenience between the corresponding properties), or a weaker relation such as some form of correlation or constant conjunction. In either of the two latter cases the distinction between internal and external facts of consciousness would be real.

[10] The issue of the correspondence between internal and external facts of consciousness should not be confused with the question concerning their intentional character. Among the internal facts of consciousness, let us consider *Mary's conscious experience, who is caught by anxiety*. In this case, it is not clear whether an external fact exists, which has approximately the same role as the hammer blow in *Mary's conscious experience, who is feeling the pain of such a blow in her toe*. The asymmetry between these two types of internal facts of consciousness is often described as the experience of anxiety's lacking a determinate content, and this may be taken as a good ground for maintaining the non-intentional character of such experiences. Whichever solution we choose for this problem, it is nevertheless a fact that *Mary's conscious experience, who is caught by anxiety* corresponds to a specific activation state of her central and/or peripheral nervous system. And it is exactly this externally ascertainable activation state that is the external fact, or the conscious function, which corresponds to Mary's internally lived anxiety.

[11] This claim of a one-to-one correspondence between internal and external facts of consciousness is not to be taken in a categorical sense, for, since it has been obtained by generalizing a number of non-systematic observations, its truth is at most limited to our world. In addition, in the light of deeper theoretical or empirical considerations, it could eventually be recognized false even in this world. Note then that accepting this claim does not prejudge in any way the question of the logical possibility of a zombie world (see sec. 4).

[12] For *strong antireductionism* I mean Chalmers' position, according to which the internal facts of consciousness, or the conscious experiences, are not reductively explainable in terms of physical facts. Two equivalent formulations of this position are: phenomenal properties are not logically supervenient on physical ones; the internal facts of consciousness (or the conscious experiences) are not logically entailed by the totality of the physical facts.

[13] I here refer to that I call the *super-hard problem of consciousness*, namely, the problem of constructing a theory of phenomenal consciousness that satisfy conditions 1-5 in the text. Chalmers (1996) attempts to solve this kind of problem. The super-hard problem is a special kind of *hard problem*, for this more general problem can be defined as the search for an adequate theory of phenomenal consciousness (where the adequacy conditions not necessarily are the ones specified by 1-5).

[14] For the moment, let us take the set of the internal facts of consciousness **CF** and the set of the external facts of consciousness **CP** as two additional primitive entities of the theory, with the further axiom that all elements of these two sets be facts. On this basis, let us then define *CF** (or *CP**) as the set of the internal (external) facts of consciousness that hold in **w***. It will be clear later on, however, that both **CF** and **CP** turn out to be definable within an appropriate extension of the axiomatic theory (def. [20] and [21], sec. 9.1).

[15] This, according to Chalmers, is equivalent to the claim that the facts of psychological consciousness are logical consequences of the physical facts or, equivalently, that the properties of psychological consciousness are logically supervenient on the physical properties.

[16] If *CF** is not empty, there is a possible world, namely the zombie world, where all physical facts of our world hold, and where no fact member of *CF** holds. Therefore, by def. [4], $\neg(F^* \models CF^*)$. Conversely, if *CF** is empty, then the consequent of def. [4] is vacuously true, and thus $F^* \models CF^*$.

[17] Doubts on this concern have been put forth, among others, by Cottrell (1999). The issue of conceivability of zombie worlds is linked to the problem of the relationships between conceivability, consistency and logical possibility. On this point, see Chalmers (2002) and Mathieson (2000). Also Block and Stalnaker (1999) and Chalmers and Jackson (2001) deal with this matter in the light of a more general question, concerning the conditions for a reductive explanation of phenomenal consciousness.

[18] The problem of the relationships between conceivability, consistency and logical possibility is quite controversial, but it gets more intricate because of a series of misunderstandings, or even plain errors of the logical kind. It is then convenient, in the first place, to get rid of such errors. Let me also make clear that, in this context, we are talking about *concepts*, not about *statements* or assertions. It is often taken for granted that the conceivability of a concept entails its consistency, and that, in turn, the latter entails the logical possibility of an object which satisfies the concept. The first entailment definitely has firm logical grounds. If a concept *C* is definable in some consistent theory *T*, then undoubtedly *C* is conceivable. Besides, because of the principle of non-creativity of definitions, the theory *T**, consisting of the conjunction of *T* with the definition of *C*, is consistent too; in such case we say that *C* is *relatively consistent* (namely, given the consistency of *T*). We can thus assert that, if a concept *C* is conceivable, in the sense that it is definable within a consistent theory *T*, then *C* is relatively consistent with respect to *T*. Still, does this entail that an object that satisfies *C* is logically possible? To answer this question we first need to make clear what we mean by logical possibility of an object which satisfies *C*. Luckily enough, this issue has a standard solution, based on the logical-mathematical usage of the notion of logical possibility. According to this usage, the logical possibility of an object that satisfies *C* consists in the relative consistency of the *statement* that asserts the existence of such object. That is to say, an object of type *C* is logically possible if, and only if, the theory $T^{*\#}$, consisting of the conjunction of *T** with the statement E_C = "there is x such that x is *C*" is consistent. But it is then evident that the relative consistency of the concept *C* in no way can entail the logical possibility of an object of type *C*. For the assertion that *C* is relatively consistent with respect to *T* just means that *T** is consistent, and this does not entail the consistency of $T^{*\#}$, which is an extension of *T**. In fact, *T** could very well entail the negation of E_C; in this case, $T^{*\#}$ would obviously be inconsistent, and thus an object of type *C* would turn out to be logically impossible.

[19] If the statement of the existence of the zombie world is not consistent with the assumed principles (*i.e.*, by the definition of note 18, if the zombie world is logically impossible), then they exclude the existence of the zombie world. If, conversely, the assumed principles are consistent with the existence of the zombie world (*i.e.*, by the definition of note 18, if the zombie world is logically possible), then such statement can be added to the principles, and the existence, in the logical-mathematical sense, of the zombie world is thus granted.

[20] Seager (2002) has pointed out that, unless one begs the question, a strong physicalist hypothesis cannot definitely be included among such auxiliary assumptions. A strong physicalist hypothesis is any hypothesis that entails that *any* fact (of our world) is reductively explainable by means of the physical facts (of our world). Such hypothesis has as an obvious consequence the physical reductive explainability of phenomenal consciousness, and it is thus inconsistent with the statement of the existence of a zombie copy of our world. However, a "proof" of this kind could not be accepted by Chalmers, for, from his standpoint, it would rather be a *reductio* of the strong physicalist hypothesis. The only type of physicalist hypothesis that can be included among the auxiliary assumptions is one that Chalmers himself would undoubtedly accept, namely, the minimal physicalist hypothesis [*R*], which *just* affirms the physical reductive explainability of psychological consciousness.

[21] It is interesting to notice that the problem we are dealing with is a typical problem of conceptual explication, in which a definition of a concept by genus (in this case: the experiences for an arbitrary subject *x*) and specific difference (in this case: whatever makes such experiences conscious) is required. Traditionally, such definitions are called *real definitions*. For the general theory of definition see Suppes (1957, ch. 8), Rogers (1971, sec. 4.4) and Hempel (1952, ch. 1); for real definitions and their relationship to explication see Hempel (1952, ch. 1, sec. 3); for the theory of explication, see Carnap (1950, ch. 1). Whenever an explication problem is in the form of the request of a real definition, it always reduces to two simpler sub-problems: (1) define the genus; (2) define the specific difference. This first moment of the solution of a problem of real definition corresponds to the second rule (rule of division) of Descartes' (1637) method, according to which to solve a difficulty we should divide it in as many simpler parts as needed. Once we have solved sub-problems (1) and (2), the original problem is solved too, as the required definition is just the conjunction of the genus definition with the one of the specific difference. Notice as well that this second moment of the solution of a problem of real definition corresponds to the third rule (rule of order) of Descartes' method, according to which, once a difficulty is divided in simpler parts, we should get back from them to the solution of the more complex original problem. In this connection, also see note 25.

[22] The last clause of the definition secures that a fact can be external only with respect to entities (the *n*-tuples of elements of the domain) that can be subject of other facts. We avoid in this way both irrelevant cases and counter-intuitive ones.

[23] The claim that *every thing has experiences* does not have anything to do with panpsychism. My position is that having an experience of a fact is identical with a particular *ontological* relationship that subsists exclusively between the fact that holds, *f*, and a system, *s*, in a privileged metaphysical position relative to *f*. Since the fact *f* consists of a property *P* and a *unique* subject to which the property *P* inheres, the aforementioned relationship is just being the subject of fact *f*. Whether or not this relation between *f* and *s* subsists is sufficient to explain the duality *internal - external*, *first person - third person*, *subjective - objective*; but it is by no means sufficient to ascribe any form of psychic life to the system *s*. The distinctive character of psychic life, in effect, is not subjectivity but, rather, whether or not *consciousness* is present, *i.e.*, whether or not a

particular type of facts, which we call *facts of consciousness*, hold. Consequently, a system *s* for which no fact of consciousness of which *s* is subject holds only has non-conscious (or, if you like, *unconscious*) experiences, and thus it is not a psychological subject (*s* is a subject exclusively in the metaphysical or ontological sense). For *s* to be a psychological subject, *s* ought to be the subject of facts of an appropriate type, namely, facts of consciousness, and at least some of these facts *should hold*.

[24] This condition brings in some self-reference within the axiomatic theory, since facts, which are members of the domain, turn out to be elements of the subject of other facts, which are members of the domain as well. We could then suspect that this situation allows the construction of an antinomy of the kind of Russell's. But this is not the case, for we can prove that the theory consisting of axioms **1** - **9.3** has a set theoretical model, and it is thus consistent with set theory. See the appendix for more details.

[25] Note that def. [14], of the relation *f* is a conscious experience for *x*, is a definition by genus and specific difference, where the genus is given by def. [6] and the specific difference by def. [11]. Def. [14] solves the problem of explicating the concept of conscious experience for an arbitrary subject *x*. On this issue, also see note 21.

[26] For example, if Mary were anesthetized, her being in neuro-physiological state *d* would not cause her being conscious of being in such state (that is to say, no consciousness of pain would take place).

[27] According to my proposal, thus, the determinate description D = "*Mary's conscious experience, who is feeling the pain of a hammer blow in her tow*" denotes exactly fact *f*. This identification is justified by the given analysis, by the intuitive meaning of description D, and by def. [16], according to which

> [D*] *f* is a self-conscious experience for Mary, whose content is her being in state *d*, which is caused by the hammer blow (*i.e.*, the content of Mary's conscious experience *f* is her experience f_d , which is caused by the hammer blow);

if we now compare the intuitive meaning of D with the characteristics of *f* asserted by [D*], it is in effect quite natural to identify the object described by D with *f*.

[28] An angelic copy of a world *w* can be thought as the dual concept of a zombie copy. For an angelic copy w^a of a possible world *w* is a world where all the internal facts of consciousness that hold in *w* hold, but where no physical fact holds.

[29] The condition that there is *f* such that $f \in CF_w$ and *f* does not hold in all possible worlds is not very restrictive, for the only angelic copies that theorem 6.2 does not exclude either are those, trivial, of a world where no internal fact of consciousness holds, or those of worlds where no internal fact of consciousness is contingent. In particular, an angelic copy of our world **w*** is ruled out by theorem 6.2 because, presumably, at least some conscious experiences of **w*** are contingent.

[30] A strong physicalist hypothesis is a principle that entails that *any* fact is reductively explainable by the set of all physical facts.

[31] The condition that there is *f* such that $f \in CF_w$ is not really restrictive, for the only angelic copies that corollary 2.7.3 does not exclude are those, trivial, of a world where no internal fact of consciousness holds. In effect, we could even add such a condition to the definition of angelic copy, and thus eliminate it from corollary 2.7.3. However, this does not seem to be convenient, because the perfect symmetry between def. [22] of angelic

copy and def. [5] of zombie copy would thus be lost. We could restore such symmetry by adding the condition that F_w be not empty to def. [5]. But then, in theorem 5.1, the biconditional would no longer hold; rather, only the implication from right to left would be granted.

[32] I thank Francesco Paoli for prompting me to consider objections of this kind.

[33] More precisely, axioms **1** - **9.3** define a type of mathematical structure or, according to Suppes' terminology (1957), *a set theoretical predicate*. Let us call such predicate C = *being a consciousness ontology*, and let us then define: a mathematical structure S *is a consciousness ontology* iff there are **D**, **P**, **Q**, **W**, **w***, **λ**, **C**, such that S = (**D**, **P**, **Q**, **W**, **w***, **λ**, **C**) and axioms **1** - **9.3** are satisfied. In general we say that an axiomatized theory T *is consistent* iff there is a set theoretical model of the theory, that is to say, iff there is a mathematical structure that satisfies the set theoretical predicate C_T defined by the axioms of T. It is interesting to notice that this definition of consistency for axiomatized theories is equivalent to the one of consistency relative to set theory of the statement E_{C_T} = "there is S such that S is C_T" (see note 18). For, by such definition, E_{C_T} *is consistent relative to set theory* iff the conjunction of set theory with E_{C_T} is consistent. Furthermore, if E_{C_T} is consistent relative to set theory, we can add E_{C_T} to such theory and thus make it one of its theorems. On the other hand, by the foregoing definition of consistency for an axiomatized theory, if E_{C_T} is provable within set theory, then T is consistent. Consequently, if E_{C_T} is consistent relative to set theory, then T is consistent. The converse holds too. Let us assume T to be consistent. Then, by the foregoing definition of consistency for an axiomatized theory, there is S such that S is C_T. Consequently, E_{C_T} is true. But, since all theorems of set theory are true, the conjunction of such theory with E_{C_T} (which is true as well) is consistent. It then follows that, if T is consistent, then E_{C_T} is consistent relative to set theory.