

# PHYSICALISM, *QUALIA* AND MENTAL CONCEPTS†

Diana I. PEREZ\*

Manuscript received: 2001.7.20.

Final version: 2002.2.3.

\* Universidad de Buenos Aires, CONICET, Bulnes 642, 1424 Buenos Aires, Argentina.

E-mail: dperez@filo.uba.ar

BIBLID [0495-4548 (2002) 17: 44; p. 359-379]

**ABSTRACT:** In this paper I shall carefully examine some recent arguments for dualism. These arguments presuppose a strong version of physicalism that I consider inappropriate. I shall try to show that, if we reformulate the thesis of physicalism according to Kim's view of physicalism (in terms of the supervenience relation), there is a third option, a version of type physicalism, where physicalism and *qualia* could be conciliated. In order to sketch this option, I shall consider the main argument against type physicalism: the explanatory gap argument, and two arguments that Kim mentions against physicalism: the inverted spectrum/zombies argument and the intrinsicality argument. I shall try to show that these three arguments depend upon a misconception of the nature of our ordinary mental concepts.

**Keywords:** physicalism, qualia, mental concepts, reductive explanation, identity theory.

## CONTENTS

Kim's road to a desperate dilemma

The way out

The Explanatory Gap argument

Two Thought experiments and the "grammar" of ordinary mental concepts

The intrinsicality argument

Conclusion

Bibliography

In his recent book *Mind in a Physical World*, Jaegwon Kim develops and defends the view that a robust physicalism must commit itself to a reductive account of mental phenomena. The kind of reduction that Kim favors in his book is what he calls "functional reduction." Unlike Ernest Nagel's theoretical reduction, functional reduction does not presuppose the existence of brute intertheoretical laws; hence, it gives a simpler format to the physicalist framework, with a few basic natural laws. To achieve a functional reduction of a given phenomenon, Kim's recipe is that "we must stop thinking of it as an intrinsic property but construe it as an extrinsic property characterized relationally, in terms of causal/nomic relations." (Kim

*THEORIA - Segunda Época*

Vol. 17/2, 2002, 359-379

1998, p. 25). Propositional attitudes have been generally thought of as extrinsic properties throughout the last three decades. But there is a rock in the road of this physicalist project: the so-called "*qualia*," the qualitative, subjective, conscious features of our mental life that seem to resist functionalization and, for that reason, reduction. As a consequence, the most obvious, directly known features of the world, our own sensations, ruin the entire enterprise of physicalism. If this line of argument is sound, then the remaining options in the mind-body problem are the following two: (a) to adopt physicalism, i.e. mind-body supervenience plus some other "deep" relation grounding it, in which case the cost is the unreality of some mental properties<sup>1</sup> or (b) to adopt dualism and reject mind-body supervenience (Kim 1998, pp. 118-20). Neither of these two options seems attractive to me, so I shall carefully examine the arguments that drove Kim to these desperate options. His arguments, in my view, presuppose a strong version of physicalism, that I consider inappropriate. In this paper I shall try to show that, if we reformulate the thesis of physicalism according to Kim's own previous view about physicalism (in terms of the supervenience relation with a nomological necessity involved), there are more interesting additional options to consider. I shall try to sketch a third option -a kind of type physicalism- where physicalism and *qualia* could be conciliated. In order to do that, I shall examine three arguments: the explanatory gap argument, the inverted spectrum argument, and the intrinsicality argument. My aim is to show that these arguments involve a notion of "phenomenal concept" that does not correspond to any ordinary mental concept we possess. If we build a theory of mind taking as a starting point our ordinary mental concepts, type physicalism is still a viable option.

### *Kim's road to a desperate dilemma*

In *Mind in a Physical World*, Kim deals with the three main problems in the philosophy of mind since Descartes "clearly and distinctly" distinguished between the mental and the physical. Although nowadays nobody accepts Descartes' substantial dualism, the biggest problems that he faced in his philosophy of mind could be reformulated in terms of properties.

1. *The Problem of the "Mark of the Mental"*: Is there any metaphysical distinction worth making between mental properties and physical properties?

2. *The Mind-Body Problem*: If the answer to the first question is affirmative, what is the metaphysical relationship between mental properties and physical properties?

3. *The Problem of Mental Causation*: Is there any causal connection between mental properties and physical properties?

The last of these three questions is central to Kim's view. He thinks that to be real is to have genuine causal work to do. So if the account that one adopts for the relation between mental and physical properties leads us to the causal impotence of mental properties, then the mental becomes unreal, and the first question cannot be formulated anymore. I shall not deal with the problem of mental causation in this paper; I shall concentrate on Kim's answer to the first two questions. I do not accept the reality criterion for properties that Kim proposes; in fact, I believe that a property's being real is something that depends on its belonging to the laws of nature, and not on its causal powers. So I propose divorcing the third problem from the first two. I shall show that the characterization of physicalism that I will adopt accepts mental properties as real, according to the definition I have just proposed.<sup>2</sup>

The answer Kim proposes to the first question is affirmative, as almost everyone would accept nowadays. While discussing whether there is a "mark of the mental," Kim argues that our conception of the mental lacks unity, and he concludes: "there are two broad basic categories of mental phenomena: sensory or qualitative states (*qualia*) like pains and sensings of colors and textures, and intentional states, like beliefs, desires, and intentions." (Kim 1996, p. 23)<sup>3</sup>. Conceding the fact that there are two different features that distinguish the mental from the physical we can ask whether these two features are present in every mental state we possess or if these two features could stand alone in a given mental state. In other terms, some of our ordinary mental concepts seem to refer to an intentional state (for example "believing that snow is white") whereas other ordinary terms seem to refer to a qualitative state (for example "pain"). But it is not obvious that a belief cannot have a phenomenal character or that pain has no representational features at all. In some recent works on *qualia*, this second claim is questioned. According to the representational theory of *qualia* our qualitative mental states like pain have also a representational character and, in some cases, it is claimed that this representational character exhausts these mental states (Dretske 1997, Tye 1995). As we will see, the problem for physicalism starts, in Kim's opinion, when we recognize that the representational aspect of sensations does not exhaust the nature of our sensations, claiming that there is something more than this: the phenomenal/ subjective aspect of our sensations. If all our qualitative states could be reduced to representational states, then there would be no problem for

physicalism. I prefer to be neutral on this point: in my opinion further empirical findings will tell us if the representational theory of *qualia* is correct. In what follows I shall accept, for the sake of the argument, that there is something more than the representational character of *qualia*, otherwise Kim's point cannot be formulated.

Thus, once having accepted the idea that there are two metaphysical features which distinguish the mental and the physical, i. e. that there are two different kinds of mental features, the second and third problems posed above have to be divided into two parts. In fact, it is plausible to maintain that the relation between intentional and physical properties is not the same as the relation between *qualia* and physical properties; and it is also plausible to affirm that intentional properties are causally efficacious while *qualia* are epiphenomenal, or vice versa.

So, the second question is whether there is a metaphysical relation between mental properties and physical properties. And given the answer to the first question, this problem has to be divided into two parts as follows: 2.1. What is the metaphysical relation between intentional properties and physical properties? 2.2. What is the metaphysical relation between qualitative properties and physical properties?

Kim refrains from answering these questions by appealing to the notion of supervenience (Kim 1998, pp. 9-15). He thinks that mind-body supervenience does not give the answer to the mind-body problem because it is compatible with a host of different answers to this problem: emergentism, type physicalism, epiphenomenalism, physical realizationism, etc. The thesis of mind-body supervenience is the weakest thesis that any physicalist is committed to; it defines "minimal physicalism."<sup>4</sup> But mind-body supervenience is a "phenomenal" relation and it has to be grounded on, and explained by, some deeper dependency relation, otherwise there would be a host of unexplained, brute laws connecting the mental with the physical. Different mind-body theories propose different explanations for the supervenient connections.

In the case of intentional properties, Kim favors what he calls "physical realizationism." It states (i) that mental properties are functional properties, and (ii) that mental properties, if they are realized, must be physically realized. According to (i), intentional properties are thought of as extrinsic, relational, functional properties. For Kim, as for D. Lewis, (1966, 1970, 1972) and N. Block (1980; 1990), functional properties are second order properties. Functional properties are the existential quantification over a set of base properties (in this case, physical properties).

Physical realizationism entails nomological supervenience, hence minimal physicalism is safe. And physical realizationism gives us an explanation for mind-body supervenience. In Kim's words:

(...) by definition, having M is having a property with causal specification D, and in systems like s, P is the property (or one of the properties) meeting specification D. (...) having M for these systems simply *is* having P (Kim 1998, p. 24).

According to Kim, the strategy sketched above consisting of "functionalizing" properties accords with the way in which reductions are achieved in other sciences. Kim gives some examples to support his claim: "temperature," "gene," and "transparent," are properties that were reduced when scientists stopped thinking of them as intrinsic properties and adopted an extrinsic, relational characterization of them in terms of causal/nomic relations (Kim 1998, p. 25). In fact, according to Kim, what scientists are looking for are not extrinsic properties but extrinsic concepts: scientific progress depends upon the possibility of giving new descriptions in terms of causal /nomic relations of the same phenomena we knew before. This is the way the story goes: once upon a time we thought about heat as an intrinsic feature of some physical objects, but with scientific progress we obtained a new set of concepts, belonging to a new scientific theory, which allow us to identify heat with the kinetic energy of particles. The fact that heat could be described extrinsically as "what is caused in normal human beings under such and such circumstances" was crucial to achieve this theoretical identification. Intentional properties are nowadays considered as relational properties too, they have been functionalized. Kim's only argument in defense of this claim is the following: "it seems to me inconceivable that a possible world exists that is an exact physical duplicate of this world but lacking wholly in intentionality." (Kim 1998, p. 101). Therefore, intentional properties could be reduced to physical properties (via functionalization), and hence a robust physicalism could be established.<sup>5</sup>

But once we have accepted that *the only way*<sup>6</sup> to construct a robust physicalism is to functionalize all non-physical concepts, the physicalist gets into trouble, because *qualia* seem to resist functionalization. Kim gives us two different arguments to support this claim. On the one hand he accepts that the conceptual/ logical possibility of the inverted spectrum entail the non-functionalizability of *qualia*.<sup>7</sup> On the other hand, *qualia* seem not to be extrinsic properties, Kim says: "*qualia* are intrinsic properties, if anything is." (Kim 1996, p. 176; Kim 1998, p. 102). But the first step to functionalize a property was to stop thinking about it as an intrinsic property and

start thinking about it as an extrinsic one. If it is not possible, that property cannot be functionalized, and hence, it can not be included within a physicalist framework.<sup>8</sup> Following this line of argumentation, we have reached the dilemma we did not want. Either (a) we conserve physicalism, but we have to reject our color sensations, our pains, and the like (that is to say, we have to accept a kind of eliminativist thesis, where the unreality of some mental properties is to be accepted); or, otherwise, (b) we are committed to dualism. Both options seem unacceptable to me.

### *The way out*

In what follows I shall point out the steps of this line of argumentation that are dubious. In the first place, I think that the best way to construct a robust physicalism is to realize that the answer to the mind body problem could be different in the case of intentional properties from in the case of qualitative properties. I shall not discuss the answer that Kim gives for the case of the intentional; I shall accept, for the sake of the argument, that physicalist realizationism as an answer to those kinds of properties. But I do not think that this answer has to be extended to the case of qualitative properties. Instead, I propose looking for another physicalist answer for this case.

In order to provide this answer, it is worth mentioning that there is a problem with the notion of "physicalism" we are considering in these arguments, because there are two different conceptions of physicalism -a stronger one and a weaker one-, and that Kim is ambiguous about which of the two notions he prefers. I shall show that the dilemma posed above could be overcome by adopting the weaker characterization (and not the stronger one).

As I said above, physicalism could be formulated in terms of supervenience, and that is Kim's favorite way: to hold a physicalist thesis about the mental is to accept mind-body supervenience<sup>9</sup>:

Mental properties supervene on physical properties in the sense that if something instantiates any mental property *M* at *t*, there is a physical base property *P* such that the thing has *P* at *t*, and necessarily anything with *P* at a time has *M* at that time.<sup>10</sup>

This definition has two different interpretations according to the way in which we interpret the modal operator involved. It could be understood as expressing a conceptual necessity<sup>11</sup> or a physical or nomological necessity.<sup>12</sup> Kim calls the last one "weak physicalism" (Kim 1996, p. 173). I propose calling the first one "strong physicalism." In my view, Kim is am-

biguous about which of these two definitions is correct. At times, he accepts that not every mental property could be related to its base property with the same modal force (Kim 1998, p. 10). At other times he criticizes weak physicalism (Kim 1996, pp. 172-77),<sup>13</sup> And, at still other times, for example when he gives the definition cited above, he affirms that the modal operator involved in the definition is a nomological necessity, i.e. that "it holds in all worlds that share with our world the same fundamental laws of nature" (Kim 1998, p. 39).

D. Chalmers, following F. Jackson, makes a similar distinction. For him, "[physicalism] is true if all the positive facts about the world are entailed by the physical facts" (Chalmers 1996, p. 42. This is also the characterization of physicalism given in Jackson 1994). "Entailment" is the relation between the mental and the physical, i.e. for physicalism to be true, the mental and the physical are to be metaphysically or conceptually connected. To the contrary, if we maintain that there could only be a lawfully (nomological or natural) connection between the mental and the physical, we are holding weak physicalism.<sup>14</sup>

Unlike Chalmers and Jackson, Kim is not committed to the view that the only way to be a robust physicalist is to accept a conceptual relation between the mental and the physical. In fact, he holds that in order to be a physicalist we have to accept two constraints: first of all, supervenience (logical *or* nomological) must be accepted; and, in the second place, the very supervenience relation has to be explained. It is clear that if we accept the model of functional reduction both constraints are met, but it is not clear to me that this is the only possible way to meet them, and I am not sure that Kim feels strongly committed to that view. In my opinion, there are still two other options; hence, Kim's desperate dilemma (which depended upon the impossibility of functionalizing *qualia*, and also on the indispensability of functionalizing everything to explain supervenience) could be overcome by adopting one of these options.

Kim claims, as I said above, that supervenience alone is not an answer to the mind-body problem because it is compatible with a great variety of answers to that problem. So, the supervenience relation itself has to be explained. The supervenience relation between the intentional and the physical was already explained by functionalization and reduction. And if we ask for an explanation to ground the qualitative-physical supervenience there are, in my view, two alternative options. The first option is a *kind of* emergentist position. Nomological supervenience warrants that physical and qualitative properties are connected by natural laws (interlevel laws).

But, it could be held that not all the supervenience (nomological) relations between the qualitative and the physical are brute facts as emergent laws were thought to be.<sup>15</sup> According to emergentism all the interlevel nomic relations are brute facts about the world. In contrast, the supervenience thesis allows that many of these interlevel connections could be explained by (i.e. deductively inferred from) some few basic interlevel laws. Within this framework some (maybe few) interlevel laws will be considered as brute laws of nature. It seems clear to me that there will always be some unexplained brute facts, if we accept the layered view of the world, and the reality of the mental (in this case, qualitative) phenomena, because it is likely that there will be some sort of interlevel nomic relations that are to be basic.<sup>16</sup> This position is a *kind of* emergentism in the sense that it accepts that some interlevel laws will remain unexplained, but it does not forbid explaining many of them. If we accept calling this position a "physicalist" one (or not) is only a question of terms.<sup>17</sup>

The second alternative -the one I favor- is to accept a type-type identity theory, in the spirit of Smart.<sup>18</sup> According to this theory, there are no mere nomic correlations between qualitative phenomena of our mental life and physiological phenomena (or some other more basic phenomena); they are simply identical. Mental terms, like "pain," and physical terms, like "C-fiber's stimulation," express different concepts without any conceptual/a priori connection between them but, following Frege's distinction, these two concepts share the same reference: they point to the same property in our world. In this view, supervenience (transmuted to ascriptive supervenience, i.e. to a relation between judgments, instead of properties<sup>19</sup>) is based on a deeper metaphysical relation: Identity. And for that very reason, identity *grounds* supervenience; hence, within this framework the two constraints that Kim settled for physicalism are met: (1) supervenience holds, and (2) it is explained by a deeper metaphysical fact.<sup>20</sup> In fact, Kim himself accepts that "type physicalism is a strong and robust materialistic doctrine" (Kim 1996, p. 62).

The trouble with this option is that it is not a widely accepted view nowadays. As Kim regrets in many works, "the identity theory was unexpectedly short lived" (Kim 1998, p. 2; Kim 1989, p. 266). Although the objections raised against the identity theory were not demolishing, most philosophers abandoned the identity theory convinced by two lines of arguments: Davidson's anomalous monism, and Putnam's multiple realizability argument. In my opinion, both arguments could be acceptable when applied to intentional properties but they do not apply to qualitative



properties. Davidson's anomalous monism is explicitly concerned only with intentional states, and so, it does not affect the possibility of identifying phenomenal properties with neurophysiological properties. And Putnam's multiple realizability argument highlights the identity of causal patterns concerning inputs and behavioral outputs among different species, but these patterns do not seem essential to the application of qualitative concepts (I shall return to this issue later).

In my view, one of the strongest reasons why most people do not accept the identity theory for qualitative features is because of the "explanatory gap" argument. In what follows I shall make some remarks about this argument, trying to show that a kind of type physicalism for qualitative properties is still an option. Finally, I shall examine the two main arguments in behalf of the non-functionalizability of *qualia* explicitly mentioned by Kim: the inverted spectrum argument and the intrinsicality argument. While examining these arguments I shall make some remarks on the nature of our mental concepts that will help us to find the way out of Kim's dilemma.

### *The Explanatory Gap argument*

The explanatory gap argument was first proposed by J. Levine 1983. It is not concerned with the problem of the identity thesis as an ontological thesis. Levine's point is an epistemological one. Briefly, in Levine's opinion, there will always exist a gap between mental (qualitative) concepts and physical concepts and, for that reason, we will never come to know if the identity thesis (*qua* ontological claim) is true or not. The gap between mental and physical concepts is derived from the fact that the qualitative character itself -the painful sensation of pain- cannot be explained within a physicalist framework independently of whether or not we accept psychophysical identity. Levine's argument on this point is based on a difference between the water/H<sub>2</sub>O case and the pain/C-fibers case. He claims that "there is an apparent *necessity* that flows from the reduction of water to H<sub>2</sub>O, a kind of necessity that is missing from the reduction of pain to the firing of C-fibers" (Levine 1993, p. 128). Levine's point is that although it is conceivable that pain can exist without C-fiber stimulation, and C-fiber stimulation can exist without pain, and that it is also conceivable that something with water's macroproperties can exist without being H<sub>2</sub>O, nevertheless it is not conceivable that H<sub>2</sub>O can fail to have the macroproperties we attribute to water. And this very asymmetry is due to the fact that

the chemical theory (appealing to the identity between  $H_2O$  and water) explains something that the physicalist theory of *qualia* left unexplained.

This asymmetry is very doubtful. I believe that the asymmetry only reveals the poverty of our neurophysiological knowledge *vis-à-vis* our chemical knowledge. In fact, " $H_2O$ " is an expression that belongs to our current chemical theory, and we understand that expression because we know (more or less) what an atom is, what a molecule is, what hydrogen is, and what oxygen is. We could also mention a lot of other elemental chemical substances. We also know that not every pair of atoms can form a molecule, that atoms have valences, and so on. And all this knowledge is what prevents us from conceiving a set of  $H_2O$  molecules that do not behave as water, i.e. that do not exhibit the macro-properties that we attribute to water. But, what about "C-fiber's stimulation"? This expression, to the contrary, does not belong to any scientific theory we learned in high school; in fact, it does not belong to any theory at all. It is not linked to other concepts we have; it means nothing to us. So the reason why we can conceive a C-fiber's stimulation without the sensation of pain, is that "C-fiber's stimulation" does not mean anything to us. My claim is that while we acquire more neurophysiological knowledge about the functioning of our brains, it will turn out to be increasingly difficult to conceive that very functioning without the associated macro-mental-properties.

I am not defending the idea that the gap will no longer exist, but that the same gap that we feel while talking about the mental and its physical base is the one that we feel when we try to relate an ordinary-commonsense- description of a given physical phenomenon (for example, "the water boils"), and its micro-physical scientific explanation. Let us take a look at the following explanation:

What is explained by the theory that water is  $H_2O$ ? Well, as an instance of something that's explained by the reduction of water to  $H_2O$ , let's take its boiling point at sea level. The story goes something like this. Molecules of  $H_2O$  move about at various speeds. Some fast-moving molecules that happen to be near the surface of the liquid have sufficient kinetic energy to escape the intermolecular attractive forces that keep the liquid intact. *These molecules enter the atmosphere. That's evaporation.* The precise value of the intermolecular attractive forces of  $H_2O$  molecules determines the vapour pressure of liquid masses of  $H_2O$ , the pressure exerted by molecules attempting to escape into saturated air. As the average kinetic energy of the molecules increases, so does the vapour pressure. *When the vapour pressure reaches the point where it is equal to atmospheric pressure, large bubbles form within the liquid and burst forth at the liquid's surface. The water boils* (Levine 1993, p. 129, my italics).

It seems to me that there is a gap between the last two sentences of this explanation (and also between the other two underlined sentences). It is the gap that we will always find between two different descriptions of a given phenomenon, one in scientific terms, the other in ordinary terms. And I do not believe that any conceptual analysis that we could give of our ordinary concepts could fill this gap.

Levine concludes his paper declaring: "what seems responsible for the explanatory gap, then, is the fact that our concepts of qualitative character do not represent (...) causal roles" (Levine 1983, p. 134). And so we come back to the same arguments that Kim mentions in order to reject the possibility of functionalizing *qualia*: in the first place, the thought experiment about the inverted spectrum, and, in the second place, the idea that *qualia* are intrinsic properties. Let us now examine these arguments.

*Two Thought experiments and the "grammar" of ordinary mental concepts*

There are two thought experiments, the inverted spectrum argument and the zombie argument, that are usually used in order to prove the falsity of physicalism. I shall concentrate on the former, but my point could be extended to the latter. The inverted spectrum argument could be stated as follows:<sup>21</sup>

(i) There are conscious (phenomenal) experiences in our world (in my words: there are true descriptions of our world that involve mental terms such as "sensation of pain," "seeming red," etc.).

(ii) We can conceive a scenario physically / functionally indistinguishable from our world where phenomenal concepts are truly applied in a different way from the way we hold them in our world. In the case of the inverted spectrum, we can conceive a situation where our phenomenal experiences are "inverted" from our actual experiences, in the sense that we could apply the same physical / functional concepts to a given situation, but changing the phenomenal concepts applied.<sup>22</sup>

(iii) If something is conceivable, then it is possible.

(iv) If the inverted spectrum (the zombie world) is possible, then, logical supervenience between the physical and the qualitative fails.

(v) Physicalism is true if and only if all the positive facts are entailed by the physical facts, i.e. if logical supervenience holds. In other words, someone who knows all the physical facts and possesses the higher-level concepts, also knows all of the higher-level facts.

So, (vi) physicalism is false.

I shall not discuss premises (i), (iii), (iv) and (v). I said above that I do not agree with this characterization of physicalism (premise (v)), and that the logical possibility of the inverted spectrum is not a counterexample of physicalism when we accept a weaker form of physicalism according to which the link between the mental and the physical is weaker than conceptual. In this paragraph, I want to discuss only premise (ii).

There are two possible formulations of (ii): one in terms of properties, the other in terms of concepts. The first is in terms of properties, i.e. in terms of the way properties are distributed in our world, and in other possible worlds (the inverted world, or the zombie world). But in my opinion, this formulation begs the question for dualism: it presupposes that there are two different kinds of properties, so qualitative properties are *supposed*; their existence is not *proved* by the argument. Probably this is the best way to interpret Chalmers' argument, because he claims in the *Introduction* to his book that he takes consciousness seriously (Chalmers 1996, p. xii), i.e. he accepts that there are conscious experiences in our world. In this case, the zombie and inverted spectrum arguments are given in order to prove that consciousness is not logically supervenient upon the physical and, from that fact, the falsity of physicalism.

The second formulation of premise (ii) is in terms of concepts or predicates.<sup>23</sup> This interpretation is more coherent with Kim's proposal: he accepts, as we saw above, that intentional *concepts* are those that could be functionalized and so he has to accept also that phenomenal *concepts* are those that cannot be functionalized. In fact, the very idea of functionalization applies to concepts, not to properties. Properties are relational/extrinsic or categorical/intrinsic independently of the way we refer to them. It makes no sense to functionalize something that has an intrinsic nature, i.e. an intrinsic property. On the contrary, we could give a functional description, i.e. a functional concept, to refer to a property no matter if it has an intrinsic or an extrinsic nature. I shall come back to this issue in the next paragraph.

In this section, I shall try to show that the very conceivability that premise (ii) proclaims rests on a misunderstanding about the nature of our ordinary mental concepts. My claim, roughly put, is that the notion of "phenomenal concept" involved in this premise does not correspond to any ordinary mental concept we possess.<sup>24</sup> As I said above, there are two features of our mental life that are generally accepted: the intentional and the qualitative. But they do not correspond one-to-one to our *ordinary* mental concepts. There are some mental concepts that seem to point exactly to the

first kind of features ("belief" is the paradigmatic case), and others that seem to point to the second kind (for example, "having a red sensation"). But most of our ordinary mental concepts conflate both kinds of features. Our concept of "pain," for example, refers to a certain state of an organism that fits a certain functional role, but also to a state that possess a certain phenomenal quality. A similar phenomenon occurs with our ordinary emotional concepts: "fear" points to a certain quality (the "sensation of fear") but also to a certain functional/ representational state, and even to a certain physical state, described in terms of levels of adrenaline, the increase of our heart rate, etc. As W. James wisely wrote:

What kind of an emotion would be left if the feeling neither quickened heart-beats nor of shallow breathing, neither of trembling lips nor of weakened limbs, neither of gooseflesh nor of visceral stirrings, were present, it is quite impossible for me to think. Can one fancy the state of rage and picture no ebullition in the chest, no flushing of the face, no dilatation of the nostrils, no clenching of the teeth, no impulse to vigorous action, but in their stead limp muscles, calm breathing, and a placid face? (James 1890, quoted from Damasio 1994).

In the third place, a similar thing occurs with our ordinary propositional attitude concepts, when they are used to refer to a conscious occurrent belief or desire: If I desire to eat an ice cream, certainly I have a disposition to eat an ice cream if I obtain one, but also I have a certain qualitative feeling, a what-it's-like-to-desire-an-ice-cream, and probably there are also certain physical modifications in me that I cannot describe right now, but probably could be described once. In my opinion, our prototypical mental concepts involve the two features of the mental, when we set apart one of them we make deviant uses of the ordinary terms.

My point here is that we cannot make sense of the idea of "phenomenal concept" without begging the question against physicalism. In other terms: our ordinary mental concepts are not phenomenal concepts in the sense that they are needed to make the world described in premise (ii) conceivable. The only way to make sense of the notion of "phenomenal concepts" is either by accepting that there are phenomenal properties (begging the question against materialism) and defining phenomenal concepts as the ones that refer to phenomenal properties, or by accepting that there exists a private language, a private concept (or concepts) that lacks external application criteria. However, since Wittgenstein developed his argument against private languages, it is hard to accept nowadays that there are concepts that lack external criteria. In any case, all our ordinary mental concepts include some relational/causal understanding of the concept (see also Hill 1991, ch. 7).

Even those who doubt the private language argument, for example D. Chalmers (Chalmers 1996, ch. 5, note 13, p. 381), at a crucial point in his argumentation, affirms

we might say that the notion of pain is ambiguous between the phenomenal and the psychological concept, or we might say that both of these are components of a single rich concept (Chalmers 1996, p. 17).

In my view, the first of these options drives us to the unacceptable thesis that there are two concepts behind our ordinary word "pain" a phenomenal concept of pain and a psychological concept of pain. But this claim seems to imply that when we are attributing pain to a third person and when we attribute pain to the first person we are using different concepts. In other words, what is meant by the word "pain" when I say "I am in pain" and when I say "He is in pain" are different things. I cannot accept that consequence. If a certain thesis drives us to the idea that I cannot attribute pain to other people meaning the same as I mean when I attribute pain to myself, we had better abandon the thesis.

So, the second option is the right one. Our ordinary mental concepts are complex, rich, they have different application criteria for the first and the third person. That is the "grammar" of our ordinary mental concepts. If we try to take apart the first and the third person concepts, we fragment our ordinary concepts, and this move has to be justified. I am not claiming that our ordinary concepts cannot change, but that any change we introduce has to be justified in one way or another. Otherwise, we become "primitive people" as Wittgenstein said:

When we do philosophy we are like savages, primitive people, who hear the expressions of civilized men, put a false interpretation on them and then draw the queerest conclusions from it (Wittgenstein 1953, §194).

### *The intrinsicity argument*

The last reason Kim mentions for his opposition to the functionalizability of *qualia* is the idea that "*qualia* are intrinsic properties if anything is." (Kim 1996, p. 176; Kim 1998, p. 102). Once again, I propose two separate interpretations of this claim: one of them in terms of concepts, the other one in terms of properties. First, let us consider the argument in terms of properties. I think that this is the idea of the majority of philosophers at this point: that qualitative *properties* are intrinsic. I would like to make some remarks on this claim. First of all, it is not clear, in my view, that

we have to accept that there are intrinsic properties at all. On the one hand, there is not unanimous agreement on which properties would count as intrinsic properties. The properties usually mentioned (Rodríguez Larreta draft, p. 3, Garrett 1995, p. 162) are: (1) secondary properties, like colors or tastes, that many people consider as relational<sup>25</sup>, (2) physical properties, although many people think that physical properties are relational, I shall return to this point later, and (3) qualitative properties whose existence we are trying to prove. On the other hand, there is no unanimous agreement about the right definition of intrinsicity, there are different characterizations of "intrinsic properties" and, in my view, none of them fits *qualia*.<sup>26</sup>

An interesting attempt to characterize intrinsic properties starts from the idea that all properties we could know are relational because all properties are known by the effects they produce on us (see Blackburn 1991). In this case, "intrinsic property" would mean an unknowable "thing-in-itself." But *qualia* are not unknowable, so this characterization is not the one we are looking for. Moreover, and this is my final remark on this point, following this characterization of intrinsicity, the problem with intrinsic properties is not a peculiar problem for *qualia*, in fact it is a problem that we can generalize as Jackson suggests when he considers the view that he calls "Kantian physicalism" (Jackson 1998, pp. 23-4). According to this view, all that a physicist could know about the physical nature of our world is what physical properties *do*. We only know, for example, that a given property produces the instantiation of another property in certain circumstances, or how a given property affects our perceptual systems and our measurement instruments in a certain way. But from the fact that our terms for the fundamental properties pick out properties via the causal relations the property enters into, it does not follow that they could not be referring to intrinsic properties. And if we could pick out properties exclusively via their relational mode of presentation, it remains impossible to know if we are picking out physical intrinsic properties, qualitative intrinsic properties,<sup>27</sup> or just relational properties "all the way down" (if we accept that there are no intrinsic properties at all). We need an additional argument to show that these relational physical (or psychological) concepts are referring to intrinsic properties, and I have not seen any of these.<sup>28</sup>

But maybe the right interpretation of Kim's argument is not in terms of properties, but in terms of concepts. From this point of view, the claim about the intrinsicity of *qualia* is a claim about the intrinsicity of qualitative concepts. As I said before, this interpretation is more coherent

with Kim's proposal because he claims that intentional *concepts* are what could be functionalized and so he has to also accept that phenomenal *concepts* are those that cannot be functionalized. In fact, the very idea of functionalization applies to concepts, not to properties: to functionalize something is to give a functional or relational description (or to construe a functional or relational concept) in order to refer to something that was referred to by an intrinsic concept before. The properties referred to by these new concepts will be relational/extrinsic or categorical/intrinsic independently of the way we refer to them.

However, it is hard to see what an intrinsic phenomenal concept could be. Probably the idea is that those concepts acquire their content from an internal ostensive definition; whether a concept applies to a sensation or not depends exclusively upon the nature of that sensation. With this account of our sensation concepts we come back to the inner ostension and the private language that Wittgenstein criticized. And, as I have tried to show above, our ordinary sensation concepts cannot be understood that way.

Another way to try to make clear what a phenomenal concept could be is following Chalmers' (forthcoming) suggestion, claiming that when we are having a red experience and we think that we are having a red experience, there are at least four phenomenal concepts involved in that thought: (1) the community relational concept ("the quality caused in normal observers in my community by red things"), (2) the individual relational concept ("the quality normally caused in me by red things"), (3) the indexical concept ("the quality I am experiencing now") and (4) the qualitative concept ([red splotch]). But, according to Chalmers, the property that makes a statement true involving each of these concepts is the same, that is to say that it is *a posteriori* necessary that (1) = (2) = (3) = (4). It is not clear why he affirms that these are four different concepts, and not just one complex concept (as I prefer to think). In my opinion our ordinary concept of "having a red sensation" involves all these four elements (or at least the first three). In fact, my additional worry about this proposal is that I cannot see why we have to add (4) to the first three relational phenomenal concepts mentioned.

What I am trying to say is that unless we accept the idea of a private language, we have to accept that all mental concepts are relational. But what is the ontological consequence we can draw from this fact about concepts? I think that we cannot throw a direct bridge between our concepts and the properties that are instantiated in our world. From the fact that all our concepts are relational, it doesn't follow that all properties are rela-



tional nor that they are intrinsic. As I said above, all the concepts (1) to (4) refer to the same property, hence, either there are relational concepts that refer to intrinsic properties or intrinsic concepts that refer to relational properties. In any case, the bridge cannot be established.

### *Conclusion*

In this paper I have tried to sketch a way out of the dilemma posed by Kim (1998), according to which the only available solutions to the mind-body problem are either (a) a physicalist eliminativist position, or (b) dualism. The third option that I was trying to sketch is based on the following thesis:

(i) The physicalist account of intentionality could be different from the physicalist account of qualia.

(ii) According to Kim's own view, to be a robust physicalist there are two constraints: (1) supervenience holds and (2) it is explained by a deeper fact: identity. So, a type physicalism for qualia is a genuine physicalist option.

(iii) The identification between the qualitative with one or another physical state depends upon the progress of scientific knowledge, it does not depend upon the possibility of giving a conceptual analysis of our ordinary mental concepts.

(iv) The explanatory gap we find between the qualitative and the physical is due to the poverty of our knowledge about the functioning of our brain: the gap will be dissolved once we reach a "complete neuroscience."

(v) The arguments that Kim (among others) mention against the functionalization of qualia (the inverted spectrum, (the zombie argument) and the intrinsicity argument) depend upon a notion of "phenomenal concept" which does not correspond to any ordinary mental concept at all, and which does not need to be accepted. If we take our ordinary mental concepts with their rich and complex "grammar" as a starting point to scientific inquiry, the best answer to the (qualitative) mind-body problem is still type physicalism.

### *Notes*

† I read previous versions of this paper at the X Seminario Interuniversitario de Filosofía y Ciencia Cognitiva: "Explaining the Mind: Causation, Reduction and Supervenience", San Sebastián, Spain, June 17<sup>th</sup> to 19<sup>th</sup>, 1999; at the Coloquio de Filosofía

Teórica de SADAF, Buenos Aires, november 19<sup>th</sup> and 20<sup>th</sup>, 1999; at Kansas State University, april, 2000, and at the Coloquio "Language, Mind and World", Tlaxcala, México, march, 14<sup>th</sup> to 17<sup>th</sup>, 2001. The paper was improved due to the helpful comments of the participants of these events, and specially because of the comments made by E. Rabossi, J. Rodríguez Larreta, M. Sabatés, L. Skidelsky, N. Stigol, J. Vergara, and I. Zuberbuller. I would like also to thank the suggestions from two anonymous referees of *Theoria*, which helped me to improved this final version.

- <sup>1</sup> This step from mind-body supervenience to mental irrealism is not obvious. It depends on: (1) the supposition that the only two physicalist options at this point are reduction or elimination, (2) the thesis that while adopting a reductive account of the mental, it has no *new* causal work to do; instead, the mental just inherits the causal power of its realizers, and (3) the idea that to be real (for a property) is to have a causal work. I will not discuss these assumptions on this occasion.
- <sup>2</sup> See Pérez (1999, ch. VI, #1), and also Sabatés (draft) where he proposes another account for the reality of properties, based on the idea of dependency.
- <sup>3</sup> This distinction between the intentional and the qualitative could be stated in terms of properties or in terms of concepts or predicates. Kim's account of functional reduction holds that what would be reduced are intentional *concepts*, hence, probably, the best formulation of this distinction is itself in terms of concepts. I shall return to this issue later.
- <sup>4</sup> Kim (1998, p. 15; pp.119-20). I shall return to physicalism later.
- <sup>5</sup> According to Kim, second order properties inherits its causal power, in any given occasion, from the first order realizer that is instantiated in that specific occasion (Kim 1998, p. 116). Hence, second order properties are causally powerfull, and so they are real, according to Kim's own view about the reality of properties.
- <sup>6</sup> From the fact that physicalist realizationism is a good answer to the mind-body problem in the case of intentional properties it does not follow that it is the only possible answer, nor that it is the best answer for qualitative properties. As I suggested above, it is likely that, given the fact that we accept that there are two different (irreducible) kinds of mental features (the intentional and the qualitative) the answer to the mind-body problem could be different in both cases. I shall return to this point later.
- <sup>7</sup> Kim (1998, pp. 101-2). He also mentions the zombie argument, but he says that this argument is "more controversial". So, in what follows, I will concentrate on the inverted spectrum argument which he thinks is the strongest one.
- <sup>8</sup> It is not clear if the two arguments are logically independent, may be the first one is used as a premise to ground the claim that phenomenal properties are intrinsic, but Kim is not explicit on this point, so I shall consider both arguments separately in the last part of the paper.
- <sup>9</sup> Supervenience and something else, I shall come back to this later.
- <sup>10</sup> Kim (1998, p. 38). It is also interesting to note that while Kim usually defines physicalism in terms of strong (local) supervenience, Chalmers and Jackson usually define physicalism in terms of logical *global* supervenience (Chalmers 1996, p. 38; Jackson 1998, p. 10).

- 11 In this paper I shall use interchangeably the notions of metaphysical, conceptual and logical necessity.
- 12 I will not distinguish between physical and nomological or causal necessity either.
- 13 I will come back to this critique later.
- 14 Chalmers (1996, p. 126). This latter position is compatible with Chalmers own view, but he prefers to call it "naturalist dualism," because he identifies physicalism with strong physicalism.
- 15 According to the distinction drawn by Beckermann (1992).
- 16 According to my reality criterion for properties, if mental (qualitative) properties belong to some basic interlevel laws of nature they are real, independently of the issue of whether they are causally efficacious or not (cf. Note 4).
- 17 May be Kim's refusing to call this answer a physicalist theory is due to the fact that it is not "reductive" enough. But, in my opinion, it fits Kim's own constraints to be a physicalist.
- 18 It is interesting to remark that the original formulation of the identity theory involved exactly the qualitative states of our mental life: the identity was established between sensations and brain processes (Smart 1959). Identity theorists thought that the rest of mental phenomena, like belief, understanding, wanting, etc. could be adequately handled by a behavioristic approach. However, they later extended the identity thesis to all kinds of mental phenomena. I owe this historical point to Eduardo Rabossi.
- 19 See Klagge (1988). N. Stigol, I. Zuberbuller and A. Botterell made me think about this problem.
- 20 I would like to remark that although it makes sense to ask for an explanation for a supervenience relation, it makes no sense to ask for an explanation for the identity relation. As Block and Stalnaker hold: "(...) it makes sense to ask for an explanation of the correlations between the two sets of events, but it does not make the same kind of sense to ask for an explanation of the identity. Identities do not have explanations. (...) The role of identities is to disallow some questions and allow others." (Block & Stalnaker, forthcoming, p. 16). In the same spirit, see Kim (1998, p. 98).
- 21 I follow Chalmers' presentation of these arguments.
- 22 In the case of the zombie situation, we could describe a given situation with exactly the same physical / functional concepts (as the actual one), but all the phenomenal concepts applied to that situation are false (i.e. there are no conscious experiences at all).
- 23 I would like to point out that one of the main defenders of these kinds of arguments, D. Chalmers, uses the notions of properties and concepts interchangeably, but, from my point of view (and it is also Kim's view) they are clearly distinguishable
- 24 By "ordinary," I mean a concept that belongs to our natural language.
- 25 See Hoy (1984): "(...) color, weight and hardness are complex dispositional properties that require mentioning relations with other objects" (pp. 286-7).
- 26 Kim's own definition is given in terms of accompaniment and loneliness: The idea is that extrinsic properties are those that imply accompaniment, whereas intrinsic properties are compatible with loneliness (Kim 1982). But it is not obvious that *qualia* are

the only properties that a "lonely person" could have, not that we could have any qualitative state at all if the world had never existed. Lewis & Langton (1998), trying to improve Kim's definition, give another (more elaborate) definition of "intrinsic property" appealing to the idea of accompaniment and duplication. The definition Lewis and Langton propose is the following:

Say that P is *independent of accompaniment* iff four different cases are possible: something accompanied may have P or lack P, something unaccompanied may have P or lack P. P is *basic intrinsic* iff (1) P and not-P are non-disjunctive and contingent, and (2) P is independent of accompaniment. Two things (actual or possible) are *duplicates* iff they have exactly the same basic intrinsic properties. P is *intrinsic* iff no two duplicates differ with respect to P.

The conclusion is the same with other definitions that we could read in the philosophical literature: there is no clear distinction between intrinsic and extrinsic properties (see, for example, Garrett 1995 and Hoy 1984).

- <sup>27</sup> Juan Rodriguez Larreta defends a point of view like this, and claims that this is the only real solution to the mind-body problem, given the fact that qualia are the only intrinsic properties that there are.
- <sup>28</sup> A similar point is made by Blackburn (1990) who poses the following dilemma: either (i) we accept that there are only dispositional properties all the way down (i.e. that there are no categorical properties at all) or (ii) we accept that the only categorical properties we can know are our own experiences, but those events, "conceived of as categorical play no role in a scientific understanding of the world, they certainly do not serve to ground anything" (p. 65).

## BIBLIOGRAPHY

- Beckermann, A.: 1992, 'Supervenience, Emergence and Reduction', in Beckermann, A., Flohr, H. & Kim, J. (eds.): 1992, *Emergence or Reduction?*, Berlin, DeGruyter.
- Blackburn, S.: 1990, 'Filling in Space', *Analysis* 50 (2), 62-65.
- Block, N.: 1980, 'Introduction: What is functionalism?', in N. Block (ed.): 1980, *Readings in the Philosophy of Psychology*, vol. 1, Cambridge, Harvard University Press.
- Block, N.: 1990, 'Can the Mind Change the World?', in G. Boolos (ed.): 1990, *Meaning and Method*, Cambridge, Cambridge University Press.
- Block, N. & Stalnaker, R.: forthcoming, 'Conceptual analysis, Dualism, and the Explanatory Gap'.
- Borst, C.V. (ed.): 1970, *The Mind-Brain Identity Theory*, London, MacMillan.
- Broncano, F. (ed.): 1995, *La Mente Humana. Enciclopedia Iberoamericana de Filosofia*, Madrid, Editorial Trotta.
- Chalmers, D.: 1996, *The Conscious Mind*, Oxford, Oxford University Press.
- Chalmers, D.: forthcoming, 'The Content and Epistemology of Phenomenal Beliefs', (<http://ling.ucsc.edu/~chalmers/papers/belief.html>).
- Damasio, A.: 1994, *Descartes' Error*, New York, Avon Books.
- Garrett, B.: 1995, 'Extrinsic/intrinsic', in J. Kim and E. Sosa: 1995, *A Companion to Metaphysics*, Oxford, Blackwell.
- Hill, Ch.: 1991, *Sensations. A Defense of Type Materialism*, Cambridge, Cambridge University Press.

- Hoy, R.: 1984, 'Inquiry, intrinsic properties and the identity of indiscernibles', *Synthese*, 61, 275-297.
- Jackson, F.: 1994, 'Armchair Metaphysics', in M. Michaelis and J. O'Leary-Hawthorne (eds.): 1994, *Philosophy in Mind*, Kluger Academic Publishers.
- Jackson, F.: 1998, *From Metaphysics to Ethics. A defence of Conceptual Analysis*, Oxford, Oxford University Press.
- James, W.: 1890, *The Principles of Psychology*, vol. 2, New York, Dover.
- Kim, J.: 1982, 'Psychophysical Supervenience', reprinted in Kim (1993).
- Kim, J.: 1989, 'The Myth of Nonreductive Materialism', reprinted in Kim (1993).
- Kim, J.: 1993, *Supervenience and Mind*, Cambridge, Cambridge University Press.
- Kim, J.: 1996, *Philosophy of Mind*, Oxford, Westview Press.
- Kim, J.: 1998, *Mind in a Physical World*, Cambridge, MIT Press.
- Klagge, J.: 1988, 'Supervenience: Ontological and Ascriptive', *Australasian Journal of Philosophy* 66/4, 461-70.
- Kripke, S.: 1972, *Naming and Necessity*, Cambridge, Harvard University Press.
- Kripke, S.: 1977, 'Identity and Necessity', reprinted in S. Schwartz (ed.): *Naming, necessity and Natural Kinds*, Cornell University Press.
- Levine, J.: 1983, 'Materialism and Qualia: The Explanatory Gap', *Pacific Philosophical Quarterly* 64, 354-361.
- Levine, J.: 1993, 'On Leaving Out What It's Like', in M. Davies, M & G. Humphreys (eds.): *Consciousness. Readings in Mind and Language*, Cambridge, Basil Blackwell.
- Lewis, D.: 1966, 'An Argument for the Identity Theory', *Journal of Philosophy* 63, 17-25.
- Lewis, D.: 1970, 'How to Define Theoretical Terms', *Journal of Philosophy* 67, 427-46.
- Lewis, D.: 1972, 'Psychophysical and Theoretical Identifications', *Australasian Journal of Philosophy* 50, 249-58.
- Lewis, D.: 1998, 'Defining 'Intrinsic'', *Philosophy and Phenomenological Research* LVIII/2.
- Pérez, D.: 1999, *La mente como eslabón causal*, Buenos Aires, Catalogos.
- Nagel, T.: 1998, 'Conceiving the impossible and the mind-body problem', *Philosophy* 73.
- Rabossi, E.: 1995, 'La tesis de la identidad mente-cuerpo', in Broncano (1995).
- Rodriguez Larreta, J.: draft, 'Fisicalismo y qualia: respuesta a Diana Pérez'.
- Sabatés, M.: draft, 'Epiphenomenalism, isolationism and causal criteria for property reality'.
- Smart, J.J.C.: 1959, 'Sensations and Brian Processes', *Philosophical Review* 68, reprinted in Borst (1970).
- Wittgenstein, L.: 1953, *Philosophical Investigations*, Oxford, Basil Blackwell.

*Diana I. Pérez* is currently a professor at the University of Buenos Aires, and an assistant researcher at the CONICET (Argentina). She holds a Ph.D. from the University of Buenos Aires (1995). She has published articles in the area of philosophy of mind and metaphysics in several journals (*Análisis Filosófico*, *Revista Latinoamericana de Filosofía*, *Manuscrito*, *Dialogos*, etc.) and a book on mental causation. Her main topics of research are the metaphysics of mind and the nature of mental concepts.