This is a pre-print version, please do not cite. The final version of this paper is forthcoming in the *Review of Philosophy and Psychology*, and has already been published online: https://link.springer.com/article/10.1007%2Fs13164-018-0424-1

Does the explanatory gap really arise from a fallacy?

Abstract: Many philosophers have tried to defend physicalism concerning phenomenal consciousness, by explaining dualist intuitions within a purely physicalist framework. One of the most common strategies to do so consists in interpreting the alleged "explanatory gap" between phenomenal states and physical states as resulting from a *fallacy*, or a *cognitive illusion*.

In this paper, I argue that we should not interpret the explanatory gap as the result of a fallacy. The explanatory gap does not arise from a fallacy or a cognitive illusion, even though it may very well arise from another kind of illusion (for example, a perceptual-like illusion). This does not imply the falsity of physicalism, but it has consequences on the kind of physicalism we should embrace.

Introduction

It is widely recognized that phenomenal consciousness seems to pose a serious metaphysical problem to physicalism. Even though we may have numerous reasons to think that phenomenal states are nothing over and above with physical states (such as brain states), this continues to strike us as extremely counter-intuitive. For example, I may be convinced by various arguments that experiencing pain is nothing but being in a certain brain state (say, C-fiber activation). However, when I focus introspectively on my current headache, I find myself deeply puzzled by this identity. How can this experience, this subjective feeling, be exactly the same thing as some electrochemical activity taking place in my brain? When we experience this deep puzzlement, it is said that we face the *explanatory gap* (Levine, 1983, 2001), or that we have a strong *intuition of distinctness* (Papineau, 2002) regarding phenomenal and physical states.

Many dualist philosophers have transformed this intuition into *arguments*, designed to show that phenomenal states are indeed distinct from physical states. Physicalist philosophers, on the other hand, have employed different strategies to deal with this intuition. One of the more widespread strategies amounts to understanding this explanatory gap as the result of a *fallacy*, or a *cognitive illusion*. When it seems to us that phenomenal states cannot be identical with

physical states (though they are), we are the victim of a fallacy – we succumb to a kind of reasoning mistake. Various theories have been suggested to explain why we systematically commit such a fallacy (Loar, 1997; Papineau, 1993, 2002, 2007; Tye, 1999).

My goal is to argue against this kind of account. My strategy will be to carefully describe some examples of psychological processes underlying *fallacies* on the one hand, and *valid reasoning* on the other hand. I will point out some psychological features that are distinctive of these two kinds of processes. I will then turn to the process underlying the intuition of distinctness, and I will show that it much more closely resembles the process usually underlying a case of valid reasoning than the process which is typical of a fallacy. This gives us a reason to think that the explanatory gap is not the result of a fallacy. However, even if we accept this conclusion, it does not mean that, if we accept that we *do* have a persistent intuition of distinctness, then this intuition is correct, in the sense that we are in phenomenal states that really are irreducible to physical states. Indeed, there is at least *one* alternative physicalist account of the intuition of distinctness, which sees this intuition as *illusory* (in a way), but locates the illusory component at another level. On this kind of account, we make no cognitive mistake when we come to the conclusion that phenomenal states cannot be identical with physical state. The illusory component of the intuition of distinctness comes from a *perceptual-like* illusion, rather than a cognitive one. Besides, this illusory component does not concern the metaphysical nature of the states we are in, but the more basic question which concerns *which states we are in*. My argument in this paper is primarily directed against the view that the intuition of distinctness arises from a fallacy, but I think it therefore indirectly gives weight to this alternative account.

In the first section, I will describe "Fallacy Accounts", *i.e.* physicalist views that see the intuition of distinctness as the result of a fallacy. David Papineau's account will be described in detail, as it is perhaps the most typical and the most elaborate example of such an account. In the second section, I will show how a physicalist can recognize the existence of the intuition of distinctness without necessarily seeing it as the result of a *fallacy*. I will then describe what I take to be the main physicalist alternative to Fallacy Accounts, which I call "Introspective Illusion Accounts". In a third section, I will point out some typical psychological features of the processes underlying *fallacies* or *cognitive illusions* (as opposed to valid reasoning), using two examples. In the fourth section, I will show, on the basis of the previous analysis, that the process leading to the intuition of distinctness looks much more like a valid reasoning than a

fallacy. In the fifth section, I will quickly come back to Introspective Illusion Accounts. The sixth section will consist in concluding remarks.

1. The explanatory gap as the result of a fallacy

Phenomenal states are states such that *there is something it is like* to be in these states. A visual sensation of green, a gustative sensation of chocolate, a burning sensation of pain on the forearm, etc., are typical examples of phenomenal states. These states bear *phenomenal properties*: properties in virtue of which these states are such that there is something it is like to be in them, and in virtue of which they have a certain *phenomenal character*. For example, a visual sensation of green is a phenomenal state. It has a phenomenal property, that we can call "phenomenal greenness", in virtue of which it is a phenomenal experience of green.[1]

Many philosophers of mind are physicalists: they think that all the mental properties we instantiate are entirely identical with purely physical properties – "physical" being here taken in an extended sense, so that it includes a vast set of properties (physical properties strictly speaking, but also physically realized functional properties, physically grounded properties, properties which logically supervene on physical properties, etc.) are included.[2] However, even convinced physicalists[3] are often puzzled by such an identity thesis in the case of phenomenal properties and phenomenal states. How can my current visual sensation of green, which instantiates phenomenal greenness, be *identical*, and fully reducible, to an objective electrochemical activity taking place in my brain? Even if we believe in it, physicalism still seems counter-intuitive, and to a certain extent *arbitrary*. This problem has been famously labelled the "explanatory gap" by Joseph Levine (Levine, 1983, 2001). David Papineau described the situation in the following way (Papineau, 2002): when trying to accept the identity of phenomenal states with physical states, we face an *intuition of distinctness*. It seems to us that the two kinds of states simply *cannot* be identical.

[1] Here I take phenomenal properties to be properties of *mental states*. Some philosophers prefer to think about them as properties of *subjects* (experiencing subjects). I do not think anything substantial for my paper bears on this distinction; what I say in this paper could be restated in this alternative framework.

[2] I want to make it clear that by "physicalism" I do not intend to refer merely to what is often called "reductive physicalism" (or "identity physicalism"), but also to other, more liberal, forms of physicalism: realization physicalism, supervenience physicalism, grounding physicalism, etc.

[3] The reasons to accept physicalism have generally mostly to do with causal considerations (Levine, 2001, Chapter 1; Papineau, 2002, Chapter 1). I won't expound them here, as my goal is not to argue in favor of physicalism.

To my mind, these two expressions refer to the same thing, and I will use "explanatory gap" and "intuition of distinctness" in an interchangeable way.[4] This explanatory gap/intuition of distinctness is what fuels the various anti-physicalist arguments that have been recently developed on the subject of consciousness (Chalmers, 1996; Jackson, 1982; Kripke, 1980), although philosophers understand this gap in various ways.

Some philosophers have tried to defend physicalism against the pull of this intuition, by showing that we should expect this anti-physicalist intuition to arise even if physicalism is true. They explain this intuition by appealing to certain purely physical features of some of the concepts we use to think about phenomenal states: the so-called "phenomenal concepts", that are notably, but not only, applied through introspection. This way of defending physicalism has been labeled the "Phenomenal Concept Strategy" (Stoljar, 2005), and it has been recently developed in many versions (Aydede & Güzeldere, 2005; Balog, 2012; Diaz-León, 2008, 2010, 2014; Elpidorou, 2013, 2016; Hill, 1997; Hill & McLaughlin, 1999; Levin, 2007; Loar, 1997; Papineau, 1993, 2002, 2007; Schroer, 2010; Sturgeon, 1994, 2000; Tye, 1999).

Many theorists amongst those following this strategy have suggested that the intuition of distinctness should be interpreted as the result of a "fallacy" (Papineau, 1993, 2002, 2007), or a "cognitive illusion" (Tye, 1999). In their view, our phenomenal concepts have some peculiar features. These features are such that we are systematically led to commit a *mistake* when reflecting on the relationship between phenomenal states and brain states: we are mistakenly led to judge that phenomenal states are distinct from brain states, even though they are not.

David Papineau, for example, states that phenomenal concepts present a "use/mention feature": every occurrence of a given phenomenal concept involves the instantiation of the phenomenal property this concept refers to, or at least of a similar property. Therefore, every time I think about a certain type of phenomenal state, using phenomenal concepts, I activate a version of this experience (or at least a "faint copy" (Papineau, 2002, p. 118) of this

[4] I think that this interpretation matches both Levine's and Papineau's opinion. See for example what Levine writes: "Whether we think of [the explanatory gap] as an explanatory gap or a distinctness gap, the problem is really the same" (Levine, 2007, p. 148). See also Papineau (Papineau, 2008, 2011) for the thesis according to which the explanatory gap has to be primarily understood as constituted by the intuition of distinctness. Some philosophers, notably David Chalmers, reject such an understanding. According to David Chalmers, the explanatory gap has primarily to be understood as a matter of lack of *a priori* derivation from physical truths to phenomenal truths, and does not rely on an additional "intuition of distinctness". I am convinced by the arguments presented by David Papineau and Joseph Levine in favor of their understanding of the explanatory gap; my paper therefore supposes that there is more to explain in the explanatory gap than a mere lack of *a priori* derivation.

experience).[5] What happens then when we consider whether a certain phenomenal state (say, an experience of pain) is identical with a certain physical state (say, a C-Fiber activation) ? We make use of two very different concepts: one of them is a phenomenal concept, whose application activates an experience of pain (or a copy of it). The other is a descriptive (physico-biological) concept which does not bring any experience of this kind when it is applied. Therefore, one way of thinking about pain has a distinctive feeling: *it is like something* to think about pain with a phenomenal concept. On the other hand, when I think about pain *as* a C-fiber activation, there is no distinctive feeling associated with my thought. For this reason, as Papineau says, "there is an intuitive sense in which exercises of material concepts 'leave out' the experience at issue. They 'leave out' […] the technicolour phenomenology, in the sense that they don't activate or involve these experiences" (Papineau, 2002, p. 170).

This is where we commit the fallacy, that Papineau calls the "Antipathetic Fallacy":[6] we systematically tend to "project" this phenomenological difference between our two thoughts on the *referents* of these two thoughts. In other words, we can't help inferring, from the fact that our physical understanding of pain "leaves out" something when compared with our phenomenal understanding of the same thing, that the first must refer to something *different* from the second – *that what is left out is the referent itself*, or at least some features of the referent. This is why it irreducibly appears to us that phenomenal concepts and concepts of brain states *must* refer to different states; that phenomenal states and brain states *must be distinct*. This explains the arising of the intuition of distinctness.

Michael Tye gave a somewhat (though not exactly) similar explanation of the explanatory gap (Tye, 1999, p. 712-713), even if he did not call the process by which the dualist intuition arises a "fallacy", but a "cognitive illusion". In this paper, I consider that "fallacy" and "cognitive illusion" refer to the same kind of psychological process: a *mistaken* process, where the mistake takes place at the level of *reasoning*. Fallacies and cognitive illusions can notably be differentiated from *perceptual-like* illusions: the first take place at the level of reasoning (the manipulation of conceptual representations), and concerns the way in which we infer (wrongly) something from something else. The second take place at a lower-level and does not involve a

[5] Papineau has changed the details of his theory over the years, though he maintained his general line of thought (Papineau, 1993, 2002, 2007).

[6] By using this term, Papineau makes a reference to the "Pathetic Fallacy", described by the critic John Ruskin – the fallacy by which we tend to falsely attribute mental states that are our own to inanimate objects. When we commit the "Antipathetic Fallacy" described by Papineau, on the other hand, we (falsely) reject to attribute phenomenological properties to purely physical states.

mistake made when manipulating conceptual representations: if they rely on *inferences*, these inferences are necessarily only *subpersonal*, and the "premises" are encoded in a non-conceptual format, and have a non-conceptual content.

Amongst the proponents of the Phenomenal Concept Strategy, Brian Loar et Katalin Balog also made a similar proposal, and both described the explanatory gap as an "illusion" (Balog, 2012; Loar, 1997, p. 30-31). Although they do not specify the kind of illusion it is, I think that they probably have in mind a kind of *cognitive* (rather than perceptual, or perceptual-like) illusion[7]. All in all, I think it is fair to say that explaining the explanatory gap as the result of a "fallacy" or a "cognitive illusion" is quite in the mainstream in recent philosophy of mind.

For reasons of simplicity, I will now call the views that see the explanatory gap as a result of a fallacy "Fallacy Accounts". Papineau's theory is, to my mind, the most typical example of a Fallacy Account. In what follows, I will argue against Fallacy Accounts, from the point of view of a physicalist.

2. If not a fallacy, then what?

Most of the debates concerning these theories have taken place with the metaphysical question in mind: are our minds purely physical? There have been various arguments designed to show that the explanatory gap should not be considered as an *illusion* – whether because it is a *real*, actual gap, or at least because we don't have good reasons to think that it actually is an illusion. So, most of the philosophical discussion about Fallacy Accounts has been focused on the problem of knowing whether or not it was possible to defend physicalism against the dualist intuition by interpreting this intuition as something illusory (Demircioğlu, 2013; Gertler, 2001; Goff, 2011; Levine, 2007; Nida-Rümelin, 2007). Some critical attention was also given to the peculiar features of phenomenal concepts which, in the various accounts, are supposed to explain the arising of the illusion (Dove & Elpidorou, 2016; Shea, 2014; Sundström, 2008).

[7] Loar describes this illusion as something that is created during "our philosophical ruminations", which seems to confirm that what he had in mind was a *cognitive illusion*, rather than a perceptual one. It is perhaps less clear in Balog's writing, as she sometimes seems to understand the illusion of the explanatory gap as something similar to a perceptual illusion. However, she does not take a clear stance on that question, and she does not explicitly analyze the explanatory gap as resulting of a kind of *perceptual-like* illusion (with everything that comparison implies).

I want to address a slightly different issue here. Let's assume, for the sake of the argument, that physicalism is true of all our mental states and properties.[8] Let's also assume that we *do* encounter a robust and irreducible intuition of distinctness, and that we *do* face an explanatory gap – even though the truth of physicalism implies that there is no gap in reality (no "metaphysical" gap). We can then ask: is the explanatory gap really the result of a *fallacy*? Is it really a *cognitive illusion*, a kind of reasoning mistake we make, that leads us to the intuition of distinctness? Or is the psychological process that gives rise to the intuition of distinctness something different? For example, couldn't it be a process where the illusory component arises in a way which is more similar to a *perceptual* illusion?

I will now try to flesh out this question by formulating at least *one* alternative to Fallacy Accounts. That is to say, I want to describe a kind of psychological process, distinct from a *fallacy* or a *cognitive illusion*, which could also be responsible for the intuition of distinctness if physicalism is true. I will then explain why the precise description of the psychological process that creates this illusion is important, even for those who are primarily interested in the metaphysics of consciousness. In the remaining sections, I will give an argument in favor of the thesis that the intuition of distinctness is *not* the product of a fallacy.

If physicalism is true and if we nevertheless face an intuition of distinctness when we consider it, this means that there *has* to be something illusory, at one point or at another, in this intuition. According to Fallacy Accounts, this illusory component is *cognitive*: it is comparable to a *reasoning* mistake. Roughly, Fallacy Accounts tend to say the following: we introspectively represent phenomenal states, and on the basis of introspection we apply *phenomenal concepts*, which allow us to *think* about our current conscious experiences. The introspective judgments then formed (of the kind: "I am now having an experience of green", thought about with a phenomenal concept) are not mistaken (at least not in a systematic way) and we can expect them to be *true* most of the time. So, introspection is a reliable process (introspection is not systematically illusory), and introspective phenomenal judgments are generally true. But when we *reflect* on the nature of phenomenal states, the peculiarity of phenomenal concepts makes it so that we make *systematic cognitive mistakes*: we systematically judge wrongly that phenomenal states are *not* physical. And it is here, at the cognitive level – at the level of *reasoning*, manipulating conceptual representations – that the illusory component is generated. This is why these accounts are *Fallacy* Accounts. In this kind of view, the explanatory gap is

[8] I think, however, that most of the ideas presented in this paper could be of interest even if someone does not embrace physicalism.

the result of a systematic reasoning mistake, which in the end must belong to the category of *fallacies* or *cognitive illusions* which are studied by psychologists of reasoning (Kahneman, 2012; Pohl, 2004; Peter Wason & Johnson-Laird, 1972).

However, the illusory component could very well arise at a different level. Let's consider for example the following story. We introspectively represent phenomenal states, and on the basis of this introspective process we apply phenomenal concepts. However, introspection is systematically inaccurate: it *misrepresents* phenomenal states. For example, it may represent phenomenal properties as having a qualitative nature that they don't have, and that nothing physical (and therefore nothing real) has (Pereboom, 2009, 2011). The judgments we then form, on the basis of introspection, when we apply phenomenal concepts, may be seen as *systematically* mistaken: they retain the illusory component that arose at the introspective level. When we start to *reason* about the referents of phenomenal concepts, we conclude that phenomenal states, understood as the states which are exactly as presented to us through introspection, cannot be identical with physical states. But this reasoning is not mistaken, and it is not a fallacy. In fact, it is quite the opposite: this reasoning is perfectly correct. In accounts of this kind, which we can label (taking inspiration from Pereboom), "Introspective Illusion Accounts", it is true that phenomenal states (as presented to us through introspection[9]) are not identical with physical states. However, it just happens to be the case that *we never are in any phenomenal states* (at least, not if we understand, by "phenomenal states", states which are exactly as presented to us through introspection). I take Pereboom's theory to be a typical example of an Introspective Illusion Account, even though there are other (mostly scientific) theories that could fit this category (Graziano, 2013; Humphrey, 2011).

I have just laid out two ways of understanding the explanatory gap in a physicalist framework. Both consider this gap as an *illusion*, though they understand the source of this illusion very differently. For Fallacy Accounts, phenomenal introspection is reliable and our introspective judgements, which use phenomenal concepts, are generally true. We *really are* in phenomenal states, such as introspection presents them to be. However, a *cognitive* illusion arises when we try to identify these states with physical states: we reason mistakenly, in a

[9] This caveat is important, because theories which endorse this explanation could try to say that phenomenal concepts have a *dual content*, following Chalmers' distinction between edenic and ordinary content (Chalmers, 2006). That's exactly what Derk Pereboom suggests (Pereboom, 2011). On this view, only phenomenal states understood as states satisfying phenomenal concepts' edenic content would be distinct from physical states – while in some other understanding phenomenal states (understood as states satisfying merely phenomenal concepts' ordinary content) could very well be identical with physical states.

systematic manner, when we think about the metaphysical nature of phenomenal states. On the other hand, according to Introspective Illusion Accounts, phenomenal introspection is inaccurate, as well as the introspective judgments using phenomenal concepts. We *never really are* in phenomenal states (such as introspection presents them to be). When we judge that we are in phenomenal states, we are simply the victims of an *introspective illusion*, which has nothing to do with a reasoning mistake. We then reason perfectly well when we judge that these "phenomenal states" (that we falsely think we are in) are different from physical states. Indeed, it is *true* that, *if these states were to obtain* (they do not), *they would not be identical with physical states*.

These are two ways to understand how a fallacious intuition of distinctness could arise in a purely physical world. I don't think that they exhaust all the ways a physicalist has to deal with this intuition. However, I think these two ways are the two main ones available for a physicalist who seeks to explain away the intuition of distinctness.[10] [11]

[10] Note that the two kinds of accounts I just described differ at two levels. First, they differ because they locate the illusory component in two different aspects of the representational content which constitutes the intuition of distinctness. Introspective Illusion Accounts say that we are wrong about *what mental states we are in*, while Fallacy Accounts say that we are wrong only when it comes to the *metaphysical relation between some of the states we are in and physical states*. Second, they also differ in the interpretations of the psychological process that gives rise to the illusory component of the intuition of distinctness: is it a cognitive illusion, or rather an introspective, perceptual-like illusion? Even though I do not think that there is anything that *logically* prevents another kind of combination of these two factors, I think that the two options I considered (Fallacy Accounts and Introspective Illusion Accounts) are the most "natural". Indeed, it seems that the question of knowing *what mental states we are in* is a "low-level" question (especially when it comes to phenomenal states), which does not rely on a lot of inference on the basis of conceptual representations (as would be required if it was the result of a fallacy), so that it would be hard to see how an illusory component could arise at this level, and yet be a fully *cognitive* illusory component. The same way, it would be hard to understand how a "perceptual-like" illusion could arise concerning a question which is very abstract and theoretical, and requires sophisticated concepts, such as the question which bears on the *metaphysical relation between phenomenal states and physical states*. Georges Rey (Rey, 1995) developed a theory that seems to imply that we are subjected to an illusion concerning *the states we are in* (we falsely think that we instantiate phenomenal properties), and that this illusion is a cognitive one, as it is the result of a kind of "projective" fallacy. However, this kind of position clearly is a minority position. Besides, I must say I am not sure I fully grasp Rey's account. Furthermore, I want to make clear that, even if I set aside the preceding qualifications, the two accounts I discuss do not exhaust the available strategies for the physicalist who wishes to deal with the intuition of distinctness. It is possible to account for the existence of dualist intuitions, neither as the result of cognitive illusions, nor as the result of perceptual-like illusions. For example, one can state that the dualist intuition is a kind of theoretical illusion which comes from some cultural prejudices (for example, Cartesian prejudices). Daniel Dennett often seems to embrace that position (Dennett, 1988, 1991). The disadvantage of this kind of position is that, if it were true, we should expect the intuition of distinctness to disappear, as we abandon our old Cartesian beliefs – and this is not obviously the case. I set aside all these details in my paper, and I simply consider that, in arguing *against* Fallacy Accounts, I give more weight to Introspection Illusion Accounts.

[11] It is important to note here that not all physicalists think that such an intuition deserves an explanation of its own. Some deny its existence (or at least its persistence on reflection), while others think that all that physicalists must explain is the absence of *a priori* derivation of phenomenal truths from physical truths, and that once this absence of *a priori* derivation has been explained the explanatory gap itself is explained, and there is a no need for some further explanation of any kind of dualist intuition (Chalmers, 2002; Papineau, 2008).

Asking which of the two kinds of accounts correctly describes the source of the explanatory gap can be seen as an interesting question in itself: a question that deserves to be answered for its own sake. However, it is important to note that the choice between these two accounts also has consequences regarding the metaphysical nature of consciousness. Indeed, if Fallacy Accounts are correct, introspection is roughly reliable: we *really* are in the kind of mental states that introspection presents us (even though we tend to commit mistakes concerning their metaphysical nature). That naturally leads to a *reductionist* view of consciousness, in which phenomenal consciousness is *real* and can be equated with a physical process (even though it is counter-intuitive, because of our tendency to commit the fallacy). On the other hand, if Introspective Illusion Accounts are correct, that means that introspection is inaccurate, and that we *never are* in mental states such as the mental states that introspection presents to us. We are then led to an *illusionist* conception of consciousness, in which we deny the existence of genuine phenomenal states (Frankish, 2016): phenomenal consciousness is the result of an introspective illusion, and we *never are* in states that are like what is presented to us through introspection. [12] The mental states we really are in have none of the features that phenomenal states seem to have through introspection, and that creates a problem when we try to understand how physical states could have these features (qualitativeness, subjectivity, etc.).

3. Fallacies and valid reasoning

I will now argue against Fallacy Accounts, and in favor of Introspective Illusion Accounts. My strategy will be to analyze the psychological process that gives rise to the explanatory gap, and to show that the reasoning on which it relies displays more similarities with processes underlying cases of *valid reasoning* than with the processes underlying *fallacies*. This will constitute an argument *against* the idea that the explanatory gap arises from a fallacy, and will indirectly weigh in favor of Introspection Illusion Accounts.

Let's examine a classic fallacy studied in psychology of reasoning, called the "conjunction fallacy" (Fisk, 2004; Kahneman, 2012; Tversky & Kahneman, 1983). A character is described to subjects in the following way: "Linda is thirty-one years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations". After

[12] Illusionism does not strictly speaking imply eliminativism, as an illusionist can always suppose that phenomenal concepts *do* refer, in the sense that their *ordinary* content is satisfied (even though their *edenic* content is not).

that, subjects are asked to say which of the two following is the most probable: (1) Linda is a bank teller and (2) Linda is a bank teller and is active in the feminist movement.

It is logically impossible for (2) to be *more* probable than (1), and it is necessarily *less* probable (if we suppose that there is even a very small probability that Linda is a bank teller *not* active in the feminist movement, which is obviously true). There is therefore no doubt that (1) is the right answer. However, according to various repeated studies, about 85 to 90% of people who took this test (undergraduate students at major universities) answered that (2) was more likely (Kahneman, 2012, p. 158). This is a classic example of a *fallacy*, or *cognitive illusion*.

My goal is not to give a general psychological account of fallacies in general, or a particular account of *this fallacy* – whether it is Kahneman's account, which relies on a *dual system* theory, or another account. I simply want to make a few remarks concerning the psychological process that people undergo when they fall prey to this fallacy. This description seems to me to be rather devoid of any theoretical commitment, though I must say that these remarks match Kahneman's descriptions and would fit perfectly well in his account of cognitive illusions; they are also partly inspired by what Rüdiger Pohl says about defining features of cognitive illusions (Pohl, 2004, p. 2-3).

The first thing I want to point out is that our tendency to commit the conjunction fallacy is embodied in an *automatic* psychological process. That means that, even when we try to inhibit it, we cannot help having the tendency to judge that (2) is more probable than (1) – even if we can prevent from judging that (2) is more probable than (1). But that also means (and this is the point I am primarily interested in noting) that this fallacy is *not* the product of careful reflection. Careful reflection about the issue at stake does *not* give rise to the fallacious judgment. On the contrary, careful reflection is what allows us to find the right answer, which is that (1) is more probable than (2) – even though reflection does not suppress the tendency to make the fallacious judgment. So: careful reflection certainly does not suppress the tendency to commit the fallacy (otherwise, this case would not be a genuine case of cognitive illusion), but it does not *produce* the fallacy either.

The second thing I want to remark is that, even though there is obviously a strong and widely shared tendency to commit this fallacy, the *right answer* to the question is perfectly intelligible on careful reflection. By that, I mean that, on careful reflection, we are perfectly able to *identify* the right answer, and we are perfectly able to understand *why* the right answer

is the right answer: why (1) is necessarily more probable than (2), given notably that (2) implies (1) but not conversely. We may still *tend* to judge that (2) is more probable than (1), but we will nevertheless find the right answer perfectly intelligible and unproblematic on careful reflection.

I take it that these two remarks (A) are independent of any particular psychological account of fallacies (even if they fit well with Kahneman's two systems theory); (B) would also apply to other typical fallacies and cognitive illusions (which I don't describe here for reasons of space), such as – to take an example of fallacy which is well-known by philosophers – the fallacy committed by a vast majority of subjects in the famous Wason Selection Task (Paul Wason, 1966).

I now want to describe a very different kind of psychological process. It is a case of *valid reasoning*, where the reasoning is made on the basis of false premises, and the false premises are obtained by way of an unreliable and illusory perceptual process. Let's suppose that a woman, Anne, is about to enter a room. Some people tell her that there is nothing in the room but a white chair; let's say she believes them. They happen to be telling the truth, as the room is really empty, except for a white chair. However, when Anne enters the room, she happens to have a hallucination of a black cat sitting on the white chair.[13] Let's also suppose that, on the basis of this (fallacious) visual experience, she forms a perceptual judgment of the kind: "there is a black cat in this room".

At this point, a little bit of reasoning (which, in most real cases, will probably happen in an extremely quick way) can show her that her previous belief, according to which there was nothing in the room but a white chair, must be abandoned in the light of this new piece of information. For example, she can reason like this: "There is a black object in the room, an object cannot be both black and white, so there is an object in the room which is not white, which means that it's false that there is nothing in the room but a white chair"; or like this: "There is a cat in the room, something cannot be both a chair and a cat, so there is an object in the room which is not a chair, which means that it's false that there is nothing in the room but a white chair", etc. In other words: a little bit of reasoning leads her to consider a contradiction between her previous belief about the room (which happened to be true) and the content of her new judgment that there is a black cat, made on the basis of her fallacious visual experience. It

[13] Perhaps she has this hallucination because she just took (without knowing it) some very elaborate psychoactive drug which causes visual hallucinations, while leaving her reasoning capacity intact; or perhaps because she has been secretly equipped with a sophisticated TMS device that directly stimulates her visual cortex. It does not really matter here.

may very well lead her to *abandon* her previous belief that there is nothing in the room but a white chair.

I now want to make a few remarks on the story I just described. The psychological process which leads Anne to have a visual experience of a black cat is misleading, and it gives rise to an illusory perceptual experience. On the basis of this illusory experience, she will then form a false belief concerning what is in the room. However, the illusory component here is not *cognitive*, but *perceptual-like*: it is her perceptual (visual) system which malfunctions here, and leads her to form this false belief. No reasoning mistake is involved here: she commits no *fallacy*, and she falls prey to no *cognitive* illusion. Notably, she makes absolutely no mistake when her (simple) reasoning allows her to understand that the belief that there is a black cat in the room contradicts the belief that there is nothing in the room but a white chair. This reasoning is perfectly valid: she is fully justified to infer, from the premise that there a black cat in the room, that it is false that there is nothing but a white chair in the room.[14] So, inasmuch as some reasoning is involved when she comes to abandon her previous true belief that there was nothing in the room but a white chair, this reasoning is perfectly valid – it does not rely on a *fallacy*, or on a *cognitive* illusion of any kind. However, the premise on which this reasoning is based is false, and it is obtained by way of a dysfunctional perceptual device. The illusory component here arises at the perceptual level: if there is an illusion here, it is a perceptual, not a cognitive illusion.[15]

I now want to point out some psychological features of the reasoning Anne uses to conclude, from the judgment that there is a black cat in the room, that it's false that there is nothing in the room but a white chair.

First, this reasoning is *sustained* by careful reflection. The more Anne (or anyone in her situation) reflects on this subject, the more it is obvious to her that, if there is a black cat in the room, then it is *impossible* that there is nothing in the room but a white chair.

Second, she can formulate many different specific arguments to reach the same conclusion. For example, as I showed previously, she can base her reasoning on the *color* of

[14] Granted a few obvious definitional truths (concerning colors, kinds of objects, etc.)

[15] Of course, there is another step between the perceptual illusion and the valid reasoning that leads to the conclusion that it is false that there is nothing in the room but a white chair: the step by which Anne "endorses" her (illusory) perception, and judges, on the basis of this perception, that there is a black cat in the room. However, I will set aside this step here, as it does not seem to be a good candidate for the localization of the source of the illusory component *per se* (it seems hard indeed to argue that, in this case, what is dysfunctional in Anne's cognitive system is her tendency to *prima facie* endorse the content of her perceptions; after all, this tendency is normally perfectly reliable and usually leads to correct judgments).

the objects considered (something cannot be both black and white, therefore if there is a black cat, then there is more than just a white chair), on the *kinds* of the objects concerned (a cat cannot be a chair), etc. Even though all of these arguments are extremely simple, it is important to note that we can easily come up with many of them.

Third, these arguments are deductive, which notably has the following consequence: even when Anne thinks hard (and, maybe, *especially* when she thinks hard), she simply cannot understand how the conclusion of these arguments could be false, granted that the premises are true. In this case: if Anne accepts that there is indeed a black cat in the room (and if she accepts some commonly shared definitions), then she cannot make sense of the idea that *it is still true that there is nothing in the room but a white chair*. The proposition according to which there is nothing in the room but a white chair simply becomes unintelligible and incoherent if she endorses the judgment according to which there is a black cat in the room. No matter how hard she tries, she cannot apprehend or picture a situation that would make both of these beliefs true.

I just described two kinds of psychological process. One is a case of *fallacy*, or *cognitive illusion*. The second is a case of valid reasoning, which leads to a false conclusion, and is based on a false premise (obtained through an illusory perceptual process). In the first case, the illusory component arises at the cognitive level. In the second case, it arises at a perceptual level. I tried to highlight some notable features which distinguish these two psychological processes. In what follows, I will focus on the psychological process which leads to the intuition of distinctness, and I will compare it to the two processes I just described, in order to argue against Fallacy Accounts.

4. The intuition of distinctness as the result of valid reasoning

I now want to focus on the psychological process which leads to the intuition of distinctness. My goal is to show that this process has more in common with the process underlying a typical case of valid reasoning than with the process underlying a typical fallacy. I thus intend to give weight to the thesis according to which the intuition of distinctness is not the result of fallacy. If we are physicalists however, we still have to say that the intuition of distinctness, broadly considered[16], has an illusory component. However, we have reasons to think that this illusory

[16] By "the intuition of distinctness broadly considered" I mean the intuition of distinctness considered with a kind of "existential import": not only the intuition that phenomenal states are not physical states, but the intuition that *we are in states* (phenomenal states) that are not physical states.

component does not arise in the way described by proponents of Fallacy Accounts. This indirectly supports the thesis that this illusory component arises in the way described by defenders of Introspective Illusion Accounts.

Let's say that I focus on my current visual experience – for example, on my current visual experience of blue, as I am looking at the painting *Bleu II* by Joan Miró. I then try to think that this experience is *identical* with some electrochemical activity currently taking place in my visual cortex. At this point, I encounter an intuition of distinctness, I face the "explanatory gap": it seems to me that the two things I am thinking about, and that I am trying to identify, *cannot really be identical*. I am deeply puzzled by this identity, and I am very reluctant to accept it, as it seems blatantly false.

Now, let's try to describe the psychological process by which I am led to judge, or at least to be strongly tempted to judge, that my phenomenal experience cannot be identical with a brain state. To begin with, one crucial thing is striking: careful reflection about the objects grasped by my two thoughts (my "phenomenal" thought and my "physical thought" – which according to the physicalist grasp the same object) *does support* the intuition of distinctness. The more I think introspectively about my current experience of blue, the more I meditate on its subjectivity, its qualitativeness, the fact that it is directly felt and experienced, etc., the more it seems obvious to me that *it simply cannot be identical* with a blunt, objective, "blind" physical process.

Of course, further reflection can convince me that physicalism is true. After all, if that were not the case, there would be no physicalists on Earth. However, this further reflection relies on other considerations: it is not merely based on my current grasp of the objects at hand. For example, this further reflection may rely on metaphysical considerations concerning ontological simplicity, or causality, etc. The important point is that, if I consider what I am led to believe simply by *carefully reflecting* on my current experience of blue on the one hand, and on an electrochemical cortical activity on the other hand, I find that such careful reflection *does lead me to the intuition of distinctness*: it leads me to the idea that my experience *cannot be* identical with a brain process.

If indeed the intuition of distinctness is, as I claim, the product of careful reflection, then the psychological process that causes it is more similar to a case of valid reasoning than to a fallacy. Actually, I even think that the intuition of distinctness not only *is* the product of careful reflection, but also that *only* careful reflection produces it. By that, I mean that the intuition of distinctness does *not* arise at first glance, when we merely think quickly and superficially about

the issue at hand. On the contrary, it only appears when we carefully reflect on the nature of the entities we are thinking about.[17] However, this point may be less easy to support, while I take it to be quite plausible that careful reflection *does indeed* produce the intuition of distinctness. When I contemplate attentively my experience of blue, when I think about its various features with great care, I find myself to be very reluctant to identify it with a sheer brain mechanism.

I think that this important point has been often been missed by physicalists working on these questions, perhaps because they have tended to conflate two close issues which should be carefully distinguished. One issue is the nature of the process that leads to the *intuition* of distinctness (intuition according to which consciousness is not physical), and the other one is the nature of the process that leads to the *belief* that consciousness is *indeed* not physical. Of course, physicalists want to say that careful reflection leads us to *abandon* the belief that consciousness is not physical, or at least that it *should* have this effect. After all, if they did not think so, why would they be physicalists? However, this should not preclude them from recognizing that the *intuition of* distinctness, which is not a belief but rather a peculiar *disposition* to believe, which is triggered in certain special conditions, *is indeed produced by careful reflection* on the concerned objects. We can accept this thesis, and yet think that, when we take into account *other considerations*, such as other arguments in favor of physicalism (based on causal considerations, for example), then we are no longer in the special conditions in which careful reflection leads us to the intuition of distinctness, and we can therefore be led by careful reflection (now focused on different objects, or on a larger set of objects) to accept physicalism.

So, this first and crucial feature of the process by which we are led to the intuition of distinctness gives us a reason to think that this process is more akin to a case of valid reasoning, than to a fallacy. Let's now focus on two other features of this process.

When I focus on my current experience of blue, it seems to me that it cannot be identical with an electrochemical activity in my visual cortex. This is precisely the intuition of

[17] I have only anecdotal evidence supporting this claim: when I teach philosophy of mind to undergraduates, I find that, even if many of them are intuitive dualists, they are rarely dualists for reasons specifically related to the hypothetical irreducibility of phenomenal states. They are often reluctant to accept physicalism simply because it seems to them that, by treating human minds as "machines", physicalists cannot account for the creativity and the freedom that human beings possess. However, after some teaching and some thought experiments, which I think aim at triggering careful reflection on the objects considered, my students often start to be puzzled by physicalism *concerning phenomenal consciousness in particular*, and they begin to encounter the intuition of distinctness as I understand it (even though they may very well accept physicalism for other reasons). I think that this anecdotal evidence weighs in favor of the claim that the intuition of distinctness is produced by, and *only* by, careful reflection on the concerned entities.

distinctness. However, if I carefully examine this situation, I find that I am not only *strongly disposed to believe* that the two entities are distinct. I also have a lot of difficulty *understanding how they could be the same*.[18] Even if I try really hard to *accept* this identity, I still am really puzzled and bewildered when I aim at representing *what it would be for this identity to be the case*. There is a sense in which the understanding of this identity systematically *eludes* me; a sense in which I simply don't find it fully intelligible.[19] In other words: I am not only strongly "pulled" towards anti-physicalism, I am also having a hard time picturing what it would mean for physicalism to be true.

Of course, there *has* to be a way to think about physicalism that makes this doctrine perfectly intelligible, but this way may correspond to different conditions of thought – for example, it may imply the use of other concepts. What I want to point out is that, when I introspectively focus on one of my current experiences, it seems to me that this experience is distinct from a brain activity, and it is very difficult for me to fully apprehend how these two things could be *identical*. So, in this respect, too, the psychological process that leads to the intuition of distinctness resembles valid reasoning more than a fallacy, as we find ourselves in trouble when we try to simply *understand* how a certain situation can be true: the fact that my mind is entirely physical while I "have" (according to introspection) subjective and qualitative experiences in one case, or the fact that there is nothing but a white chair in the room while there "is" (according to her illusory perception) a black cat in Anne's case.

Finally, there is a third point I want to highlight. When I focus on my current experience of blue, many reasoning paths can lead me to the intuition of distinctness, i.e. to the idea that this experience cannot be identical with a brain state. For example, I can focus on the *qualitative character* of my experience: nothing in my brain has such a qualitative character, so the two things cannot be identical (by Leibniz' Law). I can follow the same reasoning based on its *subjective character*, the fact that this experience of blue is inherently *for me*.[20] I can also reason

[18] It is important to note that these two psychological facts are quite distinct. Indeed, being strongly disposed to believe that P and having difficulties understanding how not-P could be true, are two different things and the former does not imply the latter. Consider for example the following fact: my current visual experience of my two hands very strongly disposes me to believe that I have two hands.. However, I have no difficulty understanding how, in spite of what I experience, I may *not* have two hands (for example, I can picture a situation in which I am hallucinating); I have no problem apprehending this possible situation.

[19] This is why many people find that the most tempting thing to say when facing physicalism is simply, as Joseph Levine wrote in conclusion of his review of Christopher Hill's (materialist) book on consciousness: "believe it if you can" (Levine, 2011).

[20] For the distinction between qualitative character and subjective character, see (Kriegel, 2005; Levine, 2001, p. 7-9).

on the basis of epistemological or modal considerations, those operative in the Knowledge Argument (Jackson, 1982), or the various Modal Arguments (Chalmers, 1996; Kripke, 1980), to reach the same conclusion.[21] In this respect, the psychological process by way of which I reach the intuition of distinctness resembles the case of valid reasoning previously described. In the example I gave, I could follow many paths to conclude, from the fact that there is a black cat in the room, that it cannot be the case that there is nothing in the room but a white chair. On the other hand, fallacies and cognitive illusions function quite differently. In a fallacy, such as the conjunction fallacy described by Kahneman and Tversky, we simply *jump* to the fallacious conclusion, in a single and simple step, which is difficult to analyze and which does not seem to allow for much variation (at least, not much carefully conducted and examined variation). So, this last feature of the psychological process underlying the intuition of distinctness gives us another reason to think that this intuition is not the result of a fallacy, but stems from something which is closer, from a psychological point of view, to a valid reasoning.

5. Where does the illusion lie?

I have argued that the psychological process leading to the intuition of distinctness is closer to a case of valid reasoning than to a fallacy. This gives us a reason to accept that this process indeed *is* a case of valid reasoning, and not a case of fallacy. Of course, this conclusion could be resisted. After all, it may be that the intuition of distinctness is caused by a fallacy of a very peculiar kind, endowed with some special psychological features in virtue of which it very much resembles a valid reasoning. Without further justification, however, such a move would be *ad hoc.* I think that the burden of proof now lies on the defender of Fallacy Accounts, if she wants to maintain that the intuition of distinctness really arises from a fallacy.

If rejecting to treat the intuition of distinctness as a fallacy meant that we were forced to *endorse* the intuition of distinctness in its strongest sense, *i.e.* to infer a "real" and metaphysical gap from the explanatory gap and to deny physicalism, then the cost of abandoning Fallacy Accounts would be very high for physicalists. Physicalist philosophers would then certainly be tempted by all kinds of *ad hoc* moves in order to save Fallacy Accounts. However, as I tried to

[21] Of course, physicalists reject the conclusion of such arguments, which mean that they have to say that *something* is wrong with these arguments. However, I think that the overwhelming majority of physicalists will grant that the problem with these arguments can hardly be understood as being simply a matter of fallacy, or cognitive illusion. Physicalists who reject these arguments have to reject one of the *premises* of these arguments (and they often go as far as to admit that these false premises still have some kind of *prima facie* rational plausibility).

show previously, there is at least one consistent alternative position for philosophers who want to account for the robustness of the intuition of distinctness, while endorsing physicalism: Introspective Illusion Accounts. According to these accounts, the psychological process which leads to the intuition of distinctness is a process of valid reasoning, which fits well with the conclusion of the comparison I previously put forth. There is an illusory component in the intuition of distinctness broadly considered, but it does not arise at the cognitive level – at the level of *reasoning*. Rather, it arises at the earlier, introspective stage.

The argument I have presented in the paper speaks against Fallacy Accounts, and indirectly in favor of such Introspective Illusion Accounts. In these accounts, we are systematically deceived by introspection, and our phenomenal judgments are therefore all *mistaken*: we never are in states of the kind presented to us in introspection. So, we indeed undergo an illusion concerning consciousness, but this illusion concerns the fact that *we are in phenomenal states*, where "phenomenal states" refers to states that are as presented to us through introspection. However, the reasoning that leads us to conclude that these states are *not* physical states is perfectly valid. Indeed, it is true that these states are endowed with properties that are *not* physical properties: they are intrinsically subjective; they have a qualitative character[22], etc. However, this does not endanger physicalism, as these properties are not instantiated by anything real.

On such a view, when we conclude that *phenomenal consciousness* (where this word refers to the states that are as presented to us through introspection) is not physical, we are perfectly right. In the same way, we are right when we judge that a black cat is not a white chair, and

[22] The defender of Fallacy Accounts may suggest a view in which the fallacy does not arise at the level of the question "are phenomenal states identical with physical states?", but at another (slightly different) level which concerns questions such as "can these features which introspection ascribe to phenomenal states, such as subjectivity, or "qualitative-ness", be purely physical features?" In this view, we would commit indeed no fallacy when we judge that phenomenal states cannot be physical, from the premise that they are qualitative and subjective *and* the premise that subjectivity and qualitative-ness cannot be physical feature. However, the fallacy emerged "earlier", when we judged (fallaciously), when reflecting on subjectivity and qualitative-ness, that subjectivity and qualitative-ness cannot be physical features. However, I think that this view could be targeted by an argument extremely similar to the one I just gave in this paper. Indeed, it could be argued similarly that the kind of process by which we come to think that subjectivity or qualitative-ness cannot be purely physical features, for as much as it involves reasoning and the manipulation of conceptual representations, is much more similar to a process of valid reasoning than to a process of fallacy (cognitive illusion), pretty much in the same respects as the process by which we come to think that phenomenal states cannot be identical with physical states (which I examined in detail in the paper). Indeed, (1) the process by which we come to judge that subjectivity and qualitative-ness cannot be physical *is indeed* the product of careful reflection; (2) it is very hard for us to understand how the content of the judgments at hand could be false (how subjectivity, for example, given the introspective grasp we have of it, could be a purely physical feature); (3) the same conclusion (that subjectivity or qualitative-ness cannot be physical) could be reached through several various arguments (we could reason on the categorical, modal, epistemological, properties of subjectivity or qualitative-ness, etc.). Thanks to […] for raising this point in correspondence.

that, if there really is a black cat in a room, then it cannot be true that there is *nothing but a white chair* in it. However, in both cases, the problem is that there isn't really a black cat, or a phenomenal state, present. We make no reasoning mistake when we reach the intuition of distinctness. The mistake is made by our introspective faculty, which represents us as being in states in which we are not. The mistake is not cognitive, and it is made *before* we even start to reason. It is an *introspective* mistake, closer to a *perceptual* mistake (and to a perceptual illusion) than to a fallacy.

I think that Fallacy Accounts do not correctly describe the psychological process that leads to the intuition of distinctness, while Introspective Illusion Accounts do a much better job. However, that does not mean that Introspective Illusion Accounts create no difficulties. First of all, they have *illusionist* consequences regarding consciousness. To many philosophers, this constitutes a disadvantage of such accounts in itself: the very idea that phenomenal consciousness is, in a certain sense, an illusion, strikes them as preposterous.

Secondly, they have the consequence that *the very idea that we are conscious*, in the sense of being in phenomenal states that are as introspection presents them, is an illusion. The reality of phenomenal states is supposed to be an illusion, roughly in the same way that the presence of a black cat sitting on a white chair in the previous example is an illusion. But here is a problem: if phenomenal consciousness is an illusion, for example if my current phenomenal experience of pain (say) is merely an illusion, then I should be able to easily comprehend that this may be the case. I am not saying that I should be easily *convinced* that it is the case, but rather that I should have no trouble *representing* to myself that it is the case, the same way that, when I visually experience a black cat (or something else), I have no trouble envisioning the hypothesis that this perception *could be illusory*, even if the cat *seems real* (and even if I believe that it is real). The problem is that, in the case of consciousness, we have a hard time figuring out what it would mean for our phenomenal states to be purely illusory, quite in the same way that we have a hard time figuring out what it would mean for them to be purely physical. This has been often noted by philosophers who have asserted that, in the case of phenomenal consciousness, it is impossible to distinguish between *appearance* and *reality* (Husserl, 1963; Kripke, 1980; Merleau-Ponty, 1968; Searle, 1997). I am not saying here that it is *true* that appearance and reality coincide in the case of consciousness, but simply that we *do have* a strong intuition that this is the case, and that we have trouble apprehending that it may not be the case. And it is not easy to understand why this intuition should arise if phenomenal states were simply *illusory*, in the same sense that my visual experience of a black cat is illusory.

So, Introspective Illusion Accounts are not without problems. I tried elsewhere to show how these problems could be tackled (see Author's article, ****). Even if it is true, however, that Introspective Illusion Accounts face serious problems, it is also true that Fallacy Accounts are not psychologically plausible. The psychological process which leads to the intuition of distinctness looks very different from the one by which we commit fallacies. This consideration gives us a reason to reject to treat the explanatory gap as resulting from a fallacy – which, again, does not mean that we should abandon physicalism, and treat this gap as the indication of a "real", metaphysical gap between two existing things.

6. Concluding remarks

In this paper, I examined the widely accepted view that the explanatory gap (or "intuition of distinctness") is nothing but the result of a fallacy. I criticized this view by pointing out that the psychological process which leads us to the explanatory gap is quite different, in many respects, from the processes usually underlying fallacies. I showed that it resembles much more the kind of psychological processes underlying *valid reasoning*. In my view, this gives us a reason to reject what I called "Fallacy Accounts" regarding the intuition of distinctness. However, I tried to show that rejecting Fallacy Accounts does not necessarily mean endorsing the intuition of distinctness, in the sense that it does not mean that we should accept that physicalism is false. We can reject Fallacy Accounts, without having to infer, from the explanatory gap, the existence of a real, metaphysical gap. Indeed, there is at least *one* alternative kind of account to Fallacy Accounts, which allows to defend physicalism by accounting for the arising of the intuition of distinctness in a purely physicalist framework: what I called "Introspective Illusion Accounts". In this kind of view, the intuition of distinctness is the product of perfectly valid reasoning, and in a way it says something *true*. In this perspective, it is indeed *true* that phenomenal states, understood as states which are exactly as they are presented to us through introspection, are not physical states. However, physicalism is still true, given that *we are not really* in such phenomenal states. In this kind of account, our illusions regarding consciousness do not arise at a cognitive level, and do not concern the metaphysical nature of consciousness; they arise at an introspective (perceptual-like level) and concern the very states in which we are supposed to be. The reasoning by which we find that phenomenal states are not (and cannot be) physical states, on the other hand, is perfectly valid.

My argument was primarily directed against Fallacy Accounts, and I hope to have shown that such accounts are not psychologically plausible. I also tried to give at the same time some reasons to embrace Introspective Illusion Accounts, which I think constitute the most plausible and the most interesting alternative to Fallacy Accounts.

References

Aydede, M., & Güzeldere, G. (2005). Cognitive Architecture, Concepts, and Introspection: An Information-Theoretic Solution to the Problem of Phenomenal Consciousness. *Noûs*, *39*(2), 197-255.

Balog, K. (2012). Acquaintance and the Mind-Body problem. In C. Hill & S. Gozzano (Ed.), *New Perspectives on Type Identity: The Mental and the Physical*. Cambridge University Press.

Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.

Chalmers, D. (2002). Consciousness and its Place in Nature. In D. Chalmers (Ed.), *Philosophy of Mind: Classical and Contemporary Readings*. Oxford University Press.

Chalmers, D. (2006). Perception and the Fall from Eden. In T. Szabo & J. Hawthorne (Ed.), *Perceptual Experience* (p. 49-125). Oxford: Oxford University Press.

Demircioğlu, E. (2013). Physicalism and Phenomenal Concepts. *Philosophical Studies*, *165*(1), 257-277.

Dennett, D. (1988). Quining Qualia. In A. Marcel & E. Bisiach (Ed.), *Consciousness in Modern Science*. Oxford University Press.

Dennett, D. (1991). *Consciousness Explained*. Penguin.

Diaz-León, E. (2008). Defending the Phenomenal Concept Strategy. *Australasian Journal of Philosophy*, *86*(4), 597-610.

Diaz-León, E. (2010). Can Phenomenal Concepts Explain The Epistemic Gap? *Mind*, *119*(476), 933-951.

Diaz-León, E. (2014). Do a Posteriori Physicalists Get Our Phenomenal Concepts Wrong? *Ratio*, *27*(1), 1-16.

Dove, G., & Elpidorou, A. (2016). Embodied conceivability: how to keep the Phenomenal Concept Strategy grounded. *Mind and Language, 31*(5), 580-611.

Elpidorou, A. (2013). Having it both ways: consciousness, unique not otherworldy. *Philosophia*, *41*(4), 1181-1203.

Elpidorou, A. (2016). A posteriori physicalism and introspection. *Pacific Philosophical Quarterly*, *97*(4), 474-500.

Fisk, J. (2004). Conjunction Fallacy. In R. Pohl (Ed.), *Cognitive Illusions. A Handbook on Fallacies and Biases in Thinking, Judgment and Memory* (p. 23-42). Hove, East Sussex: Psychology Press.

Frankish, K. (2016). Illusionism as a Theory of Consciousness. *Journal of Consciousness Studies*, *23*(11-12), 11-39.

Gertler, B. (2001). The Explanatory Gap Is Not an Illusion: Reply to Michael Tye. *Mind*, *110*(439), 689-694.

Goff, P. (2011). A Posteriori Physicalists Get Our Phenomenal Concepts Wrong. *Australasian Journal of Philosophy*, *89*(2), 191-209.

Graziano, M. (2013). *Consciousness and the Social Brain*. Oxford: Oxford University Press.

Hill, C. (1997). Imaginability, Conceivability, Possibility and the Mind-Body Problem. *Philosophical Studies*, *87*, 61-85.

Hill, C., & McLaughlin, B. (1999). There Are Fewer Things in Reality Than Are Dreamt of in Chalmers' Philosophy. *Philosophy and Phenomenological Research*, *59*(2), 445-454.

Humphrey, N. (2011). *Soul Dust: The Magic of Consciousness*. Princeton: Princeton University Press.

Husserl, E. (1963). *Ideas: a General Introduction to Pure Phenomenology*. (W. R. Boyce Gibson, Trad.). New York: Collier Books.

Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly, 32*(April), 127-136.

Kahneman, D. (2012). *Thinking, Fast and Slow*. Penguin.

Kriegel, U. (2005). Naturalizing Subjective Character. *Philosophy and Phenomenological Research*, *71*, 23-56.

Kripke, S. (1980). *Naming and Necessity*. Harvard University Press.

Levin, J. (2007). What is a Phenomenal Concept? In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press.

Levine, J. (1983). Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly*, *64*(October), 354-61.

Levine, J. (2001). *Purple Haze: The Puzzle of Consciousness*. Oxford University Press.

Levine, J. (2007). Phenomenal Concepts and the Materialist Constraint. In T. Alter & S. Walter (Ed.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press.

Levine, J. (2011). Review de Consciousness, by Christopher S. Hill. *Mind*, *120*(478), 527-530.

Loar, B. (1997). Phenomenal States (Revised Version). In N. Block, O. Flanagan, & G. Güzeldere (Ed.), *The Nature of Consciousness* (p. 597-616). MIT Press.

Merleau-Ponty, M. (1968). *The Visible and the Invisible*. (A. Lingis, Trad.). Evanston: Northwestern University Press.

Nida-Rümelin, M. (2007). Grasping Phenomenal Properties. In T. Alter & S. Walter (Ed.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press.

Papineau, D. (1993). Physicalism, Consciousness, and the Antipathetic Fallacy. *Australasian Journal of Philosophy*, *71*, 169-183.

Papineau, D. (2002). *Thinking about Consciousness*. Oxford University Press.

Papineau, D. (2007). Phenomenal and Perceptual Concepts. In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press.

Papineau, D. (2008). Explanatory Gaps and Dualist Intuitions. In L. Weiskrantz & M. Davies (Ed.), *Frontiers of Consciousness*. Oxford: Oxford University Press.

Papineau, D. (2011). What Exactly is the Explanatory Gap? *Philosophia*, *39*(1), 5-19.

Pereboom, D. (2009). Consciousness and Introspective Inaccuracy. In L. Jorgensen & S. Newlands (Ed.), *Appearance, Reality, and the Good: Themes from the Philosophy of Robert M. Adams* (p. 156-187). Oxford University Press.

Pereboom, D. (2011). *Consciousness and the Prospects of Physicalism*. Oxford University Press.

Pohl, R. (2004). Introduction: Cognitive illusions. In R. Pohl (Ed.), *Cognitive Illusions. A Handbook on Fallacies and Biases in Thinking, Judgement and Memory*. Hove, East Sussex: Psychology Press.

Rey, G. (1995). Towards a Projectivist Account of Conscious Experience. In T. Metzinger (Ed.), *Conscious Experience*. Paderborn: Ferdinand Schoningh.

Schroer, R. (2010). Where's the Beef? Phenomenal Concepts as Both Demonstrative and Substantial. *Australasian Journal of Philosophy*, *88*(3), 505-522.

Searle, J. (1997). *The Mystery of Consciousness*. New York: The New York Review of Books.

Shea, N. (2014). Using phenomenal concepts to explain away the intuition of contingency. *Philosophical Psychology*, *27*(4), 553-570.

Stoljar, D. (2005). Physicalism and phenomenal concepts. *Mind and Language*, *20*(2), 296-302.

Sturgeon, S. (1994). The Epistemic View of Subjectivity. *The Journal of Philosophy*, *91*(5), 221-235.

Sturgeon, S. (2000). *Matters of Mind*. London: Routledge.

Sundström, P. (2008). Is the mystery an illusion? Papineau on the problem of consciousness. *Synthese*, *163*(2), 133-143.

Tversky, A., & Kahneman, D. (1983). Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. *Psychological Review*, *90*, 293-315.

Tye, M. (1999). Phenomenal consciousness : the explanatory gap as a cognitive illusion. *Mind*, *108*(432), 705-725.

Wason, P. (1966). Reasoning. In B. Foss (Ed.), *New horizons in psychology 1*. Harmondsworth: Penguin.

Wason, P., & Johnson-Laird, P. (1972). *Psychology of Reasoning : Structure and Content*. Cambridge (Mass.): Harvard University Press.