The Reality of Colours and the Knowledge Argument

Linda Zsuzsa Lázár

Department of Philosophy

Central European University

In partial fulfilment of the requirements for the degree of Masters of Arts in Philosophy

Supervisor: Nenad Miščević

Budapest, Hungary

2011

Abstract

In this thesis I link together two issues, the knowledge argument against physicalism and the debate between colour realist and subjectivists. My aim is to show that the knowledge argument can be defended against one, otherwise particularly attractive, physicalist objection, the one that draws on direct realist intentionalism, which identifies the phenomenal character of experience with physical properties of the perceived object. This answer of course requires physicalist colour realism. In 1991 Boghossian and Velleman proposed a general argument against all kinds of physicalist-realist theories of colour. I argue, however, that their argument does not cover direct realistintentionalism about colour. Against this theory in particular, I propose an argument involving cases of phenomenal variation, similar to the argument that Cohen (2009) proposed for his relational theory of colour. I will defend the argument against some objections and show that the direct realist intentionalism is highly implausible, given that it must draw a distinction between veridical colour perceptions and misperceptions which is arbitrary and cannot be explained with the resources of a physicalist-causal theory of colour representation. I will conclude that the notion of 'transparency' on which the theory is based cannot be plausibly maintained, so the answer to the knowledge argument which is based on it will not succeed.

Contents

Introduction

This thesis is about two issues which are both very extensively discussed but rarely linked. The first is my main interest. It is the question whether the existence of the subjective qualitative character of experience constitutes evidence for the hypothesis that the ontology of the world is not entirely physical. The second is the question whether the objects around us that we perceive as coloured are really coloured. Stated this way, it is not obvious how these two issues are related, more specifically, how the second is significant for the first. In this thesis I will elaborate this relation, and claim that if the case for colour irrealism can be successfully defended it will provide a strong support for the dualist conclusion of the knowledge argument by reducing the available physicalist responses to it.

The subjective qualitative character of experience, or quale (in plural qualia), is a problem for physicalism mainly for two closely related reasons. The first is that qualia seem to resist functionalist reduction. Functionalism would be a particularly attractive way of making the reduction because it can plausibly accommodate multiple realisability, and if it is not available as a means to reduce qualia, then the physicalist has to fall back to much less plausible alternatives. The second is that there seems to be a direct argument to the conclusion that qualia cannot be physical properties, i.e. the knowledge argument. Probably the most famous version of it was put forward by Frank Jackson (1982). The argument is very simple: Imagine a perfect colour scientist, call her Mary, who knows everything that completed physics (understood as containing every other scientific discipline which is reducible to physics) had to say about colour and colour vision. Physical science can be learned discursively, so Mary has learned all she knows in a black-and-white environment, and she has never experienced colour. Now imagine that she gets released from this environment, and, at last, sees a coloured object. Jackson claimed that at this very moment Mary learns something new, i.e. she learns about the *subjective qualitative character of colour vision*. Since beforehand she didn't know what it was like to see colours, even though she knew the whole of physics, this bit of new knowledge is *not part of physics*. So there is more to reality than

even completed physics can grasp.

This is one of the most discussed arguments in the philosophy of mind, and physicalist philosophers offered a variety of different objections to it. These include the mere denial of the intuition that Mary learns something new (Dennett 1991), or insisting that there is an equivocation of the verb 'know' in the premises and therefore the argument is not valid (Nemirow 1990, Lewis 1990, Conee 1994). Objectors of the argument often admit that Mary acquires new knowledge, but it is either not an objective fact (Crane 2003), or an old (physical) fact under a new mode of presentation (Tye 1995). It is interesting to note that Jackson himself changed his mind about the interpretation of his thought experiment (Jackson 1995). He revoked his previous commitment to the view that the argument establishes property dualism, and adopted *representationalism* or *intentionalism* about qualia.

The intentional theory of qualia treats the perceptual states with a phenomenal quality as a form of intentionality, i.e. the quality of some (or maybe all) mental states that they are directed at objects. Just as words have meanings that reach out from the words themselves and capture things, perceptions as intentional states capture things out there in the world. The qualities represented by perceptual experiences are not the qualities of the experiences themselves, but the qualities of the intentional objects. Supposing that upon her release Mary is shown a ripe tomato, the redness she experiences is the redness of the tomato, a physical property of a physical object.

Now this is the point where the debate between colour realists and colour irrealists comes into the picture. *Colour irrealism* is exactly the position that the red Mary experiences cannot be the red of the tomato itself. This position is essentially that colour is the product of the mind in response to some environmental stimuli. There may be a property in the external objects that causally contributes to these stimuli, but the objects themselves are not coloured. This position was held by many notable theorists of colour, including Galileo, who wrote that

> I think that tastes, odours, colours, and so on are no more than mere names so far as the object in which we place them is concerned, and that they reside only in the consciousness. Hence, if the living creatures were removed, all these qualities would be wiped away and annihilated. (Galileo 1957 [1623], p. 274)

There is a lively debate going on between colour irrealists on one side, and colour realists on the other, who insist that when we have a colourful visual experience we see the colours that the external objects possess.[1] The debate between the two sides can be seen as being about whether (one particularly strong version of) the intentionalist theory of qualia can be true of colour-qualia.

A new turn in the course of this debate was brought about by Paul Boghossian and David Velleman (1991) who proposed an exhaustive classification of possible representationalist realist theories of colour, and argued that none of the types of theory obtained by way of their classification can accommodate what they call the 'epistemology and phenomenology of colour perception'. The specific epistemological feature of colour perception they have in mind is the fact that there are certain things we know about colours simply in virtue of being the subjects of colour perception, and this knowledge is not open to empirical correction. The specific phenomenological feature of colour perception they have in mind is that there is no distinction between the perceived colour of an object and the property of the perception itself (the way it feels) in virtue of which the perception is the perception of an object so coloured. They argue that there is no theory that would render colours as the objective physical properties of objects represented in the mind by intentional states that could account for these epistemological and phenomenological features simultaneously.

I argue that the debate about colour realism is highly relevant to the question whether physicalists can plausibly respond to the knowledge argument. If Boghossian and Velleman were right, then the case made by the knowledge argument for property dualism is stronger than it

[1] About the state of the debate between colour realism and irrealism see Byrne and Hilbert (2003), Maund (2008).

otherwise would be, because a supposedly very attractive way for the physicalist to resist the dualist conclusion (notably the way Jackson, the original proponent of the knowledge argument eventually chose) is blocked by colour irrealism. However, as it turns out the argument Boghossian and Velleman put forward does not cover a particularly strong, qualia-eliminativist version of intentionalism, which I will call direct realist intentionalism. So, on behalf of colour irrealism, I will offer an independent argument, somewhat similar to that put forward by Jonathan Cohen (2009) in favour of his relationalist theory of colour, to show that this kind of intentionality involves an arbitrary distinction between veridical perceptions and misperceptions, which makes it impossible for him to maintain its core concept of transparency.

This specification of the task delineates the structure of the thesis. In Chapter 1 I make some clarifications about colour irrealism, and I will also briefly review some objections to the knowledge argument to motivate the discussion of the problem of colour realism. In Chapter 2 I look into the details of the physicalist answer to the knowledge argument which I find the most appealing, the intentionalist theory of qualia. I will distinguish between different versions of intentionalism, and reviewing objections targeted at the specific versions offered as answers to the knowledge argument, I will conclude that only the version I call direct realist intentionalism has a chance to provide such an answer. In Chapter 3 I examine Boghossian and Velleman's argument for colour irrealism, and conclude that their argument does not cover the kind of intentionalism that can relevantly answer the knowledge argument. Then I give an argument against direct realist intentionalism from perceptual variation, with which I hope to establish that the distinction between veridical and non-veridical cases of colour perception to which the direct realist intentionalist must appeal cannot be plausibly maintained.

1 Colour Irrealism and the Problem of Qualia

The Relation of 'Physical' and 'Phenomenal' Colour

Colour irrealism is the theory that objects I now see in this room are not coloured. This is a striking theory. It seems to be part of the essence of a pair of jeans that it is blue. Peter and I recognise our otherwise identical backpacks by colour. The books on the shelves by the walls make the walls that would otherwise be monotonous and boring, colourful and homely. It is very hard to doubt that these things are really coloured. Colours also seem essential to the survival of many species. Some birds attract their mates with their richly coloured feathers. Some insects discriminate the flowers that give them food from those that don't by colour, and thereby colour contributes to the survival of both species. We recognise poisonous mushrooms partly by their colour, and probably so did our ancestors through many centuries. It seems that colour discrimination is a highly adaptive feature of living organisms, for which we were selected through thousands of generations. How could it be the case that the things do not really have colours?

There is no doubt that we, and many other species, have the capacity to discriminate the wavelengths of the light that is reflected from the surface of the objects in our environment. It is also clear that many things, and types of things, that were crucial for the survival of our ancestors, have surface-structures that have a steady tendency to reflect light in a certain way. Plausibly, there are microphysically describable properties upon which these tendencies supervene. It is understood that the cells in our retina that are differentially sensitive to different wavelengths send electrochemically transmitted signals to certain brain areas where these signals are processed and combined with many different other bits of information, from which a signal sent towards the skeletal muscles may arise and cause us to move in ways which further our survival. Colours in this sense, understood as the stable tendency of object surfaces to reflect

different wavelengths differently, and thereby providing us with information that we may use for moving around and surviving in our environment, certainly exist. Colour irrealism questions something else.

Is colour irrealism a theory about the subjective character of what happens to us when we undergo processes involving wavelength discrimination? "There is something it is like" for us to undergo such processes (cf. Nagel 1974). Colours, as we are primarily aware of them, are the subjective qualities of wavelength discrimination involving sensory experiences. Now, again, there is little question about the reality of colours as the subjective qualities of visual experience. I can doubt that there is an external world around me, in the way I normally take it for granted. I may be dreaming, I may be the victim of an evil demon, I can be wired up in the Matrix, in which cases external reality, to which I may have no access, may be just grey. What is unchangingly true, however, throughout these scenarios, is that I have the experience of vivid red, light and deep blue, lively yellow, comforting green and all the rest. So colour irrealism isn't a theory about the real existence or nonexistence of colours in this phenomenal sense either.

Colour realism and irrealism are rather theories about the *relation* of the above two senses of colour. There is little doubt that there are surfaces that reflect light in specific ways, or that organisms capable of discriminating different wavelengths are using this information. And there is no doubt either that we have subjective colour experience. The question is how the two are related.

Colour realists claim that the latter is dependent on the former not only in that the latter is (normally) caused by the former, but that (in normal circumstances) the subjectively experienced colours are the colours of the objects. Colour vision reveals a quality of the objects perceived, i.e. that they are themselves coloured. Colour, the subjective quality of experience, is no different from colour, the objective quality of the coloured thing that is being perceived.

The Phenomenal Character of Experience

My interest in colour irrealism comes from my interest in whether physicalism is true. The two topics are linked by what is often called the qualia problem. "Qualia" is a name for the subjectively felt character of experience. So colour, in the second of the senses introduced above, is a kind of qualia. Qualia, according to some philosophers, are a trouble for physicalism. There are multiple ways to say why this would be so.

One way goes back to our discussion of colour in the first of the two senses of colour introduced earlier. It is clear that the surfaces of environmental objects reflect light in specific ways, that our eyes are differentially sensitive to the different wavelengths in the reflected light, that the information about the composition-by-wavelength of the reflected light is transmitted to and processed by the brain, and that this gives rise to survival furthering behaviour. There seems to be no doubt that colour so understood exists. What is unclear, however, is why this purely functional process gives rise to colour in the second of the previous senses. In David Chalmers's very expressive formulation of the question: "Why is the performance of these functions accompanied by experience? [...] Why doesn't all this information-processing go on 'in the dark', free of any inner feel?" (2007, p. 228.) We should note, among other things, that evolution cannot provide an explanation for this. What is needed for survival is only the performance of the functional processes. Once the information-processing mechanism is in place, it is irrelevant whether it is accompanied by conscious subjective experience.

Perhaps the operation of wavelength-discriminatory information processing in the central nervous system of some higher animals is metaphysically bound to give rise to colour experience. But it is hard to see why this would be so. It seems perfectly conceivable that the functional processes can be performed without accompanying phenomenal consciousness. A robot built from usual electronic devices that performs all these functions is certainly conceivable. Does it have any special significance if the functional processes are realised in protein-based structures, similar to the structure of our molecular constitution? Again, why would it be so? If this isn't so,

does the robot built out of usual IT stuff have conscious colour experience? Or else, isn't it possible that we have exact physical duplicates that perform the functional processes with no accompanying colour experience (zombies) (cf. Chalmers 1996)? When the physicalist claims that conscious subjective colour experience is linked to the performance of the colour-discriminatory information processing by metaphysical necessity, our intuitions may diverge on whether we should want to give credit to such a suggestion. To me it seems very plausible that if qualitatively loaded experiential states are realised by physical states, then they are multiply realisable (it is hard to believe that then they are not realisable in silicon-based Alpha Centaurians), and the metaphysical fact that opens the way for multiple realisability is that such experiential states reduce to functional states (cf. Harman 1990). But it seems clear that not only our present knowledge of the functional processes is incapable of providing an explanation for the emergence of conscious experience, but so will be any later stage of neuroscience, if it is conceived as the further perfection of such functional explanations. This is often referred to as the explanatory gap between the understanding physical sciences give us of ourselves and phenomenal consciousness (Levine 1983).

Another way of stating why qualia are trouble for physicalism has to do with our knowledge of the phenomenal qualities of our experience. The most discussed anti-physicalist argument from our knowledge of qualia is formulated in the context of colour experience. Frank Jackson (1982) asked us to imagine Mary, a perfect colour scientist, who was kept from birth in a black-and-white environment, and educated through a black-and-white TV set. Eventually, she gets released. But only after she learnt everything physics had to say about colour and colour vision. Here "physics" is to be understood in a wide sense, including "everything in completed physics, chemistry, and neurophysiology, and all there is to know about the causal and relational facts consequent upon all this, including of course functional roles" (Jackson 1986, p. 291). Now, according to Jackson, it would be hard to deny that, when Mary gets released from her black-and-white prison, and sees, say, a ripe tomato for the first time, she learns something new. She learns

what red is like, or what it is like to see red. Even though she already had a complete physical knowledge of colours and vision (physical in the broad sense), this knowledge she obtains upon her release is something new. The upshot is that knowing what it is like to see something red is not something that could be deduced even from complete physical knowledge. So, complete physical truth is not the whole story about reality. So physicalism is false.

Some Physicalist Objections to the Knowledge Argument

Now, there are a number of ways physicalists attempted to resist this argument. It is not my purpose here to review all the physicalist answers to the knowledge argument and refute them one by one. What I am concerned with in this thesis is one particular type of answer, the one that interestingly convinced Frank Jackson himself to change his mind about his argument, i.e. answering the knowledge argument by invoking intentionalism (or representationalism as Jackson prefers to call it). Nevertheless I review the main other kinds of physicalist answers briefly in order to motivate the discussion of the answer stemming from intentionalism. For, in my view, these ways do not succeed in giving a credible answer to the knowledge argument on behalf of physicalism, which indicates that if the intentionalist answer also fails, then we have a strong case for accepting the argument's dualist conclusion.

To see how the objections work, it is useful to consider a formal reconstruction of the knowledge argument. A simple formal reconstruction can be stated as follows:

(1) Mary knows every physical fact about colour and colour perception before her release.

(2) Having left the black-and-white room she acquired new knowledge about colour.

Therefore,

(3) (from 1 and 2) the way colours look is a non-physical fact to know about colours.

(4) If not every fact is physical, then physicalism is false.

Therefore,

(5) (from 3 and 4) physicalism is false.

The physicalist may start by attacking premise 1 right away. He may try to claim that physics (in the relevant sense) will never be complete, or maybe he might claim that it is not possible for any human being to know everything expressible in the language of physics. The physicalist does not have to argue that this is the case. He just has to point out that it is a possibility, and then the knowledge argument is not a legitimate thought experiment. Howard Robinson (1996), however, pointed out that the proponent of the knowledge argument can easily avoid this objection by modifying the argument slightly, in a way which does not affect its conclusion. Robinson proposes to modify the first premise along these lines: (1') Take any set of the facts expressible in the broadly physical language we are considering that anyone can ever now, and is relevant to colour perception, and suppose that Mary knows all these facts. Robinson also claims that it is not plausible to assume that there are facts expressible in the language of physics which no one can ever know. There would be no principled reason for such an assumption. So with this modified premise the argument shows that Mary learns something new upon release relative to any finite set of physical facts, no matter how that set is delineated. This equally establishes that the phenomenal nature of colour experience cannot be expressed in the language of physics.

Another way is to deny the intuition that Mary learns something new upon her release, that is, premise two. Daniel Dennett (1991) claimed that this intuition rests on a lack of imagination. Those who have this intuition cannot imagine how rich the completed physical science of colour vision will possibly be. Dennett reports that he, with his greater powers of imagination, has the intuition that upon her release Mary won't learn anything new. He illustrates his intuition with a counter thought experiment. In his story, instead of a tomato, Mary is shown a banana when she gets released. However, her captors want to trick her, and they paint the banana blue. To much of their surprise, Mary does not react to the sight of the blue banana by exclaiming "Oh, so this is

what yellow is like". Instead she says, "Bananas are yellow but this one is blue." Her shamefaced captors ask her to explain how she knew.

> 'Simple,' she replies. 'You have to remember that I know everything – absolutely everything – that could ever be known about the causes and effects of colour vision. So of course before you brought the banana in, I had already written down, in exquisite detail exactly what physical impression a yellow object or a blue object…would make on my nervous system. So I already knew exactly what thoughts I would have. … I was not the slightest surprised by my experience of blue. … I realize that it is hard for you to imagine that I could know so much about my reactive dispositions that the way blue affected me came as no surprise.' (pp. 399-400)

It is far from clear, however, as Howard Robinson pointed out, that Mary's complete knowledge of what her reactions would be when she is shown blue amounts to knowing what blue is like, i.e. what feel would accompany the reactions. The functional account of the thought 'that is blue', Robinson says, may fall short of capturing its full content. "Mary can understand the functionally defined recognitional thought without grasping the nature of the phenomenon recognised" (1993, p. 176). This, of course, takes us back to the problem of the explanatory gap. It is hard to see how the further perfection of the functionalist understanding of colour perception would yield an understanding of the conscious phenomenon.

Others responded by acknowledging that Mary learns something new, but not in the sense of propositional knowledge. Some claimed that Mary learns just abilities to recognise, discriminate and remember colours (Lewis 1990, Nemirow 1990). Others argued that Mary's new acquisition is knowledge in the sense of acquaintance (like being acquainted with someone we have already met, or with a town we have already visited, cf. Conee 1994). Proponents of the

ability and the acquaintance hypothesis attack the argument by saying that the entailment from premises 1 and 2 to premise 3 is invalid because it rests on an *equivocation on 'know'*. Tim Crane (2003) argued, however, and I agree, that even if it is true that Mary acquires new abilities (and a similar response can be given to the acquaintance view), this is not all she learns. There is also the piece of knowledge mentioned in the second premise that can be expressed by saying that "Red looks like this". "Red looks like this" can be true or false, so it is a *proposition*. So knowing that "red looks like this" is propositional knowledge. So the inference from premises 1 and 2 to 3 is valid. For a detailed response to this kind of objection against the knowledge argument see Brie Gertler 1999.

Yet others conceded that the knowledge Mary acquires is propositional, i.e. it is a knowledge of a fact, namely the fact which Mary might express by saying "Aha, so red looks like this", but argued that this is not a new fact relative to the facts she already knew by having learnt all that physics has to teach about colour, but an old fact in a new guise, grasped under a new mode of presentation (Tye 1995). This is not a problem for physicalism, as long as physicalism is conceived as the doctrine that all facts are physical.

Tim Crane (2003) even agreed to regard Mary to have learnt about a *new* fact, and that this new fact is *not physical*, but he claimed that it does not endanger the doctrine of physicalism. For physicalism must not be a thesis about all facts being *objective facts*. "Physicalism does not need to say that *physics must state all the facts*." (2003) Some facts are *subjective*, and the fact expressed by the sentence "So red looks like this" is one of them. Physicalism, as a theory about the ontological constitution of the world, is in no danger, as long as all the objective facts, i.e. the instantiation of properties by particulars, are physical.

These objections can be viewed as essentially claiming that the inference from premises 3 and 4 to 5 in the above reconstruction of the knowledge argument is invalid, because it involves an equivocation on the term 'fact'. The objectors say that the sense of 'fact' which is involved in the statement that the existence of non-physical propositional knowledge establishes the

existence of non-physical facts, is not the same sense of fact which is involved in the statement that the existence of non-physical facts would entail the falsity of physicalism. As Michael Tye put it, "Sometimes it is used to pick out real-world states of affairs alone; sometimes it is used for such states of affairs under certain conceptualizations (2009)." The first we could call the "extensional use of 'fact'", the second the "intensional use". The physicalist can say that "extensional facts", i.e. real-world states of affairs, are all physical, and allow for "intensional facts" that have no reduction to the physical. With this move the physicalist can maintain both substance and property monism, he only has to allow for conceptual dualism. There are subjective facts, exactly the way Thomas Nagel (1974) famously observed, that cannot be given an objective, physical understanding. But accepting that the intensional fact that "red looks like this" is a non-physical (subjective) fact is consistent with the supposition that extensionally speaking it is nevertheless *identical with a physical fact*. So the knowledge argument proves only concept dualism, leaving open the possibility that the metaphysical identity thesis is true.

However, the ontological constitution of the world could only remain untouched by the existence of subjective facts if they are reducible to objective facts. So one is back with the question how a mental state with a qualitative nature can be constituted by a neural (and so deeply physical) state. For the reasons that have been stated earlier, i.e. for the concerns that arise from the plausible assumption of multiple realisability, it is very natural to assume that the phenomenal state must reduce to a functional state, which then is realised by one or another physical state. But then we are back with the puzzle that Chalmers has famously raised, because it is very hard to see why should any functional state be accompanied by a phenomenal quality.

For the purposes of this thesis I will regard the Dennettian and the 'equivocation-on-know' objections to have been successfully answered by Robinson and Gertler or Crane, respectively. In the sequel I will consider the 'equivocation-on-fact' answer as the only live option. I will also assume that if there is a way for the physicalist to avoid claiming that experiential states directly reduce to functional brain states, which for some metaphysical reason are bound to be

phenomenal, then he had better avoid it. In the next chapter, I would like to point out that intentionalism combined with realism about the intentional content of perceptual states provides the physicalist with the resources to avoid the implausibility of the functional reduction of phenomenal states that Chalmers pointed out. The reduction for which intentionalism opens the way will equally be a functionalist one, but only indirectly, through intentionality – provided that intentionality has a functionalist reduction. To return to the colour case, by way of intentionality, colour realism offers a very natural and economical way for the physicalist to propose an account on which the intensional fact that "red looks like this" is extensionally identical to a physical fact.

2 Intentionalism about Qualia

The Transparency of Experience

Colour realists are committed to the view that objects around us are really coloured. This means that the colours we subjectively experience (colour qualia) are not different from real physical properties instantiated in the objects of the real physical external world. This way qualia come out as not something over and above the regular physical properties that physicalism may allow. Quite to the contrary, the red I am now experiencing is then the red of the blanket in front of me, a physical property of a perfectly physical ontology. The fact that conscious colour experience has a certain qualitative character is not a fact to account for over and above the facts that objects are coloured, and that these colours can figure in the content of our mental states. Our mental states can grasp them by virtue of their quite remarkable feature called *intentionality*. If this is so, then colour qualia are, in one sense, *eliminated*. It is because it is not the case that colour-experiential mental states have *intrinsic* colour properties. Rather, it is the case that such mental states are *transparent* in the sense that by introspecting them we see the colour of the objects outside. So colour qualia, as intrinsic properties of certain mental states over and above the normal properties of the world, are explained away.

This is how Michael Tye, in an oft-cited passage, describes this intuition about the transparency of colour experience:

> Standing on the beach in Santa Barbara a couple of summers ago on a bright sunny day, I found myself transfixed by the intense blue of the Pacific Ocean. Was I not here delighting in the phenomenal aspects of my visual experience? And if I was, doesn't this show that there are visual qualia?
>
> I am not convinced. It seems to me that what I found so pleasing in the

> above instance, what I was focusing on, as it were, were a certain shade and intensity of the colour blue. I experienced blue as a property of the ocean not as a property of my experience. My experience itself certainly wasn't blue. Rather, it was an experience which represented the ocean as blue. What I was really delighting in, then, were specific aspects of the content of the experience.

Tye goes on explaining that this is how it is when we try to introspect the qualitative feel of experience:

> When one tries to focus on it in introspection one cannot help but see right through it so that what one actually ends up attending to is the real colour blue. (1992, p. 160)[2]

This may suggest a general, eliminativist strategy for the physicalist to deal with qualia. By analogy, just as colour realism places colour out of the mind, into the external physical world, this strategy does the same also with other kinds of qualia. The qualities revealed when we introspect our perceptual states, are the qualities that experience *represents* or *reports of*, not the qualities *of* experience. With the exception of misrepresentations, they belong to external objects. So when we are looking inside, i.e. introspect our perceptual states and experience the phenomenal qualities of experience, we are really looking outside, and experience what the objects of perception out there in the world are like.

This general strategy may be called the *intentional theory of qualia*. It treats the perceptual states with an intrinsic phenomenal quality as a form of intentionality, i.e. the quality of (some or maybe all) mental states that they are directed at objects, and that the objects and their properties

[2] See a similar explanation of the same idea of transparency in Harman 1990.

are somehow grasped by these mental states. As Tye put it, the idea is that "just as the meaning of a word is not a quality the word possesses, so the phenomenal character of an experience is not a quality the experience possesses" (2009, Section 7).

This strategy, of course, trades one problem for another. It trades the problem of accounting for the intrinsic qualitative character of experience for the problem of accounting for intentionality. Clearly, many physicalists share the view that by this move they get closer to a fully physicalist account of our mental life. It is far from clear, however, that intentionalism is an 'easy problem' for physicalism compared to the 'hard problem' of qualia.

On the first page of his (1981) Hilary Putnam famously describes an ant, which, crawling in the sand, accidentally draws some lines what we would take to be a picture of Winston Churchill. Then Putnam asks what the conditions are for this 'drawing' to be *about* Churchill. He concludes that unless the ant intentionally drew the picture, or there is an onlooker, who takes it to be a picture of Churchill, it isn't about Churchill. So for the lines in the sand to have intentionality, there must be an earlier intentionality on which it may depend. Putnam writes,

> But to have the intention that *anything*, even private language (even the words 'Winston Churchill' spoken in my mind and not out loud), should represent Churchill, I must have been able to think about Churchill in the first place. If lines in the sand, noises, etc., cannot 'in themselves' represent anything, then how is it that thought forms can 'in themselves' represent anything? Or can they? How can a thought reach out and 'grasp' what is external?

The point is that we may understand derived intentionality, but the "original intentionality" from which it derives remains a mystery.

Original intentionality can hardly be part of an ultimate ontology of the world that a physicalist may allow for. Jerry Fodor (1987) put the problem this way:

> I suppose that sooner or later the physicists will complete the catalogue they've been compiling of the ultimate and irreducible properties of things. When they do, the likes of *spin, charm,* and *charge* will perhaps appear on the list. But *aboutness* surely won't; intentionality simply doesn't go that deep. (p. 97)

So the physicalist must come up with a suggestion as to what the physical properties are on which original intentionality supervenes. I am not sure if it is an easier task than accounting for phenomenal qualia. Yet, I do not want to pursue this matter here, and I will proceed on the assumption that this problem can be solved.

In the following I will look into the details of different intentionalist theories and see whether they are valid objections against the knowledge argument, and if so, what other possibilities remain for the dualist for defence.

Intentionalist Theories

Jackson himself, as I mentioned earlier, turned away from his original thought experiment later and embraced a kind of the intentionalist (representationalist) theory regarding qualia (also referred to as the Australian view of colours). He claims (1998, 2003) to have identified why the intuition that Mary learns something upon release that is not deducible from her complete physical knowledge, i.e. how red looks like arises: it rests on a misconception of the nature of sensory experience. The misconception is that we think that there is such a thing as the intrinsic phenomenal character of experience, in this case, its intrinsic redness. The right conception, according to the intentionalist theory of experience, is that what it is for an experience to have a phenomenal character is exhausted by its having a representational or intentional character (the adjectives 'representational' and 'intentional' are used interchangeably throughout this chapter). All facts about the phenomenal character of a colour experience concern its representational

character. He supplements this thesis by the further thesis that all facts about the representational or intentional character of an experience can be deduced from the physical knowledge Mary acquired in the black-and-white room. These two theses entail that facts about the phenomenal character of colour experience can be deduced from the physical facts, so the Dennettian answer to the knowledge argument turns out to be correct eventually. Mary knows 'what it is like' to see red even before she gets released.

Following Torin Alter (2007), let us call the first thesis (that the phenomenal character of an experience is exhausted by its intentional character) J1, the second (that facts about intentional character are all deducible from discursively learned physics) J2. (Alter introduces these notations with reference to Jackson.) Let's start by getting a clear understanding of what is meant by J1. To understand what is meant by the claim that the phenomenal character of a state of mind is determined, or exhausted by, its intentional character, it is useful to distinguish between intentional object, intentional content, and the whole intentional nature of a state of mind, including the mode in which its content is being represented. Here I follow Tim Crane's (2007) explanation of these concepts.

The *object* of an intentional state is what the state is about, or directed upon. Every thought is about something, so every thought has an intentional object. The same thought could have several different objects at the same time. Consider the thought that the cat is sitting on the fence. It is about the cat and about the fence. An intentional object can be 'merely intentional' meaning that it is not real. (Consider the thought about a unicorn sitting on the fence.)

Various different intentional states can be directed upon the same intentional object. The same thing can be thought, desired, hoped and so on. There are different '*intentional modes*'.

Even in the same mode, the same object can be presented to the mind in an intentional state in different ways. My desire to play Bloons Tower Defense 4 can be presented to my mind as a desire to play a game in which colourful balloons are being popped by strategically positioned equipment, and also as a desire to play the game we played together with Peter. Every

object that is being represented is represented in some way. No object can be represented in no particular way whatever. The particular way in which the object is represented is called intentional or representational *content*. Some add that the object is either so as it is represented or it is not, so the contents of intentional states are assessable as true or false, thus, content is propositional.

Now, intentionalism is the view that an experience's phenomenal character is determined by its intentional character. One version of it, called 'pure representationalism' (Chalmers 2004) or 'pure intentionalism' (Crane 2007) is the thesis that an experience's phenomenal character is determined by its intentional *content*. A state has a phenomenal character when there is something it is like to be in that state. The representational or intentional content of a mental state is how it represents the world to be. The idea is that there is a very strong and close relationship between the two.

Pure intentionalism comes in two forms, depending on how this close relation is conceived. 'Strong pure intentionalism' simply identifies the phenomenal character of a mental state with its representational content – what it is for as experience to have a certain phenomenal character is simply to have a certain intentional content (Tye 1995). 'Weak pure intentionalism' makes a claim that is weaker then identity: it says only that phenomenal character is determined by, or supervenes on, the representational content – two experiences which share the same intentional content share also phenomenal character (Byrne 2001).

Impure intentionalism, on the other hand, has it that phenomenal character is determined by the intentional content and mode together: seeing that it is raining and hearing that it is raining have different phenomenal characters (although they arguably have the same intentional content), since seeing and hearing are different intentional modes (Crane 2007).

'Qualia theory' then can be obtained as the rejection of intentionalism, pure or impure (weak or strong). Someone can maybe accept the thesis that all mental states with a phenomenal character are intentional, but nonetheless reject the claim that their phenomenal character supervenes on, or identical with, their intentional character (content, or content plus mode),

because they hold that there are non-intentional qualitative properties which contribute to their phenomenal character, and in which two mental states with the same intentional character can differ.

Now let us see how exactly the non-deducibility intuition that is in the centre of the knowledge argument is explained away on the different understandings of J1 – corresponding to the different versions of intentionalism – in combination with J2.

Consider first the 'impure' version of intentionalism. Is J2 plausible if J1 is understood this way? Without intentionalism, that is, on the 'qualia theory', we naively thought that when released Mary learned something that she couldn't have previously deduced from her complete physical knowledge, i.e. the intrinsic phenomenal quality of seeing red. Now we hear that there is no such quality. However, according to impure intentionalism, the phenomenal character of Mary's seeing a red tomato for the first time is determined in part by the distinctive mode of intentionality in which seeing red represents the redness of the tomato. Could she deduce the distinctive phenomenal character of representing redness in that way on the ground of her complete physical knowledge acquired in the black-and-white room? As Alter (2007) has pointed out, the intuition that was at the heart of the knowledge argument before we became aware of intentionalism, i.e. that what it is like to see red cannot be deduced from discursive physical knowledge, is carried over with an equal force to this case. On impure intentionalism what it is like to see red is not the intrinsic phenomenal quality of the experience of seeing red, rather it is the phenomenal character of representing redness in a specific mode, which is generated in part by the special feature of this mode of representation that it presents redness phenomenally, unlike representing redness in a physical discourse with concepts about a type of surface structure with specific reflectance characteristics, which is void of any phenomenal feel. The knowledge argument needs to be modified only slightly. We cannot say that when released Mary learns about certain phenomenal properties of experience, but instead we can say that she learns about the first-person subjective features of the phenomenal mode of representation in which

the colour-sighted represent redness (Chalmers 2004, Alter 2007). The intuition that she couldn't have deduced it from her physical knowledge is as strong as it was in the original case.

Let us consider now pure intentionalism. This is the same thesis as what Alter calls 'ultrastrong representationalism', the thesis that *representational* properties, such as the property that something is represented in a certain mode, do not contribute to the phenomenal character of experience, rather, it is determined fully by *represented* properties. All facts about the phenomenal character of experiencing something red are determined then by (or maybe even identical with) the properties the experience represent, in this case the redness of the object we are looking at. Does what Alter said about impure intentionalism in relation to the intuitive appeal of the knowledge argument hold in this case too?

Alter thinks it does. He argues that the intuition that Mary learns something upon being released from her black-and-white room remains unaffected. She learns about some pure intentional properties. If it is indeed something new relative to her previous all-encompassing physical knowledge, then these pure intentional properties are not physical. So the knowledge argument stands as it did before, even if we assume pure intentionalism. So, Alter concludes, the question whether the knowledge argument is sound and the question whether intentionalism holds are orthogonal.

Here I disagree. Suppose that what Crane calls 'strong pure intentionalism' holds. This is the thesis that it is not only the case that the properties being represented determine the phenomenal character of the experience, but that the two are simply identical: the experience is phenomenally like what the intentional object of the experience is represented to be like. Now add to this thesis that, at least in the cases of veridical perception, the real properties of external things figure in the intentional content, in our case, the phenomenal colour quality of Mary's experience is simply the redness of the tomato she sees. This thesis we may call 'direct realism by way of intentionalism', or 'direct realist intentionalism' for short. (Cf. Brown 2010. Gilbert Harman (1990) seems to be a proponent of this view.)

Maybe Alter is right that the intuition that phenomenal redness is not deducible from discursive physical knowledge remains unaffected even if this is the case. But the physicalist may accept it, and concede that a Denettian answer will not do against the knowledge argument after all. But maybe direct realist intentionalism gives him the resources for an argument to the effect that all that it shows is that there is a gap between linguistic physicalism on the one hand, and metaphysical physicalism on the other, as following Terrence Horgan (1984) Owen Flanagan (1992) put it. All that the non-deducibility intuition, if correct, establishes is that the story that can be told about the world in the language of physics is not the whole story, and it is not the same thing as to say that there are other objects or properties or states of affairs than those constituted by the elements of a purely physical ontology. Take the example of the phenomenal redness of Mary's first colourful visual experience. Maybe it cannot be deduced even from completed physics what it is like. Nevertheless, if direct realist intentionalism holds, it is just the redness of the tomato Mary is looking at. The tomato is a perfectly well-behaving physical object. So presumably, its redness is a perfectly legitimate physical property, and the phenomenal red character of Mary's experience is just identical with it.

This would be the transparency thesis (see the quote from Tye above). The transparency thesis is that we cannot focus our attention introspectively to a phenomenal quality of our experience without thereby focusing on a quality of the object of the intentional state we are in. The qualities of experience are not features that the mind somehow adds to what is being represented in experience. These qualities are the qualities of what we see, hear, taste, smell or touch. So, in the cases when the intentional object of our perception is real, and when it is represented veridically, then these are qualities which are really 'out there' in the physical world.

The distinction between this theory and other theories of perception can be regarded as the difference between *generative* and *selective* theories (cf. Howard Robinson 1994, pp. 66-7). Both types of theories agree that perception is generated in a causal process triggered in our nervous system by the presence of an appropriately placed object having some crucial properties.

According to the 'generative' theories the result of the causal process is a neural state which is sufficient to produce the experience of perceiving the object, including the subjectively felt character of the experience. According to 'selective' theories, on the other hand, the causal process triggered by the presence of the object results in an act of perception whose content – rather than by *sui generis* qualitative features – are constituted by the features of the external world at which the act of perception is directed. The causal process enables us to pick out the (already existent) qualitative content of perception, rather than generate it (hence the term 'selective theory' in contrast to 'generative theory'). So 'direct realist intentionalism' can also be called 'selective'.

Direct Realist Intentionalism

But what can a 'selective' strong pure intentionalist (that is, a direct realist intentionalist) say about the cases of misperception or illusion? I think, he could account for them along the following lines. Let us consider first the cases of misperception. Consider that they are in an important way analogous to the cases of non-phenomenal misrepresentation. Suppose you see a horse, and in your non-phenomenal mental system of representation, which we may call your language of thought, you token the thought *horse* in response. Suppose that being a horse is a physical property. So your thought *horse* means a physical property. If there is a perfectly physical neural state that realises your thought *horse*, and if there is a physicalist reduction for what it is for your thought *horse* to mean the property of being a horse, then the property of being a horse itself will not cause any extra trouble for physicalism. It is a physical property, presented to the mind through a relation of representation which is physical, and the mental state that does the representing is physical too.

Now what if the object of your visual perception is far away from where you stand, and, although you take it for a horse, it is really a cow? (The example I borrow from Fodor 1987.) You see the cow, and in your language of thought you token the thought *horse*. The property of being

a horse, the meaning of your thought, is not instantiated now – there is no horse there, that thing is a cow. Did this misrepresentation of yours give rise to something genuinely, irreducibly mental (perhaps the property of being a horse as applicable to horses that exist only in the head)? Is this case more problematic for physicalism then the previous one was? I don't think so. Again, your thought *horse* is realised by a neural state. Supposing that the physicalist can account for misrepresentations (without ruining the account he gave for veridical representations), he can account for the relation the thought *horse* bares to the cow you actually see. Apart from these, the only remaining components in the story are the cow – a harmless physical being, and the property of being a horse, which we have previously agreed to consider a physical property. It seems that the whole story is accounted for, and in the course of accounting for it no irreducibly mental element entered the account at any point.

There is certainly an analogous story for representations with a phenomenal character. Suppose there is a neural state that realises the mental state of seeing the redness of a ripe tomato that you are looking at. Suppose that there is a physicalist reduction for the intentional relation this neutrally realised state bares to the redness of that tomato, that is, a physicalist story about how this state brings the property red in, so to speak, and presents it to your mind. Now, if we suppose, finally, that being red is a physical property, now instantiated by the tomato, a harmless physical object, then, again, we have a fully physical story of what is taking place.

Now, what if your seeing red at some place is a misrepresentation? Suppose you have been looking at a green traffic light for some time, and when you turn your head away and look at a white wall, you seem to see a red patch on it of a shape somewhat like the shape of a tomato. (This example I borrow from a discussion with Professor Ben-Yami, in which he explained how a direct realist would account for this case.) If there is a perfectly physical neural state you are in, if there is a perfectly physical account for what it is for this state to represent redness, then there is no extra problem for physicalism to account for the redness you seem to see, which can be solved only by introducing the irreducibly mental property of red applicable to objects that exist

only in one's visual field but not in reality. This is a case of misrepresentation, but the physicalist can account for it. You've been looking at the green traffic light too long and the cones in your retina that are sensitive to green light got tired. It means that when you next look at a white wall, the signal your retina will send to your brain will be as if it was red in the area of your visual field where the traffic light was previously, as it is understood that white light is composed of lightwaves with wavelengths ranging over the whole visible spectrum, and if you take green out of it, which is what in effect happens if your green cones get tired, it will look red. So there is a perfectly physical story about the relation your neural representation of redness bares to the whiteness of the wall you actually see. Apart from your neural state and this relation just mentioned, the remaining components of the story are the whiteness of the wall – which we can assume to be physical, and the property of being red, which we have previously agreed to consider physical. The whole story is accounted for, without making use of anything irreducibly mental.

I must say that the intuition that if there is an intentional object which is red, then there is redness in the story which is not instantiated by anything physical, but which is nevertheless there in some clearly non-negligible sense, so there must be some irreducible phenomenal redness involved, lingers on, or at least this is how it is with me. I also think that there is a good reason for this intuition to linger.

According to the physicalist's story, instead of being an intrinsic quality, the phenomenal redness of an experience is conceived now as a relational property it bears to its intentional object, i.e. something that is red. But in the case of misrepresentation, it is a relational property that it bears to something which doesn't really exist. So it is a relation to what? The only sensible answer to this question is that it is a relational property the experience as an intentional state usually, when everything is normal, bears to the real redness of something real. But how can a relational property be carried over to a situation in which the other relatum is not present, without it being grounded in something intrinsic? In our case, how can the relational property

that a neural state is usually (but not now) related by a relational property (called intentionality) to the real property of being red, be carried over to a situation in which the same neural state is not related to anything really red? The only conceivable answer is that it is indeed carried over by some intrinsic feature of that neural state, in virtue of which it enters into the relation of intentionality with real redness when everything is normal and redness is present. Then it is in virtue of this intrinsic feature of the neural state that it functions as the neural sign of redness. Alright, but then why is this feature (presumably the obtaining of a particular neurophysical pattern) is a sign of redness? We could answer that it is because it is somehow metaphysically bound to give rise to a phenomenally red feeling in anybody in whom it obtains. But this answer is not allowed here, because we are in the business of trying to avoid talk of intrinsic phenomenal redness, and replace it with a relational property this state usually bears to red things. But if it is not allowed to talk of intrinsic redness, and if being confronted with red things is not the only way that this state can arise (its relatedness to real redness obtains most of the time but far from always), then why is it a sign of redness?

This problem of course parallels the problem in the physicalist reduction of meaning why the thought *horse* means horse, and not, say, horse-or-cow, if it can be causally related to both horses and cows. This is a general problem of the physicalist reduction of intentionality – presumably via some causal relation. We are operating now under the assumption that this reduction can be done – an assumption to which I do not subscribe. So let us set this worry aside for the moment, and see how to proceed if this assumption is granted to the physicalist, and the account he has given about the cases of misrepresentation is accepted.

This account then can clearly be extended to cases other than misrepresentation – to imagination and hallucination. These are cases when the apparatus whose normal function is to represent things as they really are gets triggered not by an object that is for some reason or other gets misrepresented, but in other ways. As long as there is a physicalist account of how they get activated these will cause no extra problem for physicalism. The phenomenal qualities of

imagined or hallucinated objects or scenes will be those, or the combination of those qualities that the intentional states that get activated normally 'bring in' from the outside world. In order to show that accounting for these will not require reference to irreducibly mental properties the physicalist will only need to retell essentially the same story he did in the case of misrepresentations.

But all this is dependent on the truth of the thesis that the phenomenal qualities that experience presents us with in normal cases of perception are the physical qualities of the objects perceived. It works only if the qualitative nature of being conscious all derives from our capacity of being aware of the qualities of the physical world outside. This thesis, I think, should be the upshot of 'selective strong pure intentionalism' if it was to defuse the intuitions that are at the heart of the arguments from the phenomenal qualities of consciousness against physicalism, including the knowledge argument. But this is a highly suspicious thesis.

For the intentionalist strategy to work generally, all kinds of qualia must be so that they can be given such an intentional account. To establish such a claim one has to go a long way away from common sense, because prima facie qualia just don't seem to be intentional. But to oppose intentionalism about qualia as a general strategy to avoid the problem qualia pose for physicalism, one doesn't need to argue that all kinds of qualia are non-intentional. It is enough if some types are. In the case of some types of phenomenal concepts, arguably, there are no physical properties that could plausibly be thought to be the ones that are picked out by them. First of all, there are the conscious states with a qualitative character, which, at first sight at least, appear to have no object, so it is not easy to see how the intentionalist thesis could be applied to them. Bodily sensations like pain, or certain emotions, feelings or moods, like feeling gloomy for example, clearly have a discernible qualitative character, while, to say the least, it requires an explanation from the part of the intentionalist to see what their object can be. If even one of these conscious states with a qualitative character turn out to be non-intentional, then the intentionalist strategy to render the qualitative character of consciousness as something being imported, through

intentionality, from the real and qualitatively loaded world outside – which can also serve as a physicalist strategy, supposing that intentionality can be given a physicalist reduction – will fail.

Just to mention another point briefly, if qualia were really the qualities of objects perceived, we could misrepresent them. But it doesn't seem to be possible. One cannot be in mistake about the subjective character of one's experience. For example, it doesn't seem possible to be in error about whether we are in pain or not. Pointing out that my C-fibres are not firing doesn't prove me wrong if I feel pain. It is just irrelevant. Even if an MRI machine detects that the neural activity normally associated with having an experience of certain phenomenal quality, one's introspective evidence overrules the evidence provided by the machine. If one believes that *p*, then *p*, if *p* is a proposition stating the phenomenal character of an experience.

Now there is an on-going debate about whether bodily sensations, emotions and moods can all be interpreted intentionally. But I will not review this debate here, or attempt to take it to a conclusion. (But see Crane 1998 taking one side, and Aydede 2001 taking the other.) I think it can be omitted because I think there is a possibility that is equally damaging to the intentionalist-physicalist strategy, i.e. if it is the case that a large domain of perception turns out to present us with qualities which cannot be identified with qualities outside in the real world. And this is the point where the problem of colour realism comes into the picture.

Colour concepts look prima facie intentional. Colour irrealists, however, argue that they aren't. They argue that colour concepts cannot stand for qualities of objects in the external world. If other ways of dealing with the problems posed by qualia are blocked, the defensibility of colour irrealism may be decisive for the question whether physicalism can withstand these problems.

3 A Case for Colour Irrealism

In this chapter I will assess an argument that is purported to show that all types of physicalist-realist theory of colour are bound to fail. This argument is a much discussed one due to Paul Boghossian and David Velleman. In 1991 they offered a two-by-two classification of physicalist theories on the ground of how they conceive of the mental representation of colour properties, and whether they hold that colour properties are identical, or just realised by, physical surface properties. Then they famously argue that none of the four types of theory obtained by these divisions is capable of accounting for the special epistemic features of our knowledge of colours as subjects of colour experience, and for what they call the 'phenomenology of colour experience', i.e. that there is no sensation involved in colour experience over and above how objects look.

If this argument was successful that would be the end for the kind of intentionalist-realist answer to the knowledge argument we are considering. I will argue, however, that it is exactly this type of physicalist theory that escapes Boghossian and Velleman's argument. More precisely, I will argue that what Boghossian and Velleman say cuts no ice again 'Russellian' theories on which the qualitative character of colour experience does not belong with the intrinsic features of the mental states that represent colour.

I will offer an argument, however, that the direct realism involved in such intentionalist theories is highly implausible, given that it must draw on a distinction between veridical colour perceptions and misperceptions which is arbitrary and cannot be explained with the resources of a physicalist-causal theory of colour representation. Jonathan Cohen has recently offered a similar argument to support his relationalist theory of colour. The upshot of his argument is different from mine, but both of us are appealing to the same sort of phenomena, and some of the objections that have been raised against Cohen's argument can equally be raised against mine. In the final sections of the chapter I will give my arguments against these objections, and I will

conclude that the direct realist intentionalist theory of the qualitative nature of colour experience is based on a notion of 'transparency' which cannot be plausibly maintained.

A Fourfold Division of Physicalist Theories of Colour

Paul Boghossian and David Velleman (1991) proposed an argument from the special epistemic features of our knowledge of colours to the effect that our colour concepts cannot be the representations of any physical properties. There are things, they argue, that we know about colours simply in virtue of being the subjects of colour experience. We know, for example, that red and orange are properties; they are different properties but of the same kind (determinants of the same determinable); they are not as different from one another as they are different from blue; they cannot simultaneously be instantiated at the same place; they are properties that things visually appear to have; and we know when a thing appears to have these properties. All this can be known simply by reflecting on colour experience.

If it was not the case that experiences of seeing red and orange provided all the support that is necessary for these claims, then these claims would be subject to correction under the weight of future empirical discoveries. But they aren't. Nothing could count as evidence against these claims. Would such knowledge be possible if some physicalist theory of colour was true? To answer this question Boghossian and Velleman review what types of physicalist theories of colour are possible at all.

The colour-realist physicalist says that colour experience represents physical properties of objects. Now either it is the case that, on his theory, the property being represented by colour experience is identical to a microphysical property ("identity physicalism"), or it is the case that there are multiple microphysical realizations to the same colour, and the physical property represented by colour experience is the second order property of the object of having one or another of these microphysical properties ("realization physicalism").

There is also another major division. When we are seeing something red, it may be the case

that the property red itself is not an element of the content of the mental state we are in (the experience of seeing red). Rather, the grasp of the mental state on redness is mediated by an intension, or meaning, or characterisation, or mode of presentation, like in the Fregean theory of linguistic reference. The immediate content of the mental state is this intension, and it succeeds in referring because the property red uniquely satisfies the characterisation given by it. So the representation has an intrinsic property in virtue of which it is the representation of that particular physical colour property. If a physicalist theory of colour conceives the relation of the mental representation and the physical colour property along these lines, then this theory would be called "Fregean" by Boghossian and Velleman.

The alternative is that there is no such intrinsic property of the representation in virtue of which it can only be the representation of the property red. The mental state which is the representation of the property red is capable of referring to it directly, maybe in virtue of a specific sort of causal relation or covariance between the occurrences of the property red before our eyes and the occurrences of the representation (presumably a specific neural pattern in a specific region of the brain). So the mental representation is like a sign that doesn't itself give characterisation of its referent by any means and it stands for its referent, in our case a particular physical colour property, just as a proper name stands for the person it denotes. Boghossian and Velleman would call a theory which conceives of the mental representation of colours this way "Russellian".

These two divisions are independent of each other, so they distinguish between four main different types of physicalist-realist theories. Boghossian and Velleman's strategy is to go through all these types of physicalist theories of colour and see whether the kind of knowledge based simply on colour experience described above would be possible if they were true. They claim that such knowledge can exist only on "Fregean realization-physicalist" theories. This is so, because both the "Russellian" and the "identity theoretic" physicalist theories are ruled out. On Russellian and identity theoretic types of physicalist theory, visual experience represents colour without a

characterization that denotes it necessarily. In the case of Russellian theories it follows from the definition of what makes a theory "Russellian". In the case of identity theories, the reason why they are ruled out is that it is not credible that the features of our colour concepts denote microphysical properties necessarily. So such colour-representations can denote real, physical colour properties only contingently, and therefore fail to provide the appropriate introspective knowledge of the properties denoted.

But can a "Fregean" realization-theoretic physicalist theory be true of colour? Boghossian and Velleman answer this question in the negative. They say such a theory would misrepresent the phenomenology of colour experience. Visual experience does not distinguish between the perceived colour of an object and the property of the perception itself (the way it feels) in virtue of which the perception is the perception of an object so coloured. There is no sensation involved in colour vision distinct from how objects look. But a Fregean representation-theoretic physicalist theory would appeal to such a sensation: that would be the introspectible colour property of the representation itself, in virtue of which it would necessarily be the representation of the colour property it represents.

For our concern, however, it does not really matter whether what Boghossian and Velleman claim about "Fregean representation-theoretic" physicalist accounts of colour is right or not. If physicalist realists can come up only with "Fregean representation-theoretic theories", then they cannot answer the knowledge argument by taking the intentionalist-eliminativist way, or more precisely, the way we have earlier called 'direct realist intentionalism'. For if they choose a theory of this type, then they are back with the kind of qualia they wanted to eliminate – the intrinsic qualitative character of the experience itself, which cannot be placed out in the physical world that is being represented by experience.

So, to see whether the physicalist answer to the knowledge argument that invokes direct realist intentionalism about the qualitative character of experience is bound to fail in the case of colours, we only have to examine whether Boghossian and Velleman are right in claiming that

"Russellian" and "identity-theoretic" physicalist theories are ruled out by the special epistemic features of colour experience.

Considering a Possible Russellian Theory: Direct Realist Intentionalism

In Boghossian and Velleman's view the key difference between "Fregean realization physicalism" about colour, on the one hand, and the rest, i.e. Russellian theories or identity theories, on the other, is that on the latter theories

> visual experiences like yours represent colours only as a matter of contingent fact. Under the terms of these theories, an experience internally indistinguishable from your experience of seeing something as red might fail to represent its object as having that colour. The reason is that red is represented by your experience, according to these theories, only by virtue of facts incidental to the internal features of the experience. (p. 87)

On Russellian theories these facts "incidental to the internal features of experience" are of course the causal or correlational facts that make the mental (neural) sign capable of referring directly to the property red. On Fregean identity theories the question is whether the facts in virtue of which the property red uniquely satisfies the characterisation which the qualitative character of a red experience gives to it are incidental to the internal features of the experience. It seems that they are, given that an identity theorist believes that to be red is to have some microphysical surface property, so the facts in virtue of which the property red uniquely satisfies the characterisation given of it in the red experience are microphysical facts, yet these facts are not contained in the experience, a red experience does not reveal what redness consists in. Boghossian and Velleman conclude that from this we are entitled to draw the conclusion that the characterisation visual experience gives of redness is a contingent one. It does not represent what

redness is, just tracks redness in virtue of some causal relatedness, so it is related to redness pretty much as a 'Russellian' mental symbol would be related to its referent. So in this respect Russellian theories and Fregean identity theories are on a par.

This, I think, is correct. However, we do not really have to consider Fregean identity theories, for the same reason we do not have to consider Fregean realisation theories. For both types of theories being 'Fregean' may consist only in the feature that they hold that visual experience gives a characterisation of redness – regardless of whether it is identical with, or just realised by, a physical surface property – in the qualitative character of the experience (phenomenal redness). Nothing else can play this role. So on either subtypes of Fregean theories we are back with a quale, intrinsic to the experience, which the kind of intentionalism we are considering would want to place outside, in the object being represented.

So ultimately, the question whether the Boghossian-Velleman argument rules out the kind of intentionalism we are considering as a would-be answer to the knowledge argument boils down to the question whether they are right in claiming that Russellian theories of representation would be bound to fail to accommodate the special epistemology of colours. It is worth noting, however, that the crucial feature in virtue of which Russellian theories are bound to fail at this point according to Boghossian and Velleman is not unique to Russellian theories. It is the feature that the link between the intrinsic features of the mental representation of a colour property and the property itself is just a contingent one.

It seems to me that they would certainly be right if it was the case that a Russellian theory of colour representation could be conceived only with the qualitative character of colour experience being part of the intrinsic nature of the mental representation of colour. The knowledge about colours which is not subject to empirical correction, and which we possess entirely in virtue of being the subjects of colour experience, comes from our awareness of the qualitative character of colour experience. The Russellian character of the theory of representation, i.e. that the intrinsic features of the representation are linked to the real physical

nature of the colour property they represent only contingently, is relevant for the possibility of the kind of incorrigible knowledge about colours Boghossian and Velleman are considering only if the source of this knowledge, i.e. the qualitative character of colour experience is part of the intrinsic features of the representation. Otherwise the fact that the representation is Russellian has no bearing whatever on the possibility of such knowledge. But this is exactly what direct realist intentionalists deny. A direct realist intentionalist theory is the sort of Russellian theory according to which the qualitative character of the experience is *not* part of the intrinsic character of the mental representation. The mental representation itself is just a Russellian sign, it is capable of referring to its referent only in virtue of a contingent relation to it, e.g. a causal relation, the qualitative character of the experience is not 'generated' in perception, but is 'selected' from the real features of the object being perceived. So on the kind of Russellian theory we are considering, the qualitative nature of a red visual experience is not part of what is only contingently linked to the real physical nature of redness.

So it is not true, what Boghossian and Velleman say, that in consequence of the Russellian character of the representation, a representation with the same qualitative character could represent different physical surface properties, say, in a Twin Earth scenario. A mental representation which is intrinsically the same, that is, the same neural pattern in the same part of the brain, could stand for a different physical colour property on Twin Earth, but then its qualitative feel would be different, too. A proponent of the kind of intentionalist theory we are considering is not bound to accept that it is possible for the colour experience and the physical colour property being experienced to come apart (like in a Twin Earth or a qualia inversion thought experiment). On his theory, a difference in the physical colour properties being seen always result in a difference in the phenomenal quality of the experience, since the latter transparently reports of the former. Conversely, if the physical surface properties of the object perceived are held fixed, and so are the lighting conditions and the type of the perceptual apparatus of the perceiving subject, then the phenomenal concept that will represent the physical

surface property will be the same, too.

So the upshot is that the argument that Boghossian and Velleman propose for the irreconcilability of physicalism about colour and the epistemic status of the knowledge we have of colours and their relations merely in virtue of being the subjects of colour experience fails in the case of at least one type of intentionalism about colour experience, and this is exactly the kind of intentionalist theory we have earlier found to probably be fit to ground a physicalist answer to the knowledge argument.

The Problem of Distinguishing between Veridical and Non-veridical Colour Perception

But can such a theory save physicalism about colour? The key feature of such a theory, the one we have earlier called 'direct realist intentionalism' is that, on this theory, in the case of veridical perception the experienced phenomenal colour is an objective physical feature of the object being perceived – in the case of veridical perception the qualitative character of the experience is not 'generated' as a subjective feature of the perceptual state, but is 'selected' from the objective features of the perceived object. But how does this theory account for the difference between veridical perceptions and misperceptions?

Suppose that a Jonathan apple in bright light at noon and an Othello grape in the light of the setting sun seem to have exactly the same colour (red) to a particular person. His colour experiences in the two cases are introspectively indistinguishable. But, according to the direct realist intentionalist theory, it is possible only if one is a veridical representation and the other is a misperception. According to strong pure intentionalism phenomenal concepts represent the properties of intentional objects. According to the direct realist version of this theory, in the case of veridical perceptions, phenomenal concepts transparently represent some of the physical properties of the physical object being perceived (which, in this case, is the intentional object). We have seen earlier that the qualitative character of misperceptions (non-transparent representations) is explained by appealing to the original transparency claim. But how to think of

these cases of transparency, how should the veridical cases of representation be distinguished from the non-veridical cases, and what support could be given of this distinction?

In the above cases, it is plausible for the direct realist intentionalist to claim that phenomenal concept RED represents a surface physical property of the Jonathan apple veridically, and when it is tokened in response to the presence of the Othello grape, it is a case of misperception due to nonstandard lighting. So the veridical cases of the representation of physical colour properties by phenomenal concepts are identified with reference to a standard of lighting. It is somewhat plausible that we have more reason to consider standard the lighting at noon (on a clear day) than the lighting at sundown. But lighting at noon is very different in winter from what it is like in the summer. In the winter the spectral composition of sunlight is shifted a bit toward red, quite like on summer afternoons or evenings. So the surface properties which – in the sunlight at noon, on a clear day – induce the tokenings of the exactly same phenomenal colour concept will be somewhat different in winter from those that do it in the summer. Is it plausible to regard perhaps the lighting at noon on Midsummer day standard, provided that it is clear day? Well, it seems arbitrary. But even if we accept this definition of standard lighting, is it supposed to apply in both Nairobi and Stockholm? The sunlight at noon, Midsummer day in Stockholm is shifted toward the red end of the spectrum relative to the sunlight at noon, Midsummer day in Nairobi. But only one of them can be standard. Which one, and why? Suppose the standard lighting is the lighting in Nairobi, at noon, Midsummer day, if the sky is cloudless. From this it would follow that the members of the Royal Academy in Stockholm, although they are presumably all very smart and wise, will never have a single veridical colour perception in their entire lives (unless they go to Nairobi), so maybe they will never entertain a true thought about the colours of objects.

If we want to model how phenomenal concepts are related to surface reflectance properties, we would get the following picture. A phenomenal concept (PC) is related to a set of ordered pairs of surface physical properties (SP) and lighting conditions (LC): PC1 ↔ {(SP1,

LC1), (SP2, LC2), …, (SPn, LCn)}. This is uncontroversial. What is controversial, however, is whether we have principled reason to pick out one element of this set, say (SPi, LCi), and claim that when PC1 is induced by the occurrence of this, then this is a case of veridical perception, when the qualitative character of our visual experience captured by PC1 grasps the objective feature SPi of the object in front of us, whereas all the other cases are misrepresentations when we are using PC1, which is designed to pick out SPi, wrongly, due to the nonstandard circumstances.

A physicalist colour theorist may at this point choose to identify physical colour represented by PC1 with the whole set, or more appropriately with the disjunctive property {(SP1, LC1), or (SP2, LC2), or …, or (SPn, LCn)}. But this move is not available for the direct realist intentionalist whose theory we are now considering. For he has promised that he would explain the qualitative character of visual experience captured by the phenomenal colour concept as *an objective physical property of a physical object* (in the case when the perception is veridical, i.e. when the intentional object of the perceptual state is the real physical object, and it is represented by the content of the perceptual state as having the property it really has). But {(SP1, LC1), or (SP2, LC2), or …, or (SPn, LCn)} is not a property any object can have, for the reason that none of the disjuncts (SPi, LCi) is a property that can be possessed by an object. It is a property that only a pair composed of an object and a kind of lighting can have.

To make the situation even worse for the direct realist intentionalist, we may take into account the fact that the phenomenal concept that is being tokened in response to the occurrence of a surface reflectance property is a function of not only the lighting conditions, but also the kind of sensory-neural colour perceiving equipment the perceiving subject has. It would be very hard to deny that differences in the equipment exist. Presumably the colour perceiving equipments of octopuses, eagles, dolphins, humans and Martian scouts are quite different. If the direct realist intentionalist is to perform a standardisation analogous to the one of setting the default lighting conditions, then he must hold that one of the equipments is the standard

perceiving equipment. Now, if he claims that the standard equipment is that of humans, the arbitrariness of his theory becomes clear. Perhaps he would want to claim that the different equipments are just different realizations of the same function from ordered pairs of surface physical properties and lighting conditions to phenomenal concepts, i.e. that no matter if one is an eagle or a human, one's visual sensory apparatus would give rise to the same phenomenal colour concepts in response to the same surface physical properties in the same lighting conditions. But this supposition runs counter to empirical investigations that seem to show that some animal species visually discriminate between types of physical reflectance properties of surfaces which we represent by the same phenomenal colour concepts, which perhaps indicates that they have phenomenal colour concepts that we don't have. Colour blindness also can be regarded as a proof of the possibility of a different colour perceiving equipment. Maybe it can be set aside as a deficiency, which perhaps entails that this case is legitimately considered non-standard. But even then, many people, including me, report that they see slightly different shades with their two eyes. The world seen through my left eye is a little bit reddish relative to the somewhat greenish world seen through my right eye. Is that a manifest case of being non-standard, too?

To support the claim that there is a standard, but maybe multiply realisable, colour perceiving equipment, which can be considered standard because in standard lighting circumstances it reveals the true colours to the subjects that are equipped with them, the selective intentionalist would perhaps propose an evolutionary explanation. Maybe evolution can also compensate for systematically non-standard lighting conditions, so the Swedes can see the same colours as the Kenyans after all, at least at Midsummer. (Maybe the differences in misperceptions due to non-standard lighting are compensated, too.) But evolution has no interest, so to speak, to provide for such compensations. All the survival value of colour perception seems to be exhausted by discriminatory competence, seeing the right 'absolute' hues adds nothing to it.

If relativity to perceiving equipment is acknowledged, then physical colour may perhaps be

understood as sets of ordered triads of surface physical properties (SP) and lighting conditions (LC), and visual equipments (VE). Again this move is available to the physicalist in general, but not to the direct realist intentionalist whose position we are considering.

In his 2009 book, *The Red and the Real*, Jonathan Cohen gives an essentially similar argument for his relational theory of colour. Cohen is concerned with the different perceptions of the same physical surface properties (rather than the different surface properties represented by the same phenomenal concept in different circumstances), and his aim is to establish the view that if colours are to be given a physical meaning then they should be conceived as relations that hold between surface properties, lighting conditions and perceivers (rather than to establish the falsity of direct realist intentionalism about colour qualia). The key observation, however, on which both his argument and the argument given above are based, is the same. Both my opponents and his, i.e. the direct realist intentionalists and those physicalists who insist that colour is a monadic property of object surfaces, must draw on a distinction between veridical cases of perception and non-veridical ones to account for what Cohen calls 'perceptual variations', and there seems to be no ground on which to base such a distinction.

There is a standard objection to this point. The objectors concede that we cannot know which of the perceptions is veridical, yet to suppose that this lack of knowledge grounds the claim that there are *no* veridical perceptions *at all* is to think too much of the human intellect. Such gaps of knowledge disappeared before as more facts came to light, why suppose that now the situation would be any different? As Tye writes

> We do not suppose that objects do not have precise lengths because of the limitations of our measuring equipment. Why suppose that the situation is fundamentally any different for the case of colour? (2006, pp. 177-178)

Alex Byrne and David Hilbert illustrates the point with the following case:

> Imagine, as an analogy, a population of intelligent, reasonably accurate thermometers. [...] Like all measuring instruments, the thermometers are calibrated slightly differently. They all agree that the temperature right now is pretty high, around 70°F or so. But some think the temperature is 69°F, while others think it is 70°F, and yet others think it's 71°F. Some of them conjecture that being 70°F is a physical property of some kind, perhaps related to mean molecular kinetic energy. But the thermometers have no theory of intentionality that would enable them to establish conclusively that they are representing physical properties of this sort. And, since they do not have other ways of measuring temperature, they have no "independent method" of determining whether the temperature right now is exactly 70°F, or even whether it is pretty high. Still, some of these thermometers are perceiving the temperature correctly and others are not. Further, this lack of an independent method need not stop them from forming justified beliefs about the temperature. Perhaps none of them can justifiably believe that the temperature is exactly 70°F, but presumably they might justifiably believe that it is on the high side, or approximately 70°F. (2004, pp. 42-43)

Cohen responds to this objection by pointing out that the thermometer case is different from the colour case, because it is uncontroversial that temperature has a natural essence – mean kinetic energy, whereas it is controversial whether colour has (cf. 2009, p. 50). I think this move is not necessary, as it turns out if we take a closer look at the analogy, i.e. explore which elements of the thermometer story are analogous to which element of the one about colour perception.

In the above example, the existence of Temperature (capital T) understood as the mean kinetic energy of molecules is supposed to make it evident that there is such an objective

property as temperature, even though the different thermometers in the same room read different *temperatures* (italicised). By analogy, even though different perceivers perceive different *colours* (subjective phenomenal colours) looking at the same surface, there is nothing strange in supposing that there is such a thing as objective Colour, which they track slightly differently.

However, this analogy doesn't show anything to which a colour subjectivist couldn't agree. For what Temperature (mean kinetic energy of molecules) is analogous to is *a physical reflectance property of object surfaces*. Nobody doubts that there *is* such an objective property, and that (other things held constant) the subjective phenomenal character of colour experience tracks that property. It is not a matter of dispute between the colour realist and the colour subjectivist. Since the objective existence of mean kinetic energy (Temperature) is analogous to an element of the colour dispute which is undisputed, this analogy will hardly decide the dispute. What is disputed is whether one of the somewhat different phenomenal colour experiences that are induced in perceiving subjects by the presence of a surface reflectance property can be said to represent that surface property *veridically*, whereas the others cannot.

Now, what element of the thermometer case could be analogous to this? I think the relevant analogy would be if it was the case that the fact that a mercury column stretches exactly until the mark 70 carved next to it represents veridically the mean kinetic energy of the molecules in the surrounding air, but when the mark 69 or 71 are next to the top end of the mercury column, these are cases of misrepresentations.

Mercury column heights *covary* with mean kinetic energy, that much is straightforward. So the misrepresentation is *not* in the mechanism itself, it works properly, still sometimes it produces veridical representations, at other times misrepresentations, as it was admitted by Byrne and Hilbert[3]. But how could that be the case? The explanation is that the scaling is carved differently on the thermometers. But the problem is that there is no objective fact which would prescribe it

[3] "Visual mechanisms (for example, the mechanism of simultaneous contrast) are neither illusory nor veridical. Rather, it is the output of visual mechanisms-visual experiences-that are illusory or veridical. The same mechanism may produce illusory output on one occasion, and veridical output on another." (Byrne and Hilbert 2004, p. 41)

that the variation of mercury column height should be contrasted to *one specific scaling*, rather than any other. To calibrate a thermometer, one looks up in a physics textbook how one should do that. But the instructions one would find in the textbook would reflect an arbitrary convention on which physicists agreed previously. Byrne and Hilbert emphasise that it is highly important not to mix temperature with the conditions for the detecting of temperature (cf. 2003, p. 6). But if we look into physics textbooks for instructions for distinguishing between thermometers that represent temperature veridically and those that don't, we will only find references to conditions to measure temperature. The final line is that this distinction is just a matter of *convention* and there is no objective fact of the matter about which convention is right and which isn't.

Suppose that in this community of thermometers one particular thermometer measures the temperature correctly by *our* current standards. To say that the readings this particular thermometer produces are the veridical representation of temperature is analogous to the case of a Martian superscientist who visits earthlings, points to John Smith and declares that when Smith has a *blue* (phenomenal) experience (of a certain hue, in Nairobi, at noon, on a cloudless Midsummer day) it is caused by the physical surface property to be called True Blue, for he has a Martian physics book with him that contains the definition of colours. The analogy holds, because any scale of temperature invented by physicists would be arbitrary, and so is inventing "scales" for colours.

To make the analogy complete one should imagine that these states of reading 70, or 69, or 71, or any degrees Fahrenheit are perceptual states for the intelligent thermometers with a qualitative character. There is something it is like to read 69 degrees, and it is slightly different from reading 70, and so on. Suppose that one of the intelligent thermometers comes up with the theory that reading 70 Fahrenheit is the intrinsic qualitative feature of a perceptual state (and maybe suggests that it will cause problems for some reductive metaphysical theories). Can he be talked out of this view by pointing out that there are environments whose temperature is really 70 Fahrenheit, and saying that his perceptual state has no intrinsic qualitative character, it is just

that a perceptual state is an intentional state which introduces the objective property of being 70 Fahrenheit hot to the mind?

Not as long as the qualia theorist thermometer can draw a distinction between what it is for himself and fellow thermometers to read 70 Fahrenheit (phenomenal temperature) and a certain amount of mean kinetic energy of the molecules of the environment (physical temperature), which an alien scientist (i.e. a human physicist) arbitrarily decided to call 70 Fahrenheit. The qualia theorist thermometer can claim that it is a categorical distinction: mean kinetic energy is a primary property of the environment, but that the environment feels 70 Fahrenheit for some thermometers is a secondary property which would cease to exist in the moment when thermometers would cease to exist. The qualia theorist thermometer would say that his theory is that the instantiation of the primary property in the environment is one of the causal factors that contribute to the obtaining of the secondary quality in a thermometer mind. The rival theory would be the intentionalist's, which is that there is no secondary quality really, it just the primary quality represented – veridically in some cases and misrepresented in others. The qualia theorist thermometer could say that we are facing a theory choice here, and the fact that there is a variety of different primary properties which may cause the same secondary property, and the distinction that the intentionalist advocates between veridical and non-veridical cases of representation is an arbitrary one, clearly favours his theory. The mere fact that there is such a thing as objective Temperature, i.e. mean kinetic energy, does not speak to his point at all. The link established by certain amounts of mean kinetic energy and certain phenomenal *temperatures* (i.e. thermometer readings) that is supposed to single out the veridical cases of the former being represented by the latter, will be arbitrary anyway.

To sum up, the direct realist intentionalist (or the monadic physicalist) has to distinguish between cases of colour perception, veridical or transparent vs. non-veridical or non-transparent, which are realised by exactly alike physical processes. In each case there is a surface property, there is a kind of lighting, the light is reflected from the surface, and it stimulates the visual

perceiving equipment, and this process gives rise to the tokening of a phenomenal colour concept. There is nothing intrinsic in this physical story that would distinguish between the transparent and the non-transparent cases. Our theorist must rely on an arbitrary standardisation to account for that distinction, but this cannot be given a non-arbitrary definition cashed out in physical terms.

Earlier we found that Boghossian and Velleman's argument against colour-physicalism did not cover a case of theory of the type they call Russellian. This is the case when the phenomenal character of experience, our awareness of which is the source of our special knowledge of colours and relations between them, is not included among the intrinsic features of the representation, but placed 'outside', so to speak, among the properties of the objects being perceived. After the above critique of this proposal, however, we may conclude, that this case, uncovered by the Boghossian-Velleman argument is a highly implausible one, which must rely on a distinction between transparent and non-transparent representations, which cannot be given any physical account.

Conclusion

My aim in this thesis was to show that the knowledge argument can be defended against the physicalist objection which we have good reasons to regard the most attractive, direct realist intentionalism, which essentially identifies qualia with the physical properties of the perceived object. To motivate the discussion of this kind of reply to the knowledge argument, I very briefly reviewed the main other types of answer, and drawing on the arguments of Robinson (1993, 1996), Crane (2003) and Gertler (1999) I gave my reasons to consider them less hopeful than the intentionalist answer.

The intentionalist answer trades the 'hard' problem of qualia for the problem of intentionality which many consider 'easy'. I am not sure if the physicalist is really better off this way, but I put this worry aside for the purposes of this thesis. Intentionalism comes in different versions. Alter (2007) proposed an argument to the effect that an intentionalist version of the knowledge argument can be construed, and that, after all, intentionalism cannot overcome the conclusion of the knowledge argument. I argued that although Alter's argument works against 'impure' intentionalism, and also against weaker versions of pure intentionalism, the direct realist version of strong pure intentionalism survives Alter's critique, and is still a threat to the knowledge argument.

From this point on I was specifically concerned with colour as an example of qualia. I made this choice, because although colour qualia seem prima facie intentional, there is an extensive debate in the literature about whether colours can be real physical properties of object surfaces. An important part of the debate focuses on the knowledge we have about colour simply in virtue of being the subjects of colour experience. Boghossian and Velleman argued that no kind of physicalist theory of colour can account for this knowledge unless it misrepresents the phenomenology of colour. I found, however, that their argument does not cover the direct realist intentionalist theory which places the qualitative character of colour experience not among the

intrinsic features of the representation, but conceives of them as features of the object being represented captured by the representation.

So I offered an argument to the effect that the direct realism involved in such an intentionalist theory is highly implausible, given that it must draw on a distinction between veridical colour perceptions and misperceptions which is arbitrary and cannot be explained with the resources of a physicalist-causal theory of colour representation. My argument against direct realist intentionalism about colour is similar to Cohen's argument (2009) against monadic (non-relationalist) physicalism about colour. I argued that the direct realist intentionalist has no physical means to make it plausible that one of the surface properties that can cause a visual experience with a particular kind of phenomenal character is identical to the phenomenal character. The physical mechanism of visual representation is in no way different in the allegedly veridical and non-veridical perceptions. No non-arbitrary definition of veridical representations seems to be plausible, and so it turns out that the direct realist intentionalist theory of the qualitative nature of colour experience is based on a notion of 'transparency' which cannot be plausibly maintained.

Given this, I think the best prospect for direct realist intentionalism would be to claim that there might be some differences, in the end, to be discovered later in the causal process that brings about veridical and non-veridical perception. But then, given that only *one* kind of physical surface reflectance property can be identical to a phenomenal colour property, it would still have the result, that the vast majority of colour perception is misperception. The original idea of introducing intentionalism as a theory of visual perception was to explain the essentially representational nature of it. The direct realist intentionalist, who has to claim in the end that veridical perceptions are very rare, must regard this general motivation for the thesis to be misguided.

Coming back to the knowledge argument, I conclude that answering the knowledge argument by taking an intentionalist stand about qualia, which would otherwise be an attractive

alternative to the other kinds of answer to the knowledge argument, requires a kind of intentionalism which, at least in the case of colour qualia, must rely on a distinction between veridical and non-veridical cases of colour perception, which is highly implausible.

References

Alter, Torin (2007), "Does Representationalism Undermine the Knowledge Argument?", in Torin Alter and Sven Walter (eds.) (2007), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, Oxford Scholarship Online.

Aydede, Murat (2001), "Naturalism, Introspection, and Direct Realism About Pain", in *Consciousness and Emotion*, 2, pp. 29-73.

Byrne, Alex (2001), "Intentionalism Defended", in *Philosophical Review*, 110, pp. 199-240.

Byrne, Alex and David Hilbert (2003), "Color Realism and Color Science", in *Behavioral and Brain Sciences*, 26, pp. 3-64.

------ (2004), "Hardin, Tye and Color Physicalism", in *Journal of Philosophy*, 101, 1, pp. 37-43.

Boghossian, Paul and David Velleman (1991), "Physicalist Theories of Color", in *Philosophical Review*, 100, pp. 67-106.

Brown, Derek H. (2010), „Locating Projectivism in Intentionalism Debates", in *Philosophical Studies* 148, pp. 69-78.

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

------ (2004), "The Representational Character of Experience", in Brian Leiter (ed.), *The Future for Philosophy*, Oxford University Press.

------ (2007), "The Hard Problem of Consciousness", in S. Schneider, M. Velmans (eds.) *The Blackwell Companion to Consciousness*, Oxford: Blackwell.

Cohen, Jonathan (2009), *The Red and the Real*, Oxford: Oxford University Press.

Conee, Earl (1994), "Phenomenal Knowledge", in *Australasian Journal of Philosophy*, 72, pp. 136-50.

Crane, Tim (1998), "Intentionality as the mark of the mental", in Anthony O'Hear (ed.) *Contemporary Issues in the Philosophy of Mind*, Cambridge: Cambridge University Press, pp. 229-251.

------ (2003), "Subjective Facts", in H. Lillehammer and G. Rodriguez-Pereyra (eds.) *Real Metaphysics*, London: Routledge, pp. 68-83.

------ (2007), "Intentionalism", in Ansgar Beckermann and Brian P. McLaughlin (eds.) *Oxford Handbook to the Philosophy of Mind*, Oxford: Oxford University Press.

Dennett, Daniel C. (1991), *Consciousness Explained*, Boston: Little Brown and Company.

Fodor, Jerry A. (1987), *Psychosemantics*, Cambridge MA: MIT Press/A Bradford Book.

Galilei, Galileo (1957), *Discoveries and opinions of Galileo*, ed. and transl. by Drake Stillman, New York: Doubleday.

Gerler, Brie (1999), "A Defense of the Knowledge Argument", in *Philosophical Studies*, 93, pp. 317-336.

Flanagan, Owen J. (1992), *Consciousness Reconsidered*, Cambridge: MIT Press.

Fodor, Jerry A. (1987), *Psychosemantics: the problem of meaning in the philosophy of mind*, Cambridge: MIT Press.

Harman, Gilbert (1990), "The Intrinsic Qualities of Experience", in *Philosophical Perspectives*, Vol. 4, Action Theory and Philosophy of Mind, pp. 31-52.

Horgan, Terrence (1984), "Jackson on Physical Information and Qualia", in *Philosophical Quarterly* 32, pp. 127–136.

Jackson, Frank (1982), "Epiphenomenal Qualia," in *Philosophical Quarterly*, 32, pp. 127-136.

------ (1986), "What Mary Didn't Know", *in The Journal of Philosophy*, Vol. 83, No. 5, pp. 291-295

------ (1995), "Postscript", in Paul K. Moser and J. D. Trout (eds.) (1995) *Contemporary Materialism*, London: Routledge, pp. 184–189.

------ (1998), "Postscript on Qualia", in Frank Jackson (1998) *Mind, Method, and Conditionals: Selected Essays*, London: Routledge, pp. 76-79.

------ (2003), "Mind and Illusion", in Anthony O'Hear (ed.), *Minds and Persons*. Cambridge: Cambridge University Press.

Levine, Joseph (1983), "Materialism and Qualia: The Explanatory Gap", in *Pacific Philosophical*

Quarterly, 64, pp. 354-361.

Lewis, David (1990), "What Experience Teaches", in W. Lycan (ed.) *Mind and Cognition: A Reader*, Oxford: Blackwell.

Maund, Barry (2008), "Color", in Edward N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*, URL = <http://plato.stanford.edu/archives/fall2008/entries/color/> (Last accessed on 31 August 2011.)

Nagel, Thomas (1974), "What is it like to be a bat?", in *Philosophical Review*, 83, pp. 435-456.

Nemirow, Lawrence (1990), "Physicalism and the Cognitive Role of Acquaintance", in William G. Lycan (ed.) *Mind and Cognition: A Reader*, Oxford: Blackwell.

Putnam, Hilary (1981), *Reason, Truth and History*, Cambridge: Cambridge University Press.

Robinson, Howard (1993), "Dennett on the Knowledge Argument", in *Analysis*, Vol. 53, No. 3, pp. 174-177.

------ (1994), *Perception*, New York: Routledge.

------ (1996), "The Anti-Materialist Strategy", in Howard Robinson (ed.) *Objections to Physicalism*, Oxford: Oxford University Press.

Tye, Michael (1992), "Visual qualia and visual content", in Tim Crane (ed.) *The Contents of Experience*, Cambridge: Cambridge University Press.

------ (1995), *Ten Problems of Consciousness*, Cambridge, Mass.: The MIT Press.

------ (2006), "The Puzzle of True Blue", in *Analysis* Vol. 66, 291, pp. 173–178.

------ (2009), "Qualia" in Edward N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*, URL = <http://plato.stanford.edu/archives/sum2009/entries/qualia/>. (Last accessed on 31 August 2011.)