

# **Zombie-Mary and the Blue Banana**

## **On the Compatibility of the 'Knowledge Argument' with the Argument from Modality**

Tillmann Vierkant  
Max Planck Institute for Psychological Research  
Amalienstr.33  
80799 Munich  
GERMANY

vierkant@psy.mpg.de

Copyright (c) Tillmann Vierkant 2002

PSYCHE, 8(19), October 2002  
<http://psyche.cs.monash.edu.au/v8/psyche-8-19-vierkant.html>

**KEYWORDS:** Epistemic Argument, Zombies, Epiphenomenalism, Physicalism, Common Sense Realism

**ABSTRACT:** This paper is trying to show that it is not possible to use the Knowledge argument as independent evidence for the form of non-reductionism the Modal argument argues for. To show this, Jackson's famous 'Mary' thought experiment is imagined in a zombie world. This leads to the result that there are many problems in the Mary experiment, which cannot have anything to do with phenomenal Qualia, because the Zombie-Mary would encounter them as well, and once all these problems are accounted for, it is no longer clear whether a Zombie-Mary is conceivable at all. Finally, an alternative explanation for the strong non-reductive intuitions of the Mary experiment is discussed.

## **1. The Epistemic and the Modal Argument**

Frank Jackson developed one of the most famous thought experiments in the philosophy of mind to prove that *physicalism* must be false. In this paper I want to argue that this thought experiment confuses two very different ideas about non-reductionism. The first idea is *epiphenomenalism*, i.e. the idea that there are irreducible phenomenal facts which nevertheless do not influence our behavior, while the second is *common sense realism*, which I take to be the idea that an irreducible phenomenal consciousness (i.e. a consciousness that includes phenomenal facts which are neither type nor token identical with physical facts) is causally relevant for at least parts of our behavior. This second

alternative must, of necessity, doubt in some way one of the central claims of physicalism, i.e. that there is in our world a physical cause for every physical effect. If this suspicion is correct, then the thought experiment featuring the omniscient but blind-blind scientist Mary invokes intuitions which have nothing to do with the first idea of non-reductionism. Understanding the experiment in this way changes it into what Daniel Dennett has labeled an intuition pump. It would be rendered no more than one of these "fiendishly clever devices" which "deserve their fame, if only for their seductiveness" (Dennett 1991, p.282). This is to say that Mary prompts intuitions of common sense realism which obviously cannot be reconciled with the theoretical position of epiphenomenalism. And this in the light of the fact that the Mary experiment is used by many, including its inventor, to support precisely this latter position.

But first of all, let us examine Jackson's argument. Imagine this intuitively very convincing thought experiment:

"Mary is confined to a black-and-white room, is educated through black-and-white books and through lectures relayed on black-and white television. In this way she learns everything there is to know about the physical nature of the world. She knows all the physical facts about us and our environment, in a wide sense of 'physical' which includes 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. If physicalism is true, she knows all there is to know. For to suppose otherwise is to suppose that there is more to know than every physical fact, and that is what physicalism denies.

Physicalism is not the noncontroversial thesis that the actual world is largely physical, but the challenging thesis that it is entirely physical. This is why physicalists must hold that complete physical knowledge is complete knowledge simpliciter. For suppose it is not complete: then our world must differ from a world,  $W(P)$ , for which it is complete, and the difference must be in non-physical facts: for our world and  $W(P)$  agree in all matters physical. Hence, physicalism would be false at our world (though contingently so, for it would be true at  $W(P)$ ).

It seems, however, that Mary does not know all there is to know. For when she is let out of the black-and-white room or given a color television, she will learn what it is like to see something red, say. This is rightly described as learning-she will not say, "ho,hum". Hence physicalism is false." (Jackson 1986, p.291).

It is almost impossible not to feel the intuitive force of this experiment. Intuitively, all the physical knowledge in the world cannot give Mary the slightest clue about the experiential feel of seeing colors. Jackson concluded from his experiment that this feel can therefore play no role in the functional description of our world, i.e. that this feel must be epiphenomenal.

The thought experiment evoked a host of other philosophical interpretations as well, many agreeing, but even more criticizing, the conclusion of Jackson's argument. These critical voices were very heterogeneous in themselves. They ranged from a complete denial of the conclusion of the experiment (Dennett, 1991), through the claim that Mary gained a new ability but no knowledge (Nemirow, 1990), to various forms of replies that found in the Mary experiment evidence for the intricacies of the term knowledge, knowing how and knowing that, indexical arguments, opacity of knowledge, e.g. (Horgan, 1984; Perry, 1979).

There is much to be learned from this debate, but this paper focuses exclusively on the question of whether the experiment can provide independent evidence for the ontology advocated by the epiphenomenalist [1](#), as it was intended to by Jackson and as many philosophers have thought after him. This essay is trying to provide a tool that allows philosophers and non-philosophers alike not to get misled by the thought experiment as easily as in the past. It wishes to argue that the Mary experiment does not provide support for the theoretical position it was originally designed to defend and which David Chalmers [2](#) has made such a very strong and famous case for, i.e. epiphenomenalism. The Mary experiment might be compatible with this position, but cannot provide any evidence for it.

These points can be made by demonstrating that the intuitions Mary evokes relate problematically to another classical piece of ammunition within the epiphenomenalist artillery. In fact, Jackson's description of the purely physical world W(P) is strongly reminiscent of another very interesting philosophical creature.

Probably the only other being that has become as famous as the conclusive proof against reductionism is the *zombie* [3](#). A philosophical zombie is, as no philosopher ever forgets to mention, quite different from a voodoo zombie. Where the latter is essentially something with a frighteningly different appearance and terrifying behavior in comparison to a normal human being, the philosophical zombie is exactly like a human being, with one small but decisive exception. It does not have phenomenal consciousness. In some versions, philosophical zombies are only behaviorally indistinguishable from their conscious fellow beings whereas in other, more extreme cases, zombies are supposed to be exact physical replicas of their conscious twins, which therefore must be behaviorally isomorphic to their conscious counterparts as well. Many people find these creatures difficult to imagine, so it might be useful to emphasize once more: What sets them apart from human beings is nothing else but a complete lack of phenomenal consciousness. Zombies do not have *Qualia*. Qualia are conscious qualities [4](#) - this means that they are the experiential properties of our feelings, thoughts, impressions, etc. They are 'what it is like' to have all these things. The zombie argument is supposed to show that, as it is conceivable to imagine a world that is just like ours apart from the non-existence of Qualia, it cannot be possible that Qualia can be reduced to something else (typically a purely functional description) in ours.

This is not to say that zombies do not talk about Qualia! As they are behaviorally equivalent to us, they can be imagined to talk about consciousness just as much as we do.

Only in their case, the only reason for their doing so is the complex architecture in their brain. But all the neural firing that is going on in their heads does not give them the phenomenal side of consciousness.

## 2. Zombie-Mary

Now armed with this understanding of philosophical zombies, as well as intuitions from Jackson's thought experiment which show us why Qualia might be a further fact that Mary comes to know about when she leaves her black and white room, we can turn now to *Zombie-Mary* [5](#). *Zombie-Mary* is the perfect neuroscientist living in her black and white world. She acts exactly like 'normal' Mary, but does not really have black and white, or indeed any, experiences. In other words, *Zombie-Mary* inhabits one of the possible worlds where there are human beings without consciousness who behave just like we do. She lives in zombie world.

Whether or not she has got a conscious human Mary twin in our world, whose physical replica she is, does not really matter. But it is interesting nevertheless to note that for Jackson there is no significant difference between silicon and neurons. [6](#) The physical make-up of *Zombie-Mary* simply does not matter for Jackson. For him, the possibility of conscious experience is not confined to humans. It is imaginable that computers will be conscious one day. If that should happen, then it would easily be imaginable that these computers have unconscious silicon twins on W(P). Nevertheless, the force of the example is better felt if *Zombie-Mary* does indeed have a conscious Mary twin in our world.

### 2.1 *Zombie-Mary's* First Color 'Experience'

Imagine, therefore, a physical duplicate of Jackson's Mary on W(P). In so doing, it should be possible to find out whether the intuitions that normal Mary triggered had to do with the fact that she discovered something about phenomenal consciousness, or with something else. If it was because of phenomenal consciousness, then imagining *Zombie-Mary* should not trigger any of the intuitions triggered by the original experiment, because in *Zombie-Mary's* world no such thing as phenomenal consciousness exists. If these intuitions are triggered nevertheless, then this seems to show that phenomenal consciousness was not responsible for them in the first place.

So, let's first of all imagine, purely intuitively, a day in the life of *Zombie-Mary*:

*Zombie-Mary* quietly works in her black and white laboratory until one day, a fellow zombie puts a red apple into her black and white environment. What will happen? The only answer that is compatible with the theoretical concept of zombies means that the following scenario unfolds: Our zombie will jump up and shout: "Wow, that's it! That's

what it is like to see colors! I had all this knowledge about neurons. I knew exactly which neurons fire when light waves of a certain length are translated into neuronal information about the redness of an apple, but I never knew what it's like to see a red apple. Isn't the world beautiful? Redness is something so exquisite, I cannot begin to put it in words." She will do this because she is functionally completely identical to her conscious twin, and will therefore always do what this twin does. So her behavior is not surprising at all. The fascinating thing is that nothing actually seems to have caused it, as she doesn't have the phenomenal consciousness which is seemingly responsible for normal Mary's reaction. But on second thoughts, this is not quite true. Actually, there is a cause for her behavior. The cause is obviously the neurons involved in color discrimination in her unconscious brain, which had not fired up until now, but which go crazy as the light rays reflecting from the apple hit her retina.

Nevertheless, if this describes accurately how Zombie-Mary would react, then in my view the story of 'normal' Mary is deeply misleading as a thought experiment. The thought experiment suggests that the phenomenal knowledge Mary gains will make a difference to the way she leads her life, thinks about her life, feels about her life and so on. If it turns out that her zombie equivalent gains just the same "knowledge" by leaving the room, even though she obviously acquires no new phenomenology (i.e. that leaving the room means for her the same changes in her zombie thoughts, zombie feelings), then the experiment loses its intuitive force. The original Mary experiment seemed to show that consciousness makes a behavioral difference, but in fact it shows nothing of the sort. The Zombie-Mary thought experiment shows that the inmate of the black and white room must have been lacking something more than phenomenal consciousness; otherwise Zombie-Mary would not lack anything at all. But it seems clear that she does: Zombie-Mary would acquire some new knowledge as well, she would not say, "ho, hum" any more than normal Mary.

To many people, this simple story seems to be enough to view the Mary experiment in a different light, but Zombie-Mary can do more. Up to now, I have only shown that Zombie-Mary evokes intuitions that Mary evokes as well. In the next section, the Zombie-Mary experiment will show that the philosophical debate associated with the Mary problem nowadays does not discuss different solutions to the same problem that Jackson saw, but seeks to find solutions to a different problem altogether. Jackson's idea of epiphenomenal facts must be regarded as an additional and disconnected problem, which does not provide a further solution to the problems of the philosophical debate aroused by or hidden in the Mary experiment. In the section after this, I will ask the question of whether Zombie-Mary is a possibility at all.

## **2.2 Zombie-Mary in the Nida-Rümelin Room**

Imagine a slight variation on our first story. This time, Zombie-Mary's evil fellow zombies put a switch which shows a red and a green square into her black and white environment. She is told that moving the switch to red will blow her world to

smithereens, while moving it to green will bring eternal happiness (whatever that may be in a zombie world). Will she know what to do? In the normal Mary experiment, most people have the intuition that she would not know what to do.

Martine Nida-Rümelin (Nida-Rümelin, 1993), for example, imagines a room where all objects are brightly colored, but not in their natural color. If Mary were led into such a room, she would learn about the color phenomenon, but she would not know which color was which. If she would not know, then neither should the behaviorally equivalent Zombie-Mary. This seems very strange, though. Zombie-Mary knows all the facts there are to know about colors in her physical world, so she shouldn't have a problem with a coded-coded switch. If it is nevertheless imaginable that she has such a problem, then it can't have to do with her imperfect knowledge. If this is the case in Zombie-Mary's case, then the argument does not work for normal Mary either, because she might be lacking whatever it is that her zombie twin is lacking as well. This something might be many things, but is certainly not phenomenal facts, which are excluded by definition from the zombie world.

That most philosophers and non-philosophers nevertheless feel that the *Nida-Rümelin room* shows something interesting about the Mary experiment shows that most people do believe that whatever it is that the Mary experiment shows, it has to be something that has behavioral consequences. The epiphenomenalist on the other hand, now has two options. She can either claim that the Nida-Rümelin room would not work, i.e. that Mary would know where to move the switch, or she can claim that there are many things going on at once.

If she chooses the second option, this is good news for all the philosophers who attempted ability hypotheses or opacity of knowledge replies to the original thought experiment, because they now have the epiphenomenalist on their side. The epiphenomenalist then has to hold that there is a problem with being exposed to color rays for the first time which is absolutely independent of phenomenal consciousness, because otherwise Zombie-Mary could not have that problem. This puts the epiphenomenalist in an uncomfortable position, as there is now no need at all any more to postulate phenomenal facts to solve the problems of the experiment. As it would be clear that there are intuitions triggered by something that has nothing to do with phenomenal consciousness, the onus would be on the epiphenomenalist to show that it is not this something which gives rise to the possibly mistaken impression that the Mary experiment seems to be about phenomenal consciousness as well. The experiment is thus at most compatible with, but not evidence for, her position. This option is, therefore, extremely undesirable for the epiphenomenalist.

To avoid this conclusion, she could argue instead that the Nida-Rümelin room would not pose a problem because Zombie-Mary, as well as conscious Mary, would know fine well how to save the world. This option is discussed in the next section, where we will encounter a more fundamental problem.

### 2.3. Back to the Black and White Room

The claim that Mary could save the world in the Nida-Rümelin room has been made before, but the author of the claim seems to be the most unlikely candidate for a defendant of any form of non-reductionism. Daniel Dennett (Dennett, 1991) argues that most people (even hardened philosophers) understand Mary in such a way that she couldn't say which of her first two color experiences was the red one without any further clues. This impression, says Dennett, is wrong. Instead, he argues, Mary would be able to distinguish between the two colors just because of her physical knowledge. If the experimenters were to put a blue banana in her room, she would exclaim: "Don't trick me! I know bananas are not normally blue". It seems obvious that this ending to the experiment should mean that its original purpose had failed, but this not a necessary conclusion.

David Chalmers, for example, holds that Dennett's argument, far from endangering the epiphenomenalist position, even enforces the epiphenomenalist view by showing that the ability hypotheses (Nemirow, 1990) must be wrong (Nemirow claims that Mary acquires a new ability but no new knowledge in the original experiment) without ever touching the phenomenal side of Chalmers' argument p.145. Chalmers argues that the intuitions triggered by the Mary experiment are completely independent of her behavior in any room and arise only because there is the particular phenomenal feel that accompanies her behavior. To place this in the context of our Nida-Rümelin example: Dennett argues that the Nida-Rümelin room would not pose a problem for Mary, because she would have the relevant abilities to use the switch. But even if that should hold true (as we know by now it should, if the epiphenomenalist wants to have an interesting case), it does nothing to prove that she also has the knowledge of what it is like to have a color experience (Chalmers, 1996b).

According to Chalmers, Dennett's argument only hits the mark if one accepts the ability hypothesis about phenomenal experience in advance (i.e. if one accepts that having the ability to distinguish between colors is all there is to have), because only then would it be true that there would be no additional fact for Mary to learn once she has the ability to distinguish between colors. If that were the case, then Dennett's argument would show that the thought experiment is flawed, because it concludes that Mary does not have an ability which she really would have. But to accept the ability hypothesis as a premise is to accept the conclusion as a premise, and would therefore be question begging. It would mean that any argument against materialism in this context would have already failed before the experiment had even begun. As it is, Dennett's argument is beside the point, because Mary's having the ability to distinguish between colors does not mean that she has everything there is to have, because she still might be lacking the phenomenal side of this ability.

This is, as it stands, a valid argument against Dennett. [7](#) Nevertheless, there are still some puzzling results, even if we do not take the ability hypothesis premise for granted, as Chalmers suggests. This is illustrated quite well if we compare Mary to her zombie twin once more.

## 2.4. The False Beliefs of (Zombie) Mary

If Mary has the ability to recognize blue even before she leaves her room, this could mean three things:

1) The zombie, faced with a blue banana, could observe her own behavior and then conclude: I react in this specific way because my system has been exposed to waves which create a blue functional state. But this option has undesirable consequences. It means that Mary would use exactly the same deductive mechanisms, because otherwise she would not be equivalent to Zombie-Mary any more, which in turn means that her belief about what is like to see blue does not stem from the phenomenal experience, but from her clever deductions. This could now either be true, in which case physicalism would be true as well, or it could be false and Mary actually learned something from her phenomenal experience (and just managed to deceive herself about her surprise when she learned this new and unexpected phenomenal fact). Nevertheless, this alternative renders the original premise, that Mary knows everything physical there is to know about color perception, incorrect, because she would now mistakenly believe that she had learned something as a result of her deductions, when she really learned something for very different reasons. This means: She believes in a physical fact P that does not exist. Therefore, she does not know the physical fact that -P is the case. This possibility is therefore not a possibility after all.

2) Another option might be that Zombie-Mary acquires the ability because of her knowledge, but believes that she has it because of the phenomenal impression (she would believe that she just had a surprising phenomenal experience). This would save the original Mary experiment, because Mary could now rightly believe that she learned a new phenomenal fact, but it has one major fault for Zombie-Mary. She would then have acquired a false belief about her perceptions: This is because she lives in a world that is per definition only physical and believes in a non-physical reason for her belief, but this renders the premise that Zombie-Mary knows everything there is to know about perception wrong. If she holds a false belief, then she obviously does not know the true state of affairs about why she holds a certain belief connected to her perceptions.

3) It seems to me that there is only one possibility to avoid such a dilemma. One would have to argue that there is something wrong with the idea of a Zombie-Mary itself. This could be because it does not make sense to talk of knowledge in the zombie world or it could be because a Zombie-Mary or a Mary are not possible for contingent reasons, or because somebody who has complete physical knowledge would know whether she is a zombie or not. All these options have undesirable side effects, but they are not the topic of this essay.

The purpose of this essay is only to show that the Mary argument does not provide support for the zombie ontology. This goal seems achieved. Not only do the two



experiments not provide any support for each other, but they actually seem to be fully incompatible.

## 2.5. Primary Conclusions

Zombie-Mary destroyed bit by bit the idea that the Mary experiment really can be used as an argument for the ontology advocated by the zombie argument. First of all, she could show that the intuitions triggered by Mary are triggered by Zombie-Mary as well. This seems to indicate that these intuitions are triggered because of a feature both Marys share, and not by the feature that sets them apart. Secondly, Zombie-Mary could show that all the ingenious solutions that have been thought about for the Mary problem would apply to her as well and therefore, all these problems cannot be the problems that are so special in phenomenal consciousness. Finally, it turned out that there are some real problems in the very idea of Zombie-Mary, because it seems that either she or her conscious counterpart must have a false belief about their knowledge of colors. This seems to be incoherent, considering the fact that both of them know everything physical there is to know about color perception.

If my arguments succeed up to now, then I have shown that Mary does not provide an argument for epiphenomenalism at all. This is in itself an important result, but it can be made more plausible by providing a positive explanation for the intuitions triggered by the Mary experiment. If it is not epiphenomenalism, what else might it be? This is the question for the remainder of this essay.

## 3. The 'What is it Like' Argument

At the very beginning of the Zombie-Mary story, it turned out that we find the Mary experiment intuitively plausible, not because of epiphenomenalist Qualia, but because of some important changes in her life which dramatically change what it is like to be Mary. This formulation is reminiscent of Thomas Nagel's famous 'what is it like to be a bat' argument (Nagel, 1974), and I want to argue that Mary's story and the *what-is-it-like* argument do in fact invoke similar intuitions for similar reasons on one specific reading. This fact in itself is not very instructive though, because some epiphenomenalists (like Chalmers (Chalmers, 1996b)) also see the Knowledge argument as merely one form of the more general what-is-it-like argument. Nevertheless, Jackson himself believes that there is a difference between his idea of Nagel's argument and the Knowledge argument. Jackson's position might be less consistent here than Chalmers' is, but Jackson's differing intuitions about both experiments are very instructive. Jackson believes that the what-is-it-like argument points towards physicalism, but the arguments he provides for that claim could, on closer analysis, also be used to defend the non-reductionist position of common sense realism. I argue that it is this position and not epiphenomenalism which provides

the background for the incredibly strong intuitions that are triggered by what-is-it-like arguments and Knowledge arguments alike.

Jackson differentiates between his Knowledge argument and what-is-it-like arguments because Jackson understands the Mary experiment, like Chalmers, in the epiphenomenalist way I have been discussing, but has a different understanding of the what-is-it-like argument. He tries to illustrate the difference by invoking another thought experiment about a guy called Fred, who sees two different colors where other people see only red. Again, the argument seems very plausible, because Fred can consistently sort out piles of cards in red and his special color. Jackson wants to understand Fred as a similar example to Mary, which shows the intuitive plausibility of the idea that Qualia are irreducible facts, but he insists that 'what-is-it-like arguments' work in a different way:

"It is important to distinguish this argument [the what-is-it-like argument] from the Knowledge argument. When I complained that all the physical knowledge about Fred was not enough to tell us what his special color experience was like, I was not complaining that we weren't finding out what it is like to be Fred. I was complaining that there is something about his experience, a property of it, of which we were left ignorant. And if and when we come to know what this property is we still will not know what it is like to be Fred, but we will know more about him. No amount of knowledge about Fred, be it physical or not, amounts to knowledge "from the inside" considering Fred. We are not Fred. There is thus a whole set of items of knowledge expressed by forms of words like 'that is I myself who is...' which Fred has and we simply cannot have because we are not him. When Fred sees the color he alone can see, one thing he knows is the way his experience of it differs from his experience of seeing red and so on; another is that he himself is seeing it. Physicalist and qualia freaks alike should acknowledge that no amount of information of whatever kind that others have about Fred amounts to knowledge of the second. My complaint though concerned the first and was that the special quality of his experience is certainly a fact about it and one which Physicalism leaves out because no amount of physical information told us what it is.

Nagel speaks as if the problem he is raising is one of extrapolating from knowledge of one experience to another, of imagining what an unfamiliar experience would be like on the basis of familiar ones. In terms of Hume's example, from knowledge of some shades of blue we can work out what it would be like to see other shades of blue. Nagel argues that the trouble with bats et al. is that they are too unlike us. It is hard to see an objection to Physicalism here. Physicalism makes no special claims about the imaginative or extrapolative powers of human beings, and it is hard to see why it need do so." (Jackson, 1982, p.132)

Obviously, Jackson did not want his thought experiments to be compatible with physicalism. According to Jackson's interpretation of the what-is-it-like argument,

physicalism and Nagel's argument are quite compatible; therefore, his own Knowledge argument must be about something fundamentally different. The following will try to show that the difference between Fred and Fred's ability to see an additional shade of red might be less clear than Jackson suggests, even though intuitively Jackson's differentiation between a specific phenomenal percept and the perceiver seems very plausible. This will be helpful, because it will pave the way for a form of non-reductionism that does not rely on such a distinction.

Is it really true then, that perceiver and percept can be so strictly separated? To answer this question we turn now to Zombie-Mary's cousin, Zombie-Fred. Consider: Fred has a special cone in his visual system that allows him to differentiate between red one and red two. After he dies, this cone is transplanted into Anne's system. Anne tells us now what it is like to see red one and two. Two years later, the zombie detector [8](#) is found, which can distinguish zombies from other people. It turns out that Fred was a zombie, but Anne is not. Now Anne is shocked. How can it be that Fred didn't really see all the beautiful colors she can see, even though he had the same cone? She realizes that from the beginning red one and red two had nothing to do with the cone but with the system that used the cone. In Fred the cone did nothing special, because he was a zombie. She shudders, because she realizes that, when she wanted to see what Fred saw, what she wanted to see was the phenomenal blackness that a zombie sees.

Jackson leads his readers down the wrong track when he emphasizes the difference between the cone and Fred. [9](#) He gives the impression that the cone is just a tool, whereas the person is the subject. This is a classical Cartesian fallacy. The cone is no more or less a tool than Fred's whole brain, or sunglasses that Fred might have worn. Postulating a subject behind the information-processing brain is postulating a Cartesian homunculus. What Anne actually wanted to know was what it would be like to be her with such a cone, or perhaps what it would be like for Fred to have the cone. In any case, she always wanted to know what it would be like to be a person in a certain state. The problem was that she did not realize the importance of the perceiving system. She did not realize that she was not interested in what the cone does on its own, but in the cone integrated in her (or another) system. She did not realize that she had no idea what it would be like to be her with the cone. There is nothing like the perception of colors that does not presuppose the entire perceiving system.

If such a difference should be impossible, then Jackson might have been right to believe that the what-is-it-like argument could be compatible with physicalism, but he incorrectly postulated a difference between the what-is-it-like argument and the Knowledge argument. In fact, the notion both arguments are not compatible with, if they are understood as suggested above, is the notion of zombies, indeed, the zombie idea can be used to call the whole structure of the what-is-it-like argument into question. The zombie argument seems to show that it is possible to see more in a qualitative experience than a problem of an individual perspective. It seems to show that there might be something to qualitative experiences which is completely independent of the experiencing system.

## 4. Common Sense Realism

If one wants to endorse a form of non-reductionism which is compatible with the idea that experiences cannot be seen as separate entities from experiencers, then one has to believe in something which I want to call common sense realism. Common sense realism starts from the premise that people with phenomenal consciousness cannot have a zombie equivalent. If this should be true, then the Mary and Fred experiments might still point to something interesting, but zombies are out of the game. Phenomenal consciousness would, in every possible world, result in at least one functional difference between zombies and us, but this would be the end of the zombie as defined in philosophical terms. This in turn would mean that there is something deeply wrong with the epiphenomenalist idea of consciousness.

If one wants to hold that Mary learns a new fact, although she had all the physical facts available, then Mary could be understood as showing that reductive materialism must be wrong, because the right causal story must contain essentially perspectival facts, i.e. it must take into account that the world is not completely describable in third person terms, because there is a form of causation which can only be described in first person terms. But for this very same reason, epiphenomenalism insults our basic intuitions as well, because the definition of epiphenomenalism requires that phenomenal consciousness be never involved in causation. [10](#) This reading is so intuitively plausible because with it, one can construct a functional role for consciousness that satisfies common sense realism. That this role is not describable in scientific terms, for reasons of essential perspectiveness, adds to the intuitive appeal of such a solution to folk psychology.

Nevertheless, like all philosophical positions on consciousness, such a reading has not only the desired consequences, but some highly undesirable side effects as well. In this case, it is the problem that such a position has to surrender an understanding of function that is describable purely in third person perspective terms. This is a very high price to pay to justify the intuitions of the what-is-it-like argument. I do not want to attempt to defend such a position, nor am I convinced that it is worth defending, but I believe that only this non-reductive position does justice to the intuitions of Nagel's argument.

## 5. Conclusion

David Chalmers has held that different forms of non-reductionism can be combined to form a stronger argument for non-reductionism. This paper could show that this is at least not always the case. The combination of two classical non-reductionist thought experiments, i.e. the zombie and Mary experiments, showed that the Mary experiment does not, by any means, provide independent support for the ontology advocated by the zombie argument. The Zombie-Mary thought experiment showed that it is not because of epiphenomenal Qualia that we find the Mary experiment so fascinating, but because of changes in Mary's life that have to be independent of phenomenal consciousness, because

we can imagine them for Zombie-Mary as well. What is more, after separating the real problem of phenomenal facts from all the other interesting features of the Mary thought experiment, it turned out that the idea of an omniscient color scientist might actually be incompatible with the idea of a zombie equivalent of that person.

On the contrary, Mary invokes intuitions that are useful for theoretical positions like common sense realism, which are strictly incompatible with the zombie idea because they assume that phenomenal consciousness will make a functional difference in every possible world. The paper did not try to defend such a position, but it wants to point out that it is obviously impossible to have it both ways. Common sense realism is intuitively very plausible, but faces massive problems if we want to integrate it into our scientific worldview. A position like epiphenomenalism may be philosophically very sophisticated, but it has to give up the pretence that the non-reductionism it defends saves the more important parts of common sense realism along the line. It does not! As we have seen, common sense realism is plausible precisely because of the assumption that epiphenomenalism denies, i.e. the idea that our phenomenology does effect our behavior. The two non-reductionist positions do not stand united against physicalism, but are every bit as incompatible with each other as physicalism is with one of the two.

I think this is an important result, and not only because it clarifies the status of a classical thought experiment in the philosophy of mind. It is also important because many non-philosophers have been attracted to the epiphenomenalist position because of the Mary experiment, and I hope that the Zombie-Mary story provides them with a tool to check whether it really was the epiphenomenalist intuition that they found convincing.

## Notes

<1>. The conceivability of a zombie world is a necessary condition for the coherence of the epiphenomenalist position. But epiphenomenalism makes even stronger claims as well: the conceivability of a zombie world only shows that there might be a world in which function and phenomenology are completely separated, but this does not entail that our world has to be such a world, as epiphenomenalists hold. As my arguments make use of the weaker zombie argument, every result I achieve counts a fortiori against epiphenomenalism.

<2>. Chalmers does not believe that epiphenomenalism is the only possible non-reductionist position, but most of his anti-reductionist arguments are typical epiphenomenalist arguments.

<3>. David Chalmers has an excellent and extensive zombie literature page at: <http://www.u.arizona.edu/~chalmers/zombies.html>.

<4>. (Rosenthal, 1997) holds that there is something like unconscious qualitative discriminations. Whether these discriminations should be called Qualia is unclear. For present purpose, I want to avoid this debate by defining Qualia as conscious.

<5>. Zombie-Mary has been used before. Steven Ravett Brown makes the point that the Qualia explanation of the thought experiment is redundant, because we have a better and more probable physical one. I agree, but I do not believe that this is the interesting problem in the Mary case. Brown's argument is a standard argument against epiphenomenalism, but it does not take into account the intuitive plausibility of the Mary case, which stems from another source, which I believe to be what-is-it-like intuitions. Brown's argument and a couple of replies, including one from Chalmers, are available at <http://www.ai.sri.com/~connolly/psyche-list-archive/1998a/0231.html>.

<6>. Chalmers is another adherent of this argument. He proposes it explicitly in (Chalmers, 1996a).

<7>. Nevertheless, Dennett was on the right track. He assumed correctly that the question of epiphenomenalist Qualia cannot be settled in the Nida-Rümelin room.

<8>. Whether a zombie detector is conceivable is by no means clear. It might be strictly impossible to develop anything that could distinguish between zombies and normal people. But even if that should be the case, the argument doesn't suffer. It is enough that Anne realizes that Fred could have been a zombie and what that would have meant for her desire to see what Fred saw.

<9>. A similar mistake has often been made in interpreting Nagel's bat example. Martine Nida-Rümelin (Nida-Rümelin, 1993) for example, holds that we cannot understand bats, because we do not know what the percepts of the sonic radar would be like. But Nagel's point is not the problem of what it is like to have a sonic radar, but the problem of what it is like for a bat to have sonic radar.

<10>. I take Naomi Eilan to argue in that direction. (Eilan, 1995)

## References

Chalmers, D. (1996a). Absent Qualia, Fading Qualia, Dancing Qualia. In T. Metzinger (Ed.), *Conscious Experience*. Thorverton: Imprint Academic.

Chalmers, D. (1996b). *The Conscious Mind*. Oxford: Oxford University Press.

Dennett, D. C. (1991). *Consciousness Explained*. Boston: Little, Brown & Company.

Eilan, N. (1995). Consciousness and the Self. In J. Bermudez & A. Marcel & N. Eilan (Eds.), *The Body and the Self* (pp. 337-357). Cambridge, MA: MIT Press.

Horgan, T. (1984). Jackson on Physical Information and Qualia. *Philosophical Quarterly*, 32, 147-183.

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

Nagel, T. (1974). What is it like to be a Bat. *Philosophical Review*, 83, 435-450.

Nemirow, L. (1990). Physicalism and the Cognitive Role of Acquaintance. In W. Lycan (Ed.), *Mind and Cognition*. Oxford: Blackwell.

Nida-Rümelin, M. (1993). *Farben und Phänomenales Wissen*. Wien: VWG.

Perry, J. (1979). The Problem of the Essential Indexicals. *Nous*, 13, 3-21.

Rosenthal, D. M. (1997). Apperception, Sensation, and Dissociability. *Mind and Language*, 2, 206-223.