철학박사 학위논문

Consciousness, Conceivability, Possibility

- Reconsidering The Zombie Argument -

의식, 상상가능성, 가능성

- 좀비 논변의 재고 -

2019년 6월

서울대학교 대학원

철학과 서양철학 전공

문규민

Abstract

In this dissertation, I reexamine the zombie argument developed by philosopher David Chalmers. The zombies argument claims that qualia, phenomenal qualities of our conscious experience do not supervene on physical facts. Since it is widely admitted that physicalism entails mind-body supervenience, the possibility of zombies refutes all possible forms of physicalism. Because of its huge implication, the zombie argument has provoked intense debates about the nature of consciousness and physicalism. In the zombie argument, a number of thorny issues concerning semantics, metaphysics, and epistemology are entangled. I will critically examine central notions and premises of the zombie argument and investigate related issues.

This dissertation is divided into four chapters. Chapter 1 deals with the notion of conceivability deployed in the zombie argument. The notion of conceivability supposed by the zombie argument is problematic. Especially, the notion of ideal conceivability turns out to be problematic. It is doubtable that possible formulations of ideal conceivability work well. The positive conceivability is also questionable. It is too intuition-sensitive and requires a complete theory of qualia, which is not given yet. If the notion of conceivability is problematic, the zombie argument may not be able to get off the ground.

Chapter 2 covers my *reductio* argument against the first premise of the zombie argument, the ideal positive primary conceivability of zombies. The consequence of the conceivability of zombies is a disjunction of qualia epiphenomenalism, Russellian monism, and interactionist dualism. In order to show that all of the disjuncts are wrong, I argue for the thesis of cognitive intimacy of qualia. Cognitive intimacy is *a priori* true, so that cognitively alienated qualia are negatively inconceivable. However, all of the disjuncts

commit to a negative conceivability of cognitively alienated qualia. By *reductio*, zombies are not ideally positively primarily conceivable.

In Chapter 3, the second premise of the zombie argument is examined. The second premise is an application of a principle that ideal positive primary conceivability entails primary possibility(CP+). Arguing against CP+, some philosophers have attempted to parody the zombie argument. These anti-zombie arguments, however, have their own problems. The Russellian illuminati argument, which is my own version of the anti-zombie argument, avoids such problems. If the argument is sound, ideal positive primary conceivability cannot be a guide to primary possibility. Thus, even if zombies are ideally positively primarily conceivable, there is no guarantee that they are primarily possible.

Chapter 4 concerns another physicalist response against the zombie argument, the Phenomenal Concept Strategy(PCS). Relying on PCS, physicalists can avoid the conclusion of the zombie argument while accepting its central premises. According to Chalmers' master argument, however, as far as phenomenal concepts are physically explicable, they cannot explain our epistemic situation. Against the master argument, I argue that PCS can maintain its explanatory potential, insofar as our epistemic situation should be characterized in topic-neutral terms. Thus, the master argument fails and even faces its own dilemma.

If my arguments in this dissertation are successful, they will lead to a fourfold argument against the zombie argument: 1) the zombie argument is based on the problematic notion of conceivability. 2) Even if the notion of conceivability is accepted, the first premise of the zombie argument, the conceivability of zombies, is wrong. 3) Even if the first premise of the zombie argument is right, the second premise is wrong. 4) Even if the second premise is right, it does not guarantee that zombies are

metaphysically possible. Therefore, the zombie argument fails.

keywords : consciousness, qualia, conceivability, possibility, David Chalmers, the zombie argument
Student Number : 2012-30825

Table of Contents

Figures

Introduction

The purpose of this dissertation is to reexamine the zombie argument developed by philosopher David Chalmers. The argument is known as the conceivability argument against physicalism: it starts by claiming that zombies i.e. creatures that are physically identical to us but lack our phenomenal feelings, or qualia, of experience are conceivable. From zombies' conceivability, Chalmers draws their possibility. If zombies are possible, qualia do not supervene on physical facts. Since it is widely admitted that physicalism entails mind-body supervenience, the possibility of zombies refutes all possible forms of physicalism. Because of this huge implication, the zombie argument has provoked intense debates about the nature of consciousness and physicalism. Indeed, since it was first presented, the zombie argument never stops being controversial. A number of thorny issues concerning semantics, metaphysics, and epistemology are entangled in the zombie argument. In this dissertation, I will critically examine central notions and premises of the zombie argument and investigate related issues.

This dissertation is divided into four chapters. Chapter 1 deals with the notion of conceivability deployed in the zombie argument. The notion of conceivability supposed by the zombie argument is problematic. About the conceivability of zombies, Chalmers provides several distinctions: *prima facie*/ideal, positive/negative, and primary/secondary conceivability. The ideal conceivability of zombies turns out to be problematic, since possible formulations of ideal conceivability only work under the assumption of the ideal reasoner but we are not the ideal reasoner. The positive conceivability of zombies is also questionable. It is too intuition-sensitive and requires a complete theory of qualia, which is not given yet. If the notion of conceivability is problematic, the zombie argument may not be able to get off the ground.

Chapter 2 covers my *reductio* argument against the first premise of the zombie argument, the ideal positive primary conceivability of zombies. The consequence of the conceivability of zombies is a disjunction of qualia epiphenomenalism, Russellian monism, and interactionist dualism. In order to show that all of the disjuncts are wrong, I argue for a thesis of cognitive intimacy of qualia: phenomena qualities of conscious experience must be at least potentially attended to or noticed by subjects of experience under non-defective backgrounds. Cognitive intimacy is *a priori* true, so that cognitively not intimate, alienated qualia are negatively inconceivable. However, all of the disjuncts commit to a negative conceivability of cognitively alienated qualia. By *reductio*, zombies are not ideally positively primarily conceivable. If this *reductio* works, even if the notion of ideal positive primary conceivability is well-defended, the first premise of the zombie argument is false.

In Chapter 3, the second premise of the zombie argument is examined. The second premise is an application of a principle that ideal positive primary conceivability entails primary possibility(CP+). Arguing against CP+, some philosophers have attempted to parody the zombie argument. These anti-zombie arguments, however, have their own problems. The Russellian illuminati argument, which is my own version of the anti-zombie argument, avoids such problems. Since the Russellian illuminati argument supposes CP+ and draws a contradictory conclusion, it would be a *reductio* argument against CP+. If the argument is sound, ideal positive primary conceivability cannot be a guide to primary possibility. Thus, even if zombies are ideally positively primarily conceivable, there is no guarantee that they are primarily possible. Even if the first premise of the zombie argument is true, the second premise is false.

Chapter 4 concerns another physicalist response against the zombie argument, the Phenomenal Concept Strategy(PCS). Appealing to special

nature of phenomenal concepts, some philosophers have tried to explain away our problematic epistemic situation with regard to consciousness, including the explanatory gap and the conceivability of zombies. Relying on PCS, physicalists can avoid the conclusion of the zombie argument while accepting its central premises. According to Chalmers' master argument, however, as far as phenomenal concepts are physically explicable, they cannot explain our epistemic situation. But PCS can maintain its explanatory potential, insofar as our epistemic situation should be characterized in topic-neutral terms. No matter how the topic-neutrality is interpreted, PCS can explain our epistemic situation regarding consciousness. Thus, the master argument fails and even faces its own dilemma. PCS is still a viable option for physicalists. That is, even if the first and second premise of the zombie argument is right, there is 'the third way' for physicalists to reject the zombie argument.

If my arguments in this dissertation are successful, all the works in those chapters will lead to a fourfold argument against the zombie argument: 1) the zombie argument is based on the problematic notion of conceivability. 2) Even if the notion of conceivability is accepted, the first premise of the zombie argument, the conceivability of zombies, is wrong. 3) Even if the first premise of the zombie argument is right, the second premise is wrong. 4) Even if the second premise is right, it does not guarantee that zombies are metaphysically possible. Therefore, the zombie argument fails.

Chapter 1
The Zombie Argument and Conceivability

1.1 Chapter Introduction

In this chapter, first, I shall introduce the zombie argument and present its implications. The argument argues that so-called *phenomenal qualities of conscious experience,* or *qualia,* do not supervene on the physical.[1] Indeed, the debate about the zombie argument may be one of the fiercest battles in the recent history of philosophy of mind. As the debate went on, the initial version has been updated over and over. Further, as the zombie argument evolved, the notion of conceivability has also been articulated. Thus, in this chapter, I will clarify what the zombie argument is first. (Section 1.2) Some of the conceivabilities Chalmers suggests are critically examined. (Section 1.3) It will be shown that the notion of conceivability faces a number of problems

1.2 The Real Zombie Argument

Originally, the zombie argument was not that complicated.[2] It took

---

1) Throughout this dissertation, I will loosely use expressions 'phenomenal qualities of conscious experience', 'qualia', 'phenomenal properties', 'phenomenal characters', and 'phenomenal states' interchangeably. The terms such as 'phenomenal consciousness', 'consciousness', and 'experience' are roughly mean the same thing, phenomenal quality of conscious experience.

2) Chalmers himself summarizes his original version as follows:

(O1) P&~Q is conceivable.
(O2) If P&~Q is conceivable, P&~Q is metaphysically possible.
(O3) If P&~Q is metaphysically possible, materialism is false.
(O4) Materialism is false. (Chalmers, 2010, p. 142)

the most straightforward form of the conceivability argument against materialism. Through almost two decades of debates, nonetheless, it has been articulated over and over. The refined version of the zombie argument can be formalized as follows:

(N1) PTI&~Q is ideally positively primarily conceivable
(N2) If PTI&~Q is ideally positively primarily conceivable, then PTI&~Q is primarily possible
(N3) If PTI&~Q is primarily possible, then PTI&~Q is secondarily possible, or Russellian monism is true.
(N4) If PTI&~Q is secondarily possible, materialism is false.

P represents a conjunction of all microphysical truths about our world. It specifies the fundamental microphysical properties, entities, and laws in the language of microphysics. Q represents an arbitrary phenomenal truth, a truth that a certain individual or organism instantiates a certain phenomenal property. Thus, P&~Q is the statement that everything is microphysically the same as in our world, but someone or something lacks a certain phenomenal property. Here, the individual or organism who shares everything physical with us but lacks the phenomenal property is the zombie twin of us, and the world satisfied P&~Q can be considered as the zombie world. I will call the original version of the zombie argument *the old zombie argument.*

There are several features of the old zombie argument. First, the original zombie argument is grounded on the conceivability *simpliciter.* P&~Q is argued as merely conceivable, and this conceivability is not articulated at all. Further, the argument directly draws the metaphysical possibility from the conceivability. Most of all, the old zombie argument is not two-dimensional yet: the background semantic of the zombie argument, *epistemic two-dimensionalism*, does not directly constitute the old zombie argument. While the defense of the old zombie argument Kripkean cases of a *posteriori* necessity is two-dimensional (Chalmers, 1996, p. 131-134), the argument itself does not involve two-dimensionalism.

Of course, the essential steps are Premise (O1) and (O2), but both are controversial: is P&~Q really conceivable? What is conceivability? Does the conceivability *simpliciter* of the zombie world entail the possibility of the zombie world? I think these questions motivated Chalmers to update the old zombie argument. To address the questions, he had to articulate the notions of conceivability and modality deployed in the old zombie argument.

(N5) Materialism is false, or Russellian monism is true. (Chalmers, 2010, p. 161)

P and Q represent the same thing in the old zombie argument. T is a statement which precludes extra, non-physical properties. A world with ectoplasm or entelechy may satisfy P&~Q. To prevent possible complications, one can conjoin P with a "that's-all" statement T. Therefore, PT states that P holds and no non-physical, alien truths hold. P can be replaced with PT. I represents the conjunction of all indexical truths, truths about I, here, and now. Indexical truths should supplement PT in the old zombie argument, because the microphysical truths are conceptually distinct from indexical truths. For instance, even though the Laplacian demon has complete microphysical knowledge, it may lack indexical knowledge, such as the knowledge that here is Seoul. If so, it is conceivable that all of those objective truths hold but that here is not Seoul. However, the fact that here is Seoul is not metaphysically distinct from all the microphysical facts about the world. Thus, such conceivability cannot defeat materialism. If Q functions like an indexical truth, then the conceivability of zombies cannot refute materialism. To fix this loophole, I should be conjoined with PT. Let us call the refined version of the zombie argument *the new zombie argument.*

The new zombie argument seems valid, and its conclusion is impressive. The conclusion of the argument is a disjunction: materialism is false *or* Russellian monism is true. I shall fully explicate what Russellian monism is and how it works in the new zombie argument in Section 2.5 and 3.2.

Compared to the old zombie argument, there are several significant differences in the new zombie argument. First, the new zombie argument is based on the sophisticated notion of conceivability. The conceivability is

specified as ideal, positive, and primary. What this ideal positive primary conceivability is will be exhaustively analyzed in following sections. Secondly, it does not directly draw the metaphysical possibility from the conceivability. Premise (N2) connects the conceivability of PTI&~Q with the primary possibility of PTI&~Q. Here, the possibility is primary, not metaphysical. The distinction between primary and secondary modality and its relevance to epistemic two-dimensionalism will be explained in the next section. Most importantly, the new zombie argument is in itself two-dimensional. Premise (N2) can be understood only in the two-dimensional framework. And Premise (N3) concerns the idea that primary and secondary intensions of P, T, I, and Q coincide. For epistemic two-dimensionalism is constitutive of the new zombie argument, it deserves to be called not only as the conceivability argument but also the two-dimensional argument against materialism. Last, the conclusion of the new zombie argument is weaker than that of the old one. It argues that materialism is false or Russellian monism is true.

As the zombie argument has been updated, it loses its initial simplicity and involves many complications. There is no conceivability *simpliciter.* One must single out the ideal, positive, primary conceivability. Also, in distinguishing two sorts of possibility and stepping from the primary possibility to the secondary possibility, one must understand the two-dimensional framework. Furthermore, Russellian monism is deeply involved in the new zombie argument. The force of the new zombie argument comes from these complicated matters. Therefore, in critically examining the new zombie argument, one must consider those related issues carefully.[3)]

The crucial point is that the new zombie argument is *the official version*

3) However, in Chapter 3, we will see that many philosophers have failed to argue against the new zombie argument because they neglected those details and complications.

of the zombie argument. As far as I know, after providing the new zombie argument, Chalmers does not provide any updated version. It seems that he thinks he fully articulated the zombie argument against materialism. “This completes the exposition of the two-dimensional argument against materialism.” (Chalmers, 2010, p. 154) Indeed, the argument is already complicated enough. Recently, Chalmers turns his attention and focuses on related matters, such as Russellian monism. Thus, the new zombie argument should be taken as the final, genuine the two-dimensional conceivability argument. Criticizing the old zombie argument does not work, for there is more developed, sophisticated version of the zombie argument. So in this dissertation, I will only deal with the new zombie argument. From now on, ‘the zombie argument’ refers to the new zombie argument.

1.3 The Characters of Conceivability

Despite the crucial role of conceivability, somewhat ironically, debates concerning the zombie argument tend to avoid the issue of conceivability. As far as I can tell, the nature of conceivability in the zombie argument is poorly understood. However, this issue of conceivability is crucial to understand my arguments in the following chapters. Fortunately, Chalmers(2002) has provided a comprehensive and instructive work on conceivability. His distinctions are quite articulated and deserve a careful look. Furthermore, Chalmers picks out a specific kind of conceivability as a genuine guide to possibility. In this section, I shall analyze his various notions of conceivability and their problems in turn.

1.3.1 *Prima facie* VS Ideal Conceivability

According to Chalmers, conceivability can be divided into *prima facie* conceivability and ideal conceivability. A statement “S is *prima facie* conceivable for a subject when S is conceivable for that subject on the first

appearance." In other words, "after some consideration, the subject finds that S passes the tests that are criteria for conceivability." (Chalmers, 2002, p. 147) To be *prima facie* conceivable, S does not need to be under any kind of tests or a deeper consideration. All that required is mere seeming of conceivability. This idea of *prima facie* conceivability is so mundane that it carries almost no weight in philosophical debates.

The real issue is ideal conceivability. "S is ideally conceivable when S is conceivable on *ideal rational reflection.*" (Chalmers, 2002, p. 147, italics added) The problem of this formulation is, as Chalmers states, it is hard to see if such an ideal reasoner is possible at all. (ibid., p. 148) It might be the case that for every possible reasoner there is a smarter possible reasoner. Chalmers thus suggests another definition, invoking *undefeatability by better reasoning*: "S is ideally conceivable when there is a possible subject for whom S is prima facie conceivable, with justification that is undefeatable by better reasoning." (ibid., p. 148) To ideally conceive of a statement, one should have undefeatable justification. Chalmers leaves undefeatability and reasoning as primitive notions. (ibid., p. 148) This is understandable since it seems unlikely that anybody can give a full, substantive analysis of such notions.[4)]

There are several loopholes in this alternative formulation. Even if the notion of an ideal reasoner is not problematic and the initial formulation of ideal conceivability is restored, we cannot know whether many significant but controversial statements are ideally conceivable, for we are not ideal reasoner. This would seriously limit the use of ideal conceivability as a reliable guide to possibility.

---

4) In this respect, ideal conceivability is similar to knowledge. Without an exhaustive analysis of the concept of justification and truth, we can handle many cases concerning knowledge. Although both ideal conceivability knowledge lack complete, explicit analysis of defining notions, we can do many philosophical works about them.

The first problem is that even though ‘conceivable on ideal reflection’ is substituted by ‘undefeatable by better reasoning,’ I see no benefit of such substitution. The alternative formulation appears no better than the initial one. Since it exorcises the notion of an ideal reasoner, the alternative formulation might get around the possible regress of more sophisticated reasoners. However, even though there cannot be any regress of more sophisticated reasoners, there can be a possible regress of more undefeatable justifications. That is, as how sophisticated a reasoner is comes in degree, how undefeatable a justification seems to be a matter of degree. Whatever the undefeatability is, one thing is clear: when a justification becomes stronger, it becomes more undefeatable. The crucial point is that as there is no limit of more sophisticated reasoners, there is no limit of stronger justifications. The strength of justification is a function of many different factors. For instance, more evidence makes a belief more strongly justified. How belief is formed or acquired also affects the strength of justification. The better a way of forming belief gets, the stronger a justification becomes. For any given evidence, there can be further evidence. To a way of forming a belief, we can say the same thing. For any mean of belief formation, there can be a better mean of belief formation. If this is the case, there can be an infinite regress of stronger justifications. And such regress would yield an infinite regress of more undefeatable justifications, which is no better than the infinite regress of the more sophisticated reasoners.

Even if one retreats to the initial formulation, there is another problem. According to the initial formulation, in order to know whether a statement S is ideally conceivable, one must know whether S is conceivable on ideal rational reflection. The problem of this requirement is clear: since we are not ideal reasoners, we generally do not, or cannot, know whether S is ideally conceivable. In answering a similar objection, Chalmers argues

> I think that there is little reason to accept this claim. Although we are non-ideal, we can know that it is not ideally conceivable that 0=1 and that it is ideally conceivable that someone exists. We know that certain things about the world (say, that all philosophers are philosophers) are knowable *a priori* and that certain things about the world (say, that there is a table in this room) are not so knowable even by an ideal reasoner (Chalmers, 2010, p. 155).

This reply misses the point. The question is not that is there any statement that can be known to us as ideally conceivable. The question is that how can we know whether S is conceivable on ideal rational reflection, though we are non-ideal. Merely mentioning examples of the ideally conceivable cases cannot be the answer. The issue is explaining how they are known to non-ideal reasoners like us.

Further, I think there is a plausible explanation for the mentioned cases. Consider the following conditional: if S is *easily provable* as conceivable or inconceivable on *non-ideal rational reflection*, S is conceivable or inconceivable on ideal rational reflection. Though I think this conditional is almost an *a priori* truth, whether it is *a priori* or not does not matter in the current context. Once we accept the conditional, all the cases Chalmers mentions can be explained. We can know 'someone exists' is ideally conceivable even though we are not ideal-reasoners, for 'someone exists' is easily provable by forming perceptual images of a situation where someone exists. How can we know that '1=0' is inconceivable even on ideal rational reflection? Because we can know that '1=0' is easily provable as inconceivable on non-ideal rational reflection. '1=0' is easily provable as contradictory, and we know that a contradictory statement is inconceivable even under ideal reflection. If we replace 'conceivable/inconceivable' with 'knowable/unknowable *a priori*', other examples can be explained in the

same way. Chalmers' examples do not show that non-ideal reasoners can know that S is ideally conceivable or inconceivable. Rather, they suggest that the conditional in question may be true at best.

Even if the conditional is correct, it tells us nothing about how non-ideal reasoners can know whether some statements are ideally conceivable or inconceivable, if they are not easily provable as conceivable or inconceivable. There is a lot of such statements in philosophy. For instance, can a Spinozan statement 'God necessarily exists' be easily proven as conceivable? How can we easily prove whether 'there are some coincident objects' is conceivable or not? What about statements arguing for junky or gunky worlds? Are statements of radical skepticism easily provable as inconceivable? Finally, is $\Box PTI \supset Q$ (physicalism) or $PTI \& \sim Q$ (the zombie world) easily provable as conceivable? The conditional mentioned above only applies to easy, simple, or trivial statements. It cannot be applied to many sophisticated, complicated, and substantial statements in philosophy. Whether a certain philosophically substantial statement is easily provable as conceivable or not is always controversial. In such cases, we cannot know whether such statements are ideally conceivable or not.

I am not saying that we cannot conduct any idealized thought experiment. What I am saying is that such idealized thought experiment works only in *some* cases. Imagining what would be possible for an ideal reasoner was a popular tool for intellectual investigation. Theologians asked what God would know, and physicists wondered what the Laplacean demon would know. In such cases, idealized thought experiments usually yield strongly intuitive conclusions. This success nonetheless cannot be generalized to all thorny issues in philosophy. Asking what if our cognitive limitations were removed is not a silver bullet for complicated philosophical debates about conceivability. For instance, can the Laplacean demon conceive a world where metaphysical nihilism is the case? Can God conceive of his or her

absence? Can an omniscient scientist with complete physical knowledge imagine the zombie world? It is clear that answers to these questions would diverge according to one's metaphysical, epistemological, or even ideological stance. Then, what matters is not an idealized thought experiment itself. Further *arguments* for conceivability or inconceivability would be needed. Merely conducting idealized thought experiments would not settle the issue of ideal conceivability. Unless such arguments are given, one would not be able to know whether a certain statement is ideally conceivable or not. In short, at least in some difficult cases, one cannot know what is ideally conceivable by merely conducting such-and-such idealized thought experiments

For those who want to use ideal conceivability as a reliable guide to possibility, this would be bad news. They will hold that ideal conceivability entails possibility. Based on this entailment, they may attempt to argue that a certain philosophical statement is possible because it is ideally conceivable. However, in order to argue that the statement is ideally conceivable, they must explain how they can know that it is conceivable on ideal rational reflection, even though they are non-ideal reasoner. All in all, no matter how the notion of ideal conceivability is formulated, it faces several problems.

1.3.2 Positive VS Negative Conceivability

According to Chalmers, there can be negative and positive notions of conceivability. Negative conceivability is simple. It can be defined relative to knowledge or beliefs. Chalmers' definition is that "S is negatively conceivable when S is not ruled out *a priori*, or there is no (apparent) contradiction in S." (Chalmers, 2002, p. 149) For Chalmers, negative conceivability is purely *a priori* matter: to claim a statement is negatively conceivable, one should find apparent contradiction in that statement by *a*

priori process.

The disturbing kind of conceivability is positive conceivability. Chalmers claims that in order to positively conceive S, one must be able to "form some sort of *positive conception* of a situation in which S is the case." (Chalmers, 2002, p. 150, emphasis mine) In articulating the notion of positive conception, imagination plays a central role. Chalmers says "to positively conceive of a situation is imagine (in some sense) a specific configuration of properties and objects." (ibid., 150) Imagining a situation usually requires fine-grained details. Also, a sort of interpretation and reasoning is accompanied. Through these interpretative and reasoning processes, one can find out that the imagined situation is where S is the case. Then, one can say that the imagined situation *verifies* S. If the imagined situation turns out to verifies S by interpretation or reasoning of a subject, the subject can be said to imagine that S. (Chalmers, 2002) In forming a positive conception of a situation, two different processes are intertwined: One is *psychological process* of imagining a specific configuration of properties and object. Another is *rational process* of interpreting or reasoning the imagined configuration. If one can form a positive conception of a situation that verifies S, we can say that one can positively conceive S.

Varieties of positive conceivability can be provided by classifying various notions of imagination. (Chalmers, 2002, p. 150-151) Chalmers divides imaginations into two kinds. The first imagination we can easily come up with is *perceptual imagination.* Subjects can make a perceptual image that represents S as being the case. When a perceptual image relevantly resembles a perceptual experience which represents that S is the case, the image represents S as being the case. Chalmers argues that one should not confuse perceptual imagining that S with merely supposing that S, or with entertaining the proposition that S. When a subject perceptually imagines

that S, the attitude she takes is not only toward an abstract entity such as proposition but also toward a specific situation, which is in a verifying relationship with S. Reasoning about that situation, one takes it to be the one that verifies S. The situation represented by the perceptual image can be considered as an "intermediate mental object" that mediates the subject and S. (ibid., p. 150) When a subject perceptually imagines that S, the subject must form a mental object, which is a perceptual image, that verifies S. Following Yablo, Chalmers calls this special property "mediated objectual character." (Yablo, 1993)

The second sort of imagination is *modal imagination*, which is not grounded by perceptual imagery. (Chalmers, 2002, p. 151) We can clearly imagine of situations that go beyond our possible range of perception. For example, it seems impossible to make visual images of atoms or Germany's winning the Second World War. Also, many things cannot be perceived in principle, such as the invisible, untouchable, auditable, and so on. Moreover, several situations which cannot be distinguished by perception in principle also can be imagined. Consider a physical situation postulated by two theoretically different scientific hypotheses that have the same explanatory powers and testable predictions. Though perceptually indistinguishable in principle, they are entirely different objects of our imagination. It is clear that in these cases we do not or cannot form perceptual images to imagine a certain situation. Analogous to perceptual imagination, modal imagination also has a mediated objectual character. In order to modally imagining that S, one should imagine of a world, or a situation, that verifies S. In this case, a situation is a configuration of objects and properties within a world or a part of a world. Thus, there seems to be an essential difference in 'media' of imagination of a world or situation. As perceptual imagination is mediated by perceptual image, modal imagination is mediated by intuition. In perceptual imagination, a subject imagines a certain specific situation by

forming perceptual images about the situation. In modal imagination, however, intuition takes a somewhat creative role. A situation is imagined *by having "an intuition of(or as of)" the situation.* (ibid., p. 151) For instance, when a subject modally imagines that a system of basic particles exists, she cannot have a perceptual image of that system. The subject can nonetheless have an intuition of a certain configuration of particles. Once the subject has an intuition of such situation, by reflection upon the situation, she can find out whether the imagined situation of which she has an intuition is where a system of basic particles exists.[5] Being mediated by intuition, or intermediate mental object, modal imagination acquires a mediated objectual character as perceptual one does. "This objectual character […] is distinctive of positive conceivability." (ibid., p. 150)

Modal imagination is a combination of psychological and rational processes. Both processes, however, are not immune to possible mistakes and flaws. One might think that she can imagine something contradictory. She might imagine a specific situation in a somewhat sloppy manner, and misinterpret the situation as verifying a certain contradictory statement. To avoid this kind of errors, Chalmers(2002) introduces the notion of *coherency*. "S is positively conceivable when one can coherently modally imagine a situation that verifies S" and "A situation is coherently imagined when it is possible to flesh out all arbitrary missing details in the imagined situation such that no contradiction reveals itself." This is the "core notion of positive conceivability." (ibid., p. 153) Coherency involves both processes of modal imagination. Psychologically, it must be possible to fill in every missing detail the imagined situation that verifies S. Rationally, it must be impossible to find any contradiction in that fully detailed situation.

The notion of positive conceivability faces at least three problems. First,

5) As far as I can tell, "intuition", "imagined situation", and "intermediated mental object" are different expressions of the same thing.

positive conceivability is too sensitive to intuition. It is so dependent upon intuition that debates about positive conceivability may collapse into the matter of conflicting intuitions. Second, the rational process of coherent modal imagination appears to presuppose theory. If so, even when two subjects share the same intuition, as their theories involved in the rational process may differ, positive conceivability can be underdetermined. What is worse is that since we do not have a reliably agreed theory of qualia, the positive conceivbility of zombie cannot be determined.

The first problem is that positive conceivability is overly *intuition-sensitive*, as intuition is constitutive of positive conceivability. The psychological process is having an intuition of a specific situation and adding arbitrary details. This detailed intuition becomes a sort of 'input' to the rational process. In other words, the function of the psychological process is providing an "intermediate mental object" or "imagined situation" to be interpreted or reflected by subjects. The problem of this account is that intuition can diverge among different subjects. Whatever it is, having an intuition of is a sort of psychological process that is deeply rooted in subject's psychology, philosophy, and even ideology. Then, it is obvious that there can be disagreements in intuitions. Philosophers always agree to disagree in their intuitions, and it is hard to have philosophy-free or ideology-free intuitions. Some may have an intuition of a certain situation but others may not. In other words, some may have an input to their reasoning but others may not. The issue of positive conceivability seems to overly depend on intuition. Whether a statement is positively conceivable may easily boil down to a conflict among incompatible intuitions.

This intuition-sensitivity of positive conceivability directly affects the debates concerning the zombie argument. To decide whether PTI&~Q is positively conceivable, we must have an intuition of a certain situation that may or may not verify PTI&~Q first. Without the input of such intuition,

we cannot even start our interpretation, reflection, or reasoning. Here, difference of intuition comes in. On the one hand, as Hilbert and Bernays had a clear intuition of a finite configuration of symbols, some may have an intuition of a specific situation that may or may not verify PTI&~Q. On the other hand, as Brouwer and I failed to share Hilbert and Benays' intuition, some may fail to have an intuition of such situation. Deciding whether PTI&~Q is positively conceivable becomes an issue of diverging intuition which does not allow any further intellectual endeavors.

The second problem concerns the rational process of coherent modal imagination. Once an imagined situation can be fully detailed by the psychological process, in order to check whether the detailed situation reveals contradiction, the rational process of interpreting, reflecting, and reasoning comes in. But the rational process cannot start from scratch. A rational subject's interpretation, reasoning, or reflection requires various epistemic preconditions, such as background knowledge or belief. In other words, the rational process of coherent modal imagination must be *theory-laden.* Coherency of modal imagination is partly determined by a theory chosen by subjects. Then, whether the fully detailed imagined situation reveals contradiction or not becomes a matter of theory. For the rational process involves not only the initially imagined situation but also arbitrary details, the theory embedded in the rational process must be able to cover all the actual and possible details. This theory should include all kinds of science, such as physics, psychology, ethics, aesthetics, mathematics, metaphysics, etc. The theory must be complete enough to transform the initially imagined situation into a world. Let us call such theory *complete ontology.* It seems obvious that coherency of modal imagination may not be determined, as we do not have the complete ontology yet. Unless the final, conclusive complete ontology is given, the question of positive conceivability cannot have a conclusive answer. Moreover, it even seems possible that

there are multiple complete ontologies. Ontologies might be significantly different among different thinkers. For instance, two ideal reasoners sharing their detailed imagined intuition but have different complete ontologies. Even though they share their psychological process of coherent modal imagination, one ideal reasoner may find that the detailed imagined situation is contradictory, but the other may not. There would be an intuitively indistinguishable but rationally different detailed imagined situation. In short, positive conceivability of S can be *underdetermined by intuition*.[6)]

This underdetermination of positive conceivability provokes another problem for the positive conceivability of zombies. If certain complete ontology is needed, what would it be like? To decide whether a certain configuration of properties is contradictory or not, we must know what those configured properties *are* first. Hence, our complete ontology must include

---

6) Some might object that the underdetermination cannot occur between the ideal reasoners. This seems to be a mistake. Even in ideal cases, the theory-ladenness of rational process does not go away. Ideal reasoners are ideal only in the sense that they are free of all contingent cognitive limitations. Even if ideal reasoners' cognitive capacities are unlimited, this does not determine what complete ontology they would have. Moreover, what makes positive conceivability underdetermined is complete ontology, not cognitive capacities. It is clearly possible that cognitively unlimited reasoners lack any complete ontology. Then, it would also be possible that two ideal reasoners have different complete ontologies. Suppose that there are two Laplacean demons. Both are cognitively unlimited and have the same psychology. However, they suffer a sort of ontological conflict. One of them is a dualist, but the other is a materialist. Consider they are engaging in a debate concerning whether PTI&~Q is positively conceivable. *Ex hypothesi*, two demons share the same intuition, so that both have the same detailed imagined situation. Due to their incompatible but equally complete ontologies, however, their rational reflection upon the share situation cannot be the same. Under the dualist demon's ideal rational reflection, the shared situation is revealed as coherent and verifying PTI&~Q. On the other hand, under the materialist demon's ideal rational reflection, the shared situation is revealed as contradictory in somewhere and not verifying PTI&~Q. The dualist demon will argue that PTI&~Q is positively conceivable, but the materialist one will not. Even for ideal reasoners, the question of positive conceivability of PTI&~Q remains opened.

metaphysical nature of such properties. If so, in order to know whether an intuition of situation where all physical properties fixed but an arbitrary phenomenal property is omitted is contradictory or not, one must know what phenomenal properties are first. That is, the required complete ontology should include *a complete theory of phenomenal consciousness.* Obviously, we do not have such theory yet. We only have a few competing hypotheses at best, and whether zombies are positively conceivable depends on which hypotheses will win. If phenomenal properties of consciousness turn out to be some sort of functional or representational properties, even if we have an intuition of the zombie world, we would interpret it as contradictory. We, however, do not know which theory is right about the nature of phenomenal properties yet. Without complete theory of phenomenal consciousness, one cannot rationally process any intuition of the zombie world. Without rational processes of interpretation or reasoning, one cannot know PTI&~Q is positively conceivable. Therefore, the first premise of the zombie argument cannot get off the ground.

1.3.3 Primary VS Secondary Conceivability

The third distinction is primary and secondary conceivability. Chalmers claims that S is primarily conceivable(or epistemically conceivable) when it is conceivable that S is actually the case" and "S is secondarily conceivable when S conceivably might have been the case". (Chalmers, 2002, p. 157) Primary conceivability is also called epistemic conceivability, and secondary conceivability subjunctive conceivability. Primary conceivability is based on the idea that the actual world *might be* different in various ways. These ways the actual world might be can be thought of as epistemic possibilities, which is roughly defined not being ruled out *a priori*: "it is epistemically possible that S if the hypothesis that S is not ruled out *a priori*." (ibid., p. 157) For example, 'Hesperus≠Phosphorus' is epistemically possible in that

the actual world might be the world in which 'Hesperus≠Phosphorus' is the case. Secondary conceivability is grounded on the different idea that there are several different ways the actual world *might have been.* These ways the actual world might have been can be considered as metaphysical possibilities, which are usually determined *a posteriori*. For instance, 'Hesperus≠Phosphorus' is metaphysically impossible in that the actual world cannot have been the world in which 'Hesperus≠Phosphorus' is the case.

Primary and secondary conceivability correspond to *a priori* and *a posteriori* respectively. When we primarily conceive S, how the actual world has turned out is temporarily suspended. The only thing that matters is being ruled out *a priori* or not. Although the watery stuff in our world turns out to be H_2O, we certainly can think of a specific situation where the watery stuff in our world turns out to be XYZ. For the imagined situation is consistent and reveals no contradiction, we can rationally judge that the imagined situation verifies 'water≠H_2O'. Secondary conceivability is different. When we secondarily conceive S, we cannot suspend how the actual world has turned out. Rather, we can conceive S only after we know how the actual world turns out. Philosophers, appealing to Kripkean modal error scenarios, usually say that 'water≠H_2O' is not even conceivable. The fact that water actually turns out to be H2O determines the referent of the term 'water' trans-worldly. Secondary conceivability is thus necessarily *a posteriori* matter.

1.3.4 The Characters of Conceivability

Chalmers chooses *ideal primary positive conceivability* as a genuine guide to primary possibility. "Ideal primary positive conceivability entails primary possibility." (Chalmers, 2002, p. 171) According to the analyses so far, ideal primary positive conceivability can be defined as follows:

S is ideally primarily positively conceivable when (i) S is *prima facie* conceivable, with justification that is undefeatable by better reasoning and (ii) a situation where S is actually the case can be coherently modally imagined.

Condition (i) states that S must be ideally conceivable, and (ii) claims that S should be positively and primarily conceivable. It is worth emphasizing that only this specific kind of conceivability matters in the zombie argument. One can draw a substantial modal claim only from ideal primary positive conceivability. It is the key to modality. Therefore, in this section, based on my analysis, I will summarize three main characters of ideal primary positive conceivability. These characters will play crucial roles in my argument and analysis in the following chapters.

First, ideal primary positive conceivability is *psychological.* This feature comes from positive conceivability. For S to be positively conceivable, a situation where S is the case must be coherently modally imagined. As a result of this imagination, subjects can have an intuition of the situation. By adding arbitrary details to the initially imagined situation, the imagined situation becomes a complete world. Having an intuition of a situation and arbitrary detailing consist of the psychological process of coherent modal imagination. Whatever they are, having an intuition and detailing are a matter of psychology. It involves certain psychological processes. In short, ideally primarily positively conceiving S is essentially psychological.

Ideal primary positive conceivability is also *rational.* All three kinds of conceivability are grounded by rational notions in one way or another. We have seen that ideal conceivability is grounded by essentially rational notions, such as undefeatability and reasoning. Positive conceivability is also based on rational processes. In coherent modal imagination, once the psychological process of having an intuition and detailing is done, only

thing that left is the rational process of interpretation, reasoning, and reflection of the detailed imagined situation. In the end of this process, the detailed imagined situation may reveal itself as verifying S. The interpretation, reasoning, and reflection are clearly rational notions. Likewise, primary conceivability requires such rational processes. For all we know *a priori*, if imagined actual situations that may or may not verify S are consistent and reveals no contradiction, there seems to be no reason why the actual world cannot turn out that S is the case. This process must be rational. Thus, all of the three kinds of conceivability are rational in their nature.

Last, ideal primary positive conceivability is *digital*. Conceivability is always an *all-or-nothing* matter. This all-or-nothing nature is found in all kinds of conceivability. It is obvious that ideal conceivability is all-or-nothing, as long as it is defined in terms of ideal rational reflection. On ideal rational reflection, S is conceivable or not. S cannot be hard or easy to be conceivable. Also, positive conceivability should be all-or-nothing. If a certain specific situation is coherently modally imagined and turns out to verify S, it is positively conceivable. If it does not verify S, S is positively inconceivable. S is verified or not. There is no middle ground. Primary conceivability also should be all-or-nothing for the same reason. If it is coherently modally imaginable that the actual world turns out to be where S is the case, S is primarily conceivable. If it is not, S is primarily inconceivable. Therefore, ideal primary positive conceivability is digital.

There is no gray area or middle ground between the conceivable and the inconceivable. If someone feels that there may be degrees in ideal primary positive conceivability, it may be because she conflates conceivability with *probability of actual conceiving*. Of course, there may be probability of actually conceiving S, and probability is a matter of degree. The issue of conceivability, however, is not how probable actually conceiving S is. The

issue is whether S is conceivable or not. Talking about whether S is difficult or easy to conceive misses the point of conceivability. Although it can be hard or easy for some actual subjects to conceive S, S itself cannot be hardly or easily conceivable.

These three characters of ideal positive primary conceivability will play crucial roles in the following chapters. Though I have pointed out several problems, I shall assume that the notion of ideal primary positive conceivability is consistent enough to be used in the zombie argument.[7]My aim in the following chapters is showing that even if the notion of conceivability is unproblematic, the central premises of the zombie argument do not hold.

---

7) Terminological notes: in the following chapters, I will loosely use expressions ‘conceivability of zombies’ or ‘conceivability of the zombie world’ and ‘conceivability of PTI&~Q’ interchangeably. So when I use an expression ‘imagine a certain situation’ without any special note, it is synonymous with ‘coherently modally imagine a certain situation.’

Chapter 2
The Inconceivability of Zombies

2.1 Chapter Introduction

In this chapter, I shall argue against the first premise of the zombie argument. The consequence of the conceivability of zombies is a disjunction of three different theses. I will show that all disjuncts are wrong. As for a background, in Section 2.2, the cognitive intimacy thesis will be introduced and defended. In Section 2.3, the first disjunct of the consequence of the conceivability of zombies, qualia epiphenomenalism, is rejected. It will be shown that qualia epiphenomenalism must allow a negative conceivability of a negatively inconceivable scenario. The second disjunct, Russellian monism will be critically examined and rejected in Section 2.4. Like qualia epiphenomenalism, Russellian monism entails a negative conceivability of a negatively inconceivable scenario. Finally, in Section 2.5, interactionist dualism, which is the last disjunct of the consequence of the conceivability of zombies, will be tackled. Interactionist dualism can be rejected in such a way that qualia epiphenomenalism and Russellian monism are refuted. This will complete my *reductio* argument against the conceivability of zombies.

2.2 Cognitive Intimacy

Cognition often follows through experience: when we have experience, not always but usually, we are aware of our experience. This close relationship between experience and cognition is the topic of this section.

2.2.1 Cognitive Intimacy

About the nature of phenomenal qualities of conscious experience, I suggest the following thesis:

Cognitive Intimacy: under non-defective backgrounds, phenomenal qualities of conscious experience must be potentially paid attention to or noticed by a subject of experience.

What cognitive intimacy states is simple: unless a subject of conscious experience is under defective background conditions, qualia must be in a position to be attended to or noticed by the subject. Conversely, any subject of conscious experience must be in a position to pay attention to or notice qualia of his or her experience. This thesis needs some clarifications.

First, paying attention to and noticing are cognitive *processes.* They are not cognitive *states*, such as judgments, beliefs, and knowledge. Cognitive processes are also distinct from cognitive *contents*, which represent the world or self. Indeed, in cognitive and clinical psychology, cognitive contents and cognitive processes are considered as two independent variables. Cognitive process can be treated as sort of *mental act.* When we pay attention to or notice something, we, in a cognitive sense, *do* or *act* upon that thing. What makes cognitive processes special is this active nature.

Cognitive processes can be classified in various ways. If cognitive processes involve information about mental states, let us call them *introspective* cognitive processes. There is a related question of how cognitive processes operate or how they are driven. Cognitive processes can be *top-down/control-driven* or *bottom-up/stimuli-driven.* If they are top-down/control driven, a subject's attending to and noticing would be *active, reflective* or *higher-order.* On the other hand, if cognitive processes are bottom-up/stimuli driven, they would be *passive, pre-reflective* or *first-order*. Most of all, cognitive processes themselves can either be *conscious* or *unconscious*. While these distinctions are crucial in empirical research, cognitive intimacy thesis stands neutral on such distinctions. In

what follows, ‘cognitive processes’ is used to refer to two introspective cognitive processes: attending to and noticing.

Second, what backgrounds are and how they can be compromised is relative to theoretical and empirical advances of cognitive science and neuroscience. The term “backgrounds” can refer to both varieties of computational processes and their physical substrates. Relative to theoretical and empirical advances, we may identify what the backgrounds for cognitive processes are and when they go abnormal. It must be noted that the non-defective backgrounds cover not only normal but also *ideal* conditions. For not only normal subjects but also ideal creatures without cognitive limitation, phenomenal qualities of their conscious experience must be potentially attended to or noticed. The Laplacian demon for instance, must be able to pay attention to the painfulness when he is suffering severe migraine.

Third, it is worth emphasizing that cognitive intimacy is very weak. It claims that qualia do not need to be actually attended or noticed by subjects of experience. Instead, they must be *potentially* attended or noticed. Clearly, there seems to be a lot of qualia that actually slip out our range of attention or notice. For example, if someone is so distracted by a flurry of office activities, she may not actually appreciate the taste of coffee she is sipping. She has a certain gustatory experience, nonetheless. The quality of taste can be attended to or noticed by her. If she focused on the taste, or the taste itself were somehow intensified, she could attend to or notice them. Many qualia are, or can be, out of our scope of attention and notice. I take this possible or actual dissociation between cognition and experience as data. Such dissociation does not bother cognitive intimacy. What the thesis claims is that qualia ‘might have been’ paid attention to or noticed by a subject if we ‘were’ in the non-defective backgrounds. The crucial point is that even if subjects are in the non-defective backgrounds, qualia do

not have to be actually paid attention to or noticed. All that matter is that they *can* be so. Compared to higher order theory of consciousness, cognitive intimacy is clearly weaker. (Rosenthal, 1986; 1993; 2005) According to the theory, to be phenomenally conscious, there must be actual higher-order states of the first order states. Cognitive intimacy exactly denies such commitment. No actual higher-order states are needed at all. All that needed is potential cognitive processes.[8)]

2.2.2 Arguments for Cognitive Intimacy

Why should we accept cognitive intimacy? I think there are at least three *a priori* reasons that are based on conceptual analysis of phenomenal qualities of conscious experience.

The first reason comes from the *phenomenal* part of phenomenal quality of conscious experience. Almost in all contexts, implicitly or explicitly, a phenomenal quality of experience is something that can *appear* or *reveal itself to someone.* Further, how phenomenal qualities can appear to or reveal themselves to subjects of experience determines their nature, what they essentially are. As Strawson clearly pointed out, phenomenal qualities are "properties which are of such a kind that their whole and essential nature as properties can be and is fully revealed in sensory-quality experience given

---

8) The idea that qualia must be potentially attended to or noticed by subjects of experience resonates with what Nagel says about *what-it-is likeness.* (Nagel, 1974) There is nonetheless a crucial difference between Nagel's analysis and cognitive intimacy. Nagel appears to argue that a phenomenal quality of experience is something that *is* like for a subject. However, cognitive intimacy claims that a phenomenal quality is something that *can* be like for a subject. In other words, Nagel seems to think that being actually appeared to the subject is necessary *and* sufficient for something to be a phenomenal quality of experience. I think, however, the actual appearance to the subject may be sufficient but not necessary for phenomenal qualities. In this sense, cognitive intimacy *potentializes* Nagel's idea of "the subjective character of experience", transforming what-it-*is*-likeness into what-*can*-*be*-likeness**.** (ibid., p. 436)

only the qualitative character that that experience has." (Strawson, 1989, p. 224) For instance, the phenomenal redness of the ripe tomato is phenomenal in virtue of the fact that it can appear to or reveal itself to Mary. This perspectival character is constitutive of the nature of phenomenal qualities. And in order to appear to or reveal themselves to subjects, qualia must be at least potentially attended to or noticed by the subjects. That is, *being phenomenal entails being perspectival, and being perspectival entails cognitive intimacy of phenomenal qualities.* I think cognitive intimacy must be taken to all of those who hold that phenomenal quality of experience is in itself perspectival. As far as we maintain the notion of phenomenal quality described so far, denying cognitive intimacy would always bring an unintelligible consequence that there can be an appearance that never appears to anyone or a revelation that can be revealed to no one.

The second reason is that the *conscious* part of phenomenal qualities of conscious experience seems to entail cognitive intimacy of qualia. If certain *mental* properties are instantiated by conscious experience, they are in the conscious level of mind.[9] In other words, if a mental property is of conscious experience, it must be something that one can be *conscious of*. If certain mental properties cannot be attended to or noticed, there is only one possible explanation why it is so: it is because they are in the *unconscious* level of mind. This consideration strongly suggests that if phenomenal qualities are of conscious experience at all, they must be cognitively intimate. *Being of conscious experience implies being in the conscious level of mind, and being in the conscious level of mind entails cognitive intimacy of phenomenal qualities.* If so, denying cognitive intimacy of phenomenal qualities combines two incompatible claims: on the one hand, as far as

9) This does not mean that we are in a position to introspect every property of conscious experience. Properties such as being produced by certain neural mechanisms, being maintained by molecular structures, or causing certain physiological effects can be neither noticed nor attended by subjects.

phenomenal qualities are of conscious experience, they must be properties that we can be conscious of. On the other hand, insofar as phenomenal qualities are not cognitively intimate, they are properties that we cannot be conscious of.

Last, the fact that we can be *certain* about phenomenal qualities of our experience entails cognitive intimacy. I think it is undeniably true that we *can* be certain about phenomenal qualities at least. Yes, sometimes when one is "out of her mind" or "losing it", one might not be certain about his or her own experience. In that case, one should ask his or herself 'am I really experiencing this?' Even if this is possible, it never bothers the possibility of being certain about what her experience is like. If qualia can be cognitively not intimate, however, one cannot be certain about what it is like to have that experience in principle. Even when we are highly focused on normal or even enhanced backgrounds, there always will be a skeptical scenario that may falsify our belief about our own conscious experience. For example, even when Mary encounters the ripe tomato, she cannot be certain about what her visual experience is like. While she is not deranged, there is still a skeptical scenario that she is not visually experiencing phenomenal red *alone*. Maybe she is experiencing the phenomenal red *with* a phenomenal yellow or even phenomenal rainbow. In that case, the yellow or rainbow qualia cannot be noticed and entirely hidden from Mary's perspective. Mary cannot be certain that she is seeing only the phenomenal redness, since there can always be a mixture of noticed qualities and unnoticeable ones. This situation generalizes to every possible experience. There always will be varieties of skeptical scenarios that defeat one's belief about what she experiences. This is not merely implausible but also wrong. We clearly can be certain that we are experiencing only certain things and nothing else. Indeed, except some mathematical or logical truths, phenomenal qualities are only things that we can be certain about.

It seems that qualia must be potentially attended to or noticed by subjects of experience. The conceptual analysis of the notion of a phenomenal quality of conscious experience tells us that being non-defective, being phenomenal, and being of conscious experience entail cognitive intimacy of qualia. Moreover, the idea that phenomenal qualities of experience can be hidden from a subject's cognition is incompatible with the possibility of certainty of experience. While the evidence of cognitive intimacy is overwhelming, strong counterexamples are hard to find. All these considerations lead to the conclusion that we should accept cognitive intimacy.

What I want to emphasize is *a priori* status of cognitive intimacy. Note that in defense of cognitive intimacy, I have never appealed to anything empirical but solely relied on *a priori* reasons and conceptual analyses. I believe cognitive intimacy is not something that can be confirmed or refuted *a posteriori*. It is justified *a priori*, and I believe it is an *a priori* truth. This is also Chalmers' point. He says "there is not even a *conceptual possibility* that a subject could have a red experience like this one without having any epistemic contact with it: to have the experience is to be related to it in this way." (Chalmers, 1996, p, 107, italics added) This intimate cognitive relation also partially explains why qualia have been characterized as *immediately accessible*.[10] It is *a priori* true that qualia are cognitively

10) Seager(2016a) provides a good summary of the traditional characterizations of qualia. Among such characterizations, the fourth essential property is important in the current context: "(4) Qualia are *immediately accessible.* The minimal explication of this notion is that we are non-inferentially aware of our modes of consciousness, of the way that things currently seem to us." (ibid., p. 165) Cognitive intimacy can explain the immediate accessibility. Immediate accessibility means that qualia are cognitively accessed from the first-person perspective of subjects without any inference or empirical observation. Once we direct our attention inward and focus on our own experience, we are directly aware of what it is like to have that experience. This being aware of experience by introspection is in itself a cognitive process. Saying that qualia are immediately accessible to a subject is another way of saying that qualia must be paid attention to or noticed by the subject. In other

intimate. In other words, it is conceptually impossible that qualia are *cognitively alienated*.

2.3 Epiphenomenalism and Dull Jane

In the previous section, I have set my theses about the nature of phenomenal qualities and cognitive processes. In this section, my examination of the zombie argument is started. I shall summarize the most common reaction to the zombie argument and Chalmers' reply. Perry(2001) has raised an issue of qualia epiphenomenalism against the conceivability of zombies. Chalmers' first response is denying that the conceivability of zombies entails qualia epiphenomenalism. He points out there are many type-B materialists who accept the conceivability of zombies but are not committed to qualia epiphenomenalism. Second, he argues that even if the conceivability of zombies entails qualia epiphenomenalism, there is a sophisticated version of qualia epiphenomenalism that evades all the criticisms against qualia epiphenomenalism. In Section 2.3.2, I shall argue that Chalmers' first reply does not work, for type-B materialists do not actually accept the conceivability of zombies in the relevant sense. And I will provide a *reductio* argument against qualia epiphenomenalism in Section 2.3.3. In Section 2.3.4, possible objections are examined and rejected.

2.3.1 The Conceivability of Zombies and Epiphenomenalism

The most common and instant reaction to the zombie argument is that the conceivability of zombies entails qualia epiphenomenalism. A representative case is John Perry's critiques on the zombie argument. (Perry, 2001) Let us accept the causal closure of the physical and absence of overdetermination. In conceiving zombies, one must fix all the physical properties but subtract a particular phenomenal property. If so, there cannot be any causal role for

---

words, cognitive intimacy says that qualia are essentially immediately accessible.

the subtracted phenomenal property to play. The only way for the zombie world can be conceivable seems to be committing to qualia epiphenomenalism. Qualia epiphenomenalism, however, has never been preferred by many, as it is too counterintuitive.

Chalmers is well aware of this objection and provides a threefold response: 1) doubting the entailment; 2) biting the bullet; 3) diluting the implication. The first response comes from an observation that there are a number of physicalists who deny qualia epiphenomenalism but admit the conceivability of zombies. The second reply involves Chalmers' positive hypothesis about consciousness. The third appears to be the strongest response, which expands the debate to cover interesting views in metaphysics of consciousness. The second reply seems to repeat the traditional defense of epiphenomenalism.[11] so, I take this third reply as his

---

11) Chalmers' second reply is to argue that we may bite the bullet of epiphenomenalism. His *naturalistic dualism* can be seen as a sophisticated form of qualia epiphenomenalism. Chalmers' strategy is arguing that naturalistic dualism is not vulnerable to several criticisms usually raised against epiphenomenalism. Naturalistic dualism argues that consciousness naturally, not logically, supervenes on physical properties and phenomenal properties are fundamental. If so, there must be fundamental psychophysical laws that connect the phenomenal domain and the physical domain. Naturalistic dualism is nonetheless a version of qualia epiphenomenalism. It claims that all the causal and explanatory roles are taken by the physical. Thus, though Chalmers insists on avoiding the title 'epiphenomenalism,' naturalistic dualism must be taken as a special kind of qualia epiphenomenalism. Chalmers' attitude about qualia epiphenomenalism is dubious. He shows his dubious position by stating "I do not describe my view as epiphenomenalism" and "But the view implies at least a weak form of epiphenomenalism, and it may end up leading to a stronger sort" in a single paragraph. (Chalmers, 1996, p. 160) Actually, he goes further. Chalmers says "Any view that takes consciousness seriously will at least have to face up to a limited form of epiphenomenalism." (ibid., p. 158) Despite his ambiguous attitude, I could not find any reason not to think that Chalmers in fact commits to qualia epiphenomenalism.

Naturalistic dualism can dodge a number of problems qualia epiphenomenalism usually faces. First, it can explain away our intuition about causal efficacy of phenomenal properties. Naturalistic dualism claims that what actually causes

real response to the objection from qualia epiphenomenalism. I shall examine the first two replies in this section, the full assessment of the third one will be taken independently after this section.

2.3.2 Type-B Materialism and The Conceivability of Zombies

Chalmers' first response is denying that qualia epiphenomenalism is entailed by the conceivability of zombies. He claims that the conceivability of zombies is "accepted by many "type-B" materialists (those who accept an epistemic gap between the physical and the phenomenal but deny an ontological gap), all of whom deny epiphenomenalism: e.g., Ned Block, Chris Hill, Joe Levine, Brian Loar, and many others" and "their mere existence" shows that the conceivability of zombies has nothing to do with qualia epiphenomenalism. (Chalmers, 2004, p. 183) If the conceivability of zombies already builds in qualia epiphenomenalism, "it would require that [philosophers who accept the conceivability of zombies] be deeply irrational, or have deeply divided minds." (ibid., p. 183) The conceivability of zombies, Chalmers explains, "rests partly on prima facie conceivability intuitions that many share, and partly on deeper considerations concerning the absence of any conceptual linkage between microphysical concepts (which are structural-functional in nature) and phenomenal concepts (which

---

behavioral effects are the physical states, not the phenomenal states. Even so, by virtue of natural supervenience, there should be strong regularities among phenomenal and physical states' behavioral effects. It is natural to infer causalities from such regularities. Further, naturalistic dualism can assimilate the evolution of consciousness. According to naturalistic dualism, there must be a certain physical feature which actually brings adaptive behaviors. As a matter of fundamental laws, there must be an experience which naturally supervenes on that physical feature. Due to natural supervenience, when the physical feature is selected, the consciousness also will be selected. As a result, consciousness will have its own evolutionary history, which is parallel to the history of its physical substrates. How can consciousness can evolve through natural selection can be explained as a matter of fundamental psychophysical laws.

are not). In both cases, […] their support presupposes nothing about epiphenomenalism." (ibid., p. 183)

First of all, "the mere existence" of type-B materialists cannot be the counterexample against the charge of qualia epiphenomenalism, since type-B materialists commit to a wrong kind of conceivability. To see this, we should focus on in what sense type-B materialists admit the conceivability of zombies. As I explained in the previous chapter, the zombie argument requires the *positive* conceivability of zombies. In order to positively conceive PTI&~Q, one must coherently modally imagine a situation that verifies PTI&~Q. One must have an intuition of and reflect upon a situation that verifies. The crucial point is that all of these psychological and rational processes involve verifying *situations, specific configurations of properties and objects*. When type-B materialists claim that they accept the conceivability of zombies, however, they do not commit to any verifying situation. They accept the conceivability of zombies merely in the sense that PTI&~Q reveals no apparent contradictions. In other words, what they really admit is the *negative* conceivability of zombies. To my knowledge, they do not care about having an intuition of and reflecting upon a situation that verifies PTI&~Q. If so, type-B materialists do not accept the conceivability of zombies in the relevant sense.[12)]

Even after Chalmers distinguished positive conceivability from negative one, there are type-B materialists who fail to grasp the notion of positive conceivability. Some prominent type-B materialists complain about the distinction between positive and negative conceivability. For example, Joseph Levine confesses

12) There is also a historical reason for type-B materialists' ignorance of positive conceivability. In fact, Type-B materialists' arguments and Chalmers' reply was before Chalmers articulates his notion of conceivability. So, the debate only focused on the conceptual coherency of zombies.

> My problem is this. I don't really see the difference between positive and negative conceivability. So take this example. Chalmers claims that while the falsity of certain unprovable mathematical hypotheses (e.g., the Continuum Hypothesis) are negatively conceivable (their unprovability means that their truth or falsity is not *a priori*), they are not positively conceivable. But why not say that merely by entertaining the statement expressing the mathematical hypothesis itself one has thereby positively conceived it? How is this different from conceiving of a zombie? Of course if positive conceivability were restricted to what could be imagined, in the sense of calling up the relevant perceptual image, the distinction would make clear sense. But Chalmers doesn't want positive conceivability restricted that much. So what then determines when a description corresponds to the positively conceivable and when it doesn't? […] For Chalmers, positive conceivability is supposed to be a distinctive mental act. But my problem, as expressed above, is that *I don't really understand what this distinctive mental act is or how to determine when a statement is subject to it.* (Levine, 2011, italics added)

Levine's complaint is instructive: some type-B materialists do not even understand what positive conceivability is. Those type-B materialists would not be able to positively conceive PTI&~Q, even though they claim that they accept the conceivability of zombies. How can one positively conceive a statement without knowing what it is to positively conceive?

Therefore, Chalmers' first reply to the claim that the conceivability of zombies entails qualia epiphenomenalism fails. The examples of type-B materialists are irrelevant, as type-B materialists do not catch the right sort of conceivability that is supposed by the zombie argument. Unless there is any independent proofs that type-B materialists admit the positive conceivability of zombies, Chalmers' first replay to the charge of qualia epiphenomenalism does not work.

2.3.3 The Story of Dull Jane

Chalmers argues that even if the conceivability of zombies entails qualia epiphenomenalism, as far as qualia epiphenomenalism is consistent, there is no reason to reject it. While qualia epiphenomenalism seems consistent, when it is carefully cashed out, it will get into troubles. In this section, it will be argued that qualia epiphenomenalism entails the negative conceivability of a scenario that qualia cannot be attended to or noticed by subjects of experience. This violates cognitive intimacy, so that we have a *reductio* argument against qualia epiphenomenalism.

To start, it is crucial to understand how cognitive processes can occur. Cognitive processes are grounded by diverse activities of *information processing*: conceptualization, categorization, storage, retrieval of information, and so on. And it is hard to see how something that makes no changes or differences can generate, transmit, transform, and storage information. In this sense, these changes or differences, which can be called *traces*, are media or vehicles of information processing. How can these traces implement diverse information processing? The only thing we can think of is *causation*: all activities of information processing are supposed to be implemented by multiple procedures operating through causal chains involving traces in cognitive systems. Once information is assumed to be processed by physical systems, (our biological brain) traces must be physical. It is widely agreed in cognitive science that information processing should be realized by physical causation. However, cognitive systems can be non-physical. In this case, information is not processed by brains. It would be processed by immaterial souls, so that traces must be non-physical. The information processing by souls should be implemented by causal changes of non-physical traces.

Can qualia epiphenomenalism accept this immaterial kind of information processing? It seems that there is no *a priori* reason for qualia

epiphenomenalism not to allow immaterial souls, non-physical traces and causation. However, this move has a price. If qualia epiphenomenalism adopts information processing by non-physical causation, this non-physical causation cannot supervene on microphysical facts. If they supervene on microphysical facts, qualia epiphenomenalism cannot be compatible with the conceivability of zombies. Remind that when PTI&~Q is claimed to be conceivable, T means that there are only microphysical and indexical facts and *nothing* else. But if souls or non-physical causal chains supervene on microphysical facts, when P holds, there must be *something* else, namely facts about souls or non-physical causation. So T cannot hold. Souls' or non-physical causation's supervenience on microphysical facts violates the that's-all clause in PTI&~Q. As far as souls or non-physical causation supervenes on microphysical (plus indexical) facts, qualia epiphenomenalism is incompatible with the conceivability of PTI&~Q. In order for qualia epiphenomenalism to be compatible with the conceivability of zombies, souls or non-physical causation must not supervene on microphysical facts.

Now, all the mentioned ideas can be turned into a *reductio* argument against qualia epiphenomenalism. The driving idea is that if qualia epiphenomenalism is the case, it is at least consistent that qualia can lost their cognitive intimacy. For *reductio*, let us suppose that qualia epiphenomenalism is right. Let us further assume that the super-scientist Mary is under not defected backgrounds. When she escapes from her achromatic room and sees the ripe tomato, certain physical differences occur in her brain and cause red qualia. Since qualia are supposed to be epiphenomenal, however, Mary's red qualia cannot make any physical traces in her brain. Instead, they make non-physical traces in Mary's soul. Due to these non-physical traces and their causation, the red qualia can be attended to or noticed by Mary. On the other hand, there is Jane, who is a perfect physical doppelganger of Mary. The only difference is that Jane has no

soul. This is conceivable because souls do not supervene on microphysical facts. When she escapes from her achromatic room and sees the ripe tomato, her brain causes the same red qualia with Mary. Like Mary's red qualia, Jane's red qualia cannot make any physical traces. The crucial difference is that unlike Mary's, Jane's red qualia cannot make any non-physical traces either. Jane has no soul, so that there is nothing on which non-physical traces are registered. As a result, no causal changes can be occurred in Jane's cognitive system even when she acquires the red qualia. Since there cannot be any new causal chain at all, there can be neither information processing nor cognitive processes involving Jane's red qualia. So the red cannot be paid attention to or noticed by her. All rich experience but no cognitive process makes Jane a dull girl. Let us call the full description of this situation *the story of dull Jane.* As far as qualia epiphenomenalism is true and compatible with the conceivability of zombies, this story must be coherent. If the story of dull Jane is coherent, it cannot be ruled out *a priori* and should be negatively conceivable.

However, as far as cognitive intimacy is true *a priori*, the story of dull Jane is negatively inconceivable. In the story of dull Jane, the phenomenal redness is cognitively alienated. In Section 2.2.3, I have argued that cognitively alienated phenomenal qualities are incoherent and conceptually impossible. The dull Jane story commits to exactly such qualia, and it is conceptually incoherent. If so, the story should be ruled out *a priori* and not even negatively conceivable. By *reductio*, qualia epiphenomenalism is not only counterintuitive but also wrong. It is wrong in that it entails what is negatively inconceivable is conceivable.

2.3.4 On Non-causal Epistemic Relations

Against the critique thus far, qualia epiphenomenalists would claim that *qualia-involved cognitive processes*, namely paying attention to or noticing

phenomenal qualities, are exceptional. They would reply that at least in attending to and noticing phenomenal qualities, no information processing and causal difference are needed. Instead, there can be *non-causal epistemic relations* that enable qualia-involved cognitive processes. If there is such relation, Jane's red qualia do not have to be cognitively alienated. Although the appeal to non-causal epistemic relation seems to work at first sight, however, in what follows, I shall argue that it does not. After pointing out a problem of verifiability and falsifiability of non-causal epistemic relations, I will show why such relations cannot account for cognitive intimacy of qualia.

I think there is a principled reason to believe that no non-causal epistemic relation can help to explain qualia-involved cognitive processes. In order to show this, first, I shall reveal the cognitive structure of the most well-known kind of non-causal epistemic relations, *acquaintance.* Another kind of non-epistemic relations, which is *self-representation*, shares this cognitive structure.[13] I will argue that the cognitive structure generalizes to any kind of non-causal epistemic relations and this is why non-causal epistemic relation cannot account for qualia-involved cognitive processes.

Chalmers has pointed out an interesting aspect of acquaintance. In his theory of phenomenal concepts and beliefs, The formation of *direct phenomenal concepts* is based on the *cognitive act of attention* to phenomenal qualities of experience they pick out: "The clearest cases of direct phenomenal concepts arise when a subject attends to the quality of an experience and forms a concept wholly based on the attention to the quality, 'taking up' the quality into the concept. (Chalmers, 2003, p. 235) This

13) Though 'self-presentation' or 'self-manifestation' would be better to grasp the idea of appearance or revelation of phenomenal qualities, I will stick to the term 'self-representation', because some philosophers already have coined the term to grasp the phenomena of appearance or revelation. See (Kriegel, 2009).

cognitive act of attention is called *demonstration*. Demonstration can be characterized as a cognitive relation between subjects and phenomenal qualities that enables the formation of direct phenomenal concepts. Chalmers identifies acquaintance with this relation: "acquaintance has been characterized only as that relation between subjects and properties that makes possible the formation of direct phenomenal concepts". (ibid., p. 248) Then, acquaintance with phenomenal qualities is attending to phenomenal qualities, or acquaintance must be grounded by attention. Gertler(2001) also has independently developed a similar account of phenomenal concept, according to which a phenomenal state is introspected when it is "embedded" in another state and this state receives *demonstrative attention*.

Indeed, there is a very strong intuition that acquaintance with a phenomenal quality is *determined* by paying attention to the quality, in that it will vary directly as a function of that attention in cases where that attention varies while all other physical, cognitive, and phenomenal background conditions are fixed, and that it will not vary independently of that attention in such cases. Furthermore, across a wide range of possible cases in which the attending to the quality is varied while background properties are held constant, the acquaintance relation with the quality will co-vary with attention to that quality. In this sense, acquaintance with qualia can be said to be at least partially *constituted* by attending to those qualia. Acquaintance with qualia always starts by paying attention to those qualia, and when one attends to qualia, she is already acquainted with those qualia. Then, then acquaintance with qualia does not ground paying attention to those qualia. The opposite would be closer to the truth. Acquiantance is constitutively grounded by cognitive process of attention. If something cannot be attended even unconsciously,[14] one would not be in a position to

14) In Section 2.2.1, I emphasized that paying attention and noticing themselves are mental acts that can be either conscious or unconscious.

be acquainted with it. If something is at least unconsciously attended, one is already acquainted with it. This is *the cognitive structure* of acquaintance.

This cognitive structure applies another form of non-causal epistemic relations, self-representation. Some philosophers have been argued that once subjects have qualia, qualia *present themselves* to subjects. In this case, unlike acquaintance, we do not actively engage in "direct access" to qualia. Rather, it would be better to say that qualia appear or reveal themselves to us, and we are passively involved in cognitive *reception* of qualia. However, in virtue of what such reception is possible? In order for subjects to receive qualia, they must be equipped with some degree of notice. If a subject is so drowsy or preoccupied that she cannot, consciously or even unconsciously, notice anything, what would appear or reveal to her mind? Certainly, nothing. Nothing will appear or reveal itself to the subject's mind because she is not cognitively ready for the appearance or revelation. For example, even if we have pain, if our mind is so deflected or focused on something else, the pain will not appear or reveal itself to us as painful. If all backgrounds are fixed, noticing qualia would directly determine self-representation of those qualia: if one notices a phenomenal redness, it presents itself to her. If one does not notice the phenomenal redness, it cannot manifest itself to her. As acquaintance with qualia necessarily starts with attending to those qualia, self-representation of qualia always ends with noticing those qualia. Noticing to qualia, therefore, partially constitutes self-representation of those qualia.[15)]

Given the cognitive structure, we can understand why *any* attempt to explaining cognitive intimacy of qualia in terms of non-causal epistemic relations to qualia is doomed to fail. In order to be non-causally epistemically related to a particular phenomenal quality, subjects may

15) A similar analysis has already been provided in Section 2.2.2 for cognitive intimacy.

actively and reflectively access to their qualia or passively and pre-reflectively receive them. Acquaintance, or "direct access", is a typical case of the first, and self-representation, or "revelation", is representative of the latter. However, saying that someone directly accesses to something without attending to that thing sounds unintelligible. Saying that something reveals itself to someone but she does not notice that thing does not make sense. Whatever non-causal epistemic relation is, it must be partially constituted by qualia-involved cognitive processes. If this is the case, no matter what kind of non-causal epistemic relations qualia epiphenomenalism adopts, they cannot make qualia to be cognitively intimate. No non-causal epistemic relations can make qualia to be potentially paid attention to or noticed, because such relations hold only when qualia are actually attended to or noticed. How can a relation render something accessible, if the relation comes after when that thing is accessed?

In contrast, information processing theory of qualia-involved cognitive processes does not face such circularity. In accounting for how qualia can be attended to or noticed, the theory would simply appeal to some set of physical processes of forming, storing, transmitting, or retrieving or information. No qualia-involved cognitive processes are required in this process. Of course, there must be enabling or background conditions for the information processing, but they are not qualia-involved cognitive processes that should be explained. For a detailed, low-level explanation, one can specify the background conditions in neural terms. For a more abstract, high-level explanation, one can do the same thing in computational terms. Either way, information processing would make qualia cognitive intimate without presupposing attending to or noticing qualia.

All in all, I think there are good reasons to doubt that non-causal epistemic relation can account for cognitive intimacy. The cognitive structure of non-causal epistemic relations implies that all possible kinds of such

relations must be constituted by qualia-involved processes, so that it cannot account for how such qualia-involved processes are possible. Therefore, there seems to be no reason to believe that non-causal epistemic relation will make Jane's new qualia to be cognitively intimate.

2.3.5 Objections and Replies

There may be possible objections to the story dull Jane and my *reductio* argument. In this section, I shall consider three possible objections and argue that they are not successful.

Objection 1: your *reductio* argument works only if you already presupposed a *causal theory of knowledge.* According to the theory, in order for a belief about something to be justified, that thing must be causally responsible in formation of the belief. Likewise, according to your argument, in order for qualia-involved cognitive processes to occur, qualia must cause changes or differences. However, the causal theory of knowledge fails to be a general principle about knowledge and justification. Even Goldman, the one who first brought the theory in the field of epistemology, abandoned his own view in the face of various criticisms. If so, qualia epiphenomenalism can be safe from your argument, for there is plenty of reasons to reject its background epistemology.

Reply: this objection stems from a natural misleading of my argument. Note that I have never said anything about how experience justifies belief about it. The problem I raised runs deeper than knowledge or justification. The point of my argument is not that for phenomenal beliefs to be justified, experience must be mediated through appropriate causal connection to the beliefs. Rather, the point is that for qualia-involved cognitive processes to be occurred, information about qualia must be processed. In turn, information

about qualia to be processed, qualia must involve certain causation. The background theory of my argument is not a causal theory of knowledge, but information processing theory of cognition. Even if the causal theory of knowledge can be rejected, my argument should not be, as the theory is irrelevant to the story of dull Jane.

Objection 2: as your argument stresses that phenomenal qualities must cause changes or differences to be attended to or noticed, it might be taken as a variety of *causal theory of reference.* However, it is not obvious that the theory can be generalized to all cases of reference. Paradigmatic counterexamples include abstract objects, logical relation, future events, and fictional characters, which cannot be causally connected to a subject. As we are always talking and thinking about these things, there seems to be no strong reason to adopt a causal theory of reference. Your argument would lose much of its force, if it is built on a causal theory of reference.

Reply: this objection conflates cognition with reference. My argument claims that something must cause changes or differences to trigger information processing grounding qualia-involved cognitive processes. It does not argue that it must do so to be referred to. Cognitive processes and reference are conceptually distinct. Cases where cognition and reference come apart will make this point clear.

It is easy to find cases of *cognition without reference.* Such cases typically involve either *initial* or *sloppy cognitive processes.* To refer to something, we must have intensions or make thoughts about it. When we encounter something for the first time, on the other hand, we pay attention to or noticed it without any intension or thought about it. In such case, we come to have intensions or thoughts about that thing *in virtue of* initial noticing, attending to or noticing it. Initial cognitive processes always come

before the reference by intensions or thoughts. In this sense, they can be a case of cognition without reference. Moreover, even when a subject pays attention to or notices something, her attention and notice might not be enough. Attending to and noticing come in degrees. Only sufficient cognitive processes can make us have intensions or thoughts. This constraint of degrees suggests that there can be cognitive processes that cannot result in having intension or thoughts. Such cognitive processes would not be accompanied by any reference. These insufficient or sloppy cognitive processes can be called cognition without reference.

We also have examples of *reference without cognition.* Cognitive processes are not necessary for reference. Rather, what seems really needed is, as Chalmers noted, having *intensions* or *thoughts* about referents. (Chalmers, 1996, p. 201) For instance, suppose that I have a thought <the tallest man in the world will be taller than me>. Do I have to pay attention to or notice the actual tallest person in the world? Obviously, I do not. I cannot even do so simply because I never met him. Further, the fact that we remember people who passed away or imagine future events shows that we can have intensions or thoughts about things that are not present. However, we cannot notice or pay attention to something that is not present. In cases of abstract objects or relations, it becomes more evident that no cognitive processes are required to reference. To have a thought <5 is an odd number>, I do not have to notice or pay attention to the number 5 as an abstract object. Without any notice or attention to disjunctions or material conditionals, I can talk and think about them. In all these case, cognitive processes are not prerequisites for reference. The only thing that matters is having intensions or thoughts, or deploying concepts at best. The provided cases show that cognition and reference can come apart. My argument involves only cognitive processes, not reference.

Objection 3: you have argued that any non-causal epistemic relations must be partially constituted by attention or notice. This may not be the right description about the relationship between subjects and phenomenal qualities, however. Qualia epiphenomenalist can argue that the relationship between subjects' mind and their phenomenal qualities is so epistemically intimate that it would not be mediated by any cognitive process. Qualia are not objects 'apart' from our mind. There seems to be no 'distance' between our mind and qualia. This consideration makes a logical space for a non-causal epistemic relation to qualia that is not mediated by any attention or notice. Relying on this epistemically intimate relation, qualia epiphenomenalism would be able to account for cognitive intimacy of qualia.

Reply: the problem of this objection is that it is hard to see how there can be such epistemically intimate relation. As I have argued in the previous section, there seems to be two ways how subjects' mind can be epistemically related to phenomenal qualities: we can either actively access to our qualia or passively receive our qualia. When we actively access to our qualia, we must pay attention to those qualia. Epistemic access always requires consciously or unconsciously attending. If we pay no attention to something, in what sense we epistemically access to that thing? Indeed, epistemic access and attention are so tightly related that they are sometimes treated interchangeable. On the other hand, if we passively take our qualia, we should notice them. Epistemic reception without conscious or unconscious notice seems unintelligible. If we epistemically receive something, it implies that we somehow notice that thing. If this is the case, the epistemically intimate relation without attention or notice seems to make no sense. In such relation, our mind would neither epistemically access to nor epistemically receive qualia. How can we be epistemically related to our qualia, if we neither access nor receive our qualia? I think there is little

reason to call such relation ‘epistemic’ or ‘intimate’. That relation would be neither active nor passive. I do not see how such epistemically ‘middle’ way is even possible. If we do not access or receive qualia, in what sense our mind is related with them at all? I believe this is one of the reasons why those philosophers who endorse non-causal epistemic relations characterize such relations in terms of attention or notice.[16] All things considered, there seems to be no such ‘epistemically intimate’ relation that makes qualia cognitively intimate.

All in all, I conclude that qualia epiphenomenalism is wrong in that it entails an incoherent scenario, the story of dull Jane. The story implies that Jane cannot attend to or notice the red qualia she gains, even if she is in not defected cognitive backgrounds. Above all, the dull Jane story implies that Jane’s qualia lack cognitive intimacy. But this cannot be the case. Therefore, if my *reductio* argument against qualia works, qualia epiphenomenalism turns out to be false.

Even if qualia epiphenomenalism is wrong, there is the last resort left for Chalmers. As I show in Section 2.4.1, he can dilute the epiphenomenal implication by considering other hypotheses. Replying to Perry’s analysis, Chalmers claims

> Furthermore, the Russellian monist view is a nonepiphenomenalist view that we have seen is compatible with the conceivability of zombies in the

16) Chalmers himself admits that the relation between qualia and subjects’ mind is intimate. “This relation would seem to be a peculiarly intimate one that is made possible by the fact that experiences lie at the heart of the mind rather than standing at a distance from it, and it seems to be a relation that carries the potential for conceptual and epistemic consequences. We might call this relation *acquaintance*.” (Chalmers, 2010, p. 285) However, as we have seen, he also emphasizes that such ‘peculiarly intimate’ relation between experience and mind essentially involves attention to qualia.

> relevant sense. Finally, even Cartesian interactionist dualism, in which consciousness certainly plays a causal role, is compatible with the conceivability (and possibility) of zombies. On such a view, physically identical beings without consciousness will presumably have large causal gaps in their functioning (or else will have some new element to fill those gaps), but there is nothing obviously inconceivable about such causal gaps. (Chalmers, 2010, p. 156)

Chalmers can insist that the conceivability of zombies is still viable because it can be supported by other positions instead of qualia epiphenomenalism. *Russellian monism* and *interactionist dualism* are candidates. What this claim really amounts to is that the conceivability of zombies entails not just qualia epiphenomenalism but qualia epiphenomenalism or Russellian monism or interactionist dualism.

This response seems to the strongest one, for one must show that all of its disjuncts are false in order to refute the conceivability of zombies. For I had argued against qualia epiphenomenalism in this section, what is left is showing why both Russellian monism and interactionist dualism are wrong. In the following sections, I shall argue that both seemingly non-epiphenomenalist views face serious problems.

2.4 Russellian Monism and Flipping Inscrutables

In recent years, Russellian monism is getting attention from philosophers of mind. Roughly, Russellian monism claims that at the fundamental level of the physical, there is a certain sort of properties that somehow responsible for phenomenal qualities of our conscious experience. These properties are considered to be intrinsic/categorical properties that ground the structural/dispositional ones. Proponents claim that Russellian monism is compatible with the conceivability of zombies and can be a viable alternative of traditional physicalism. In this section, however, I shall argue

that a deeper analysis of what Russellian monism commits to shows that it is wrong. To show this, first, I shall summarize Russellian monism's minimal commitments. (Section 2.4.1) Second, it will be argued that Russellian monism faces a *reductio*. (Section 2.4.2) Several possible objections are examined and rejected. (Section 2.4.3) If my argument works, the optimistic assessment of Russellian monism, which is now pervasive to philosophy of mind and metaphysics, should be seriously reconsidered.

2.4.1 The Basics of Russellian Monism

The long history of debates surrounding the Hard problem of consciousness made many philosophers to think that they meet the dead end. This pessimism has prompted many philosophers of mind to find radical alternatives, making theories of consciousness more fertile. Russellian monism, or "type-F monism" in Chalmers' terminology, is definitely such hypothesis which seems getting more and more intellectual fever these days. (Chalmers, 2010, p. 133-137) Yet even until now, what should be counted as minimal, basic elements of Russellian monism has not been clearly addressed. So I want to draw several essential commitments of Russellian monism first. Then some optimism and skepticism will be sketched.

Russellian monism starts by noticing a conceptual or epistemic limit of physics. Physics only can find out such-and-such structural or dispositional properties of the physical but cannot 'see through' what actually ground those properties. This idea originated from Russell's famous remarks of the nature of physics. In his *Analysis of Matter*, Russell states

> It is not always realized how exceedingly abstract is the information that theoretical physics has to give. It lays down certain fundamental equations which enable it to deal with the logical structure of events, while leaving it completely unknown what is the intrinsic character of the events that have the structure. We only know the intrinsic character of events when they

> happen to us. Nothing whatever in the theoretical physics enables us to say anything about intrinsic character about events elsewhere. They may be just like the events that happen to us, or they may be totally different in strictly unimaginable ways. All that physics gives us is certain equations giving abstract properties of their changes. But as to what it is that changes, what it changes from and to—as to this, physics is silent. (Russell, 1959, p. 17-18)

Russell's remark points out that what physics can give us about the physical world is merely abstract relations and nomic or causal profiles of fundamental entities. It does not, and may be cannot, teach us the nature of relata or what exactly those entities are. According to Russell, though physics can access to the *structure* and *dynamics* of the physical world, physics teaches us nothing about *what* is structured and *why* there are such dynamics. Taking this view on the nature of physics seriously is the central reason why all the variants of Russellian monism are called 'Russellian.' Sharing Russell's pessimism about physics should be taken as the first hallmarks of Russellian monism.

Russellian monism assumes that the unknown intrinsic properties of basic physical entities ground dispositional properties of those entities in a way that categorical properties ground dispositional ones.[17] These properties are not only unknown but *unknowable* by physical sciences in principle. Since the alleged properties are supposed to be intrinsic and ground dispositional

17) I intentionally simplified the situation, because many heavy issues are involved in characterizing Russellian monism. At least four distinctions should be noted: (i) extrinsic vs. intrinsic, (ii) dispositional vs. categorical, (iii) relational vs. non-relational, and (vi) structural-and-dynamic vs. non-structural-and-non-dynamic properties. As many of readers would have noticed, I implicitly have been assimilating all of them. There can be, and actually have been, a number of thorny metaphysical debates concerning these distinctions. Nonetheless, I just remain temporarily neutral on those issues and use all these distinctions more or less interchangeably in this dissertation. For some related points, see (Alter and Nagasawa, 2015, p. 427-432)

properties of fundamental physical entities, they are considered 'the intrinsic natures' of basic entities or "categorical grounds" of their dispositions. Due to their unknowability, those intrinsic natures or categorical grounds deserve to be called *inscrutables.*[18] The term roughly means that they cannot be 'read off' by physical sciences. Positing inscrutables, therefore, should be the second hallmark of Russellian monism.

Last, Russellian monism claims that inscrutables *'give rise to,' 'generate,'* or *'ground'* phenomenal properties of conscious experience. According to Russellian monism, inscrutables are responsible for phenomenal qualities we feel when we have conscious experience. There are a number of distinctions concerning what inscrutables really are and how they are responsible for our phenomenology. First, inscrutables can be either *phenomenal* or *protophenomenal.* Inscrutables can be phenomenal in such a way that our qualia are. Or, they can be protophenomenal in that they are not phenomenal in themselves but *jointly ground* the phenomenal qualities of higher order systems. Second, inscrutables can generate or ground phenomenal properties by *constitution* or *emergence*. Qualia may be

---

18) The term 'inscrutable' is, as far as I know, introduced in (Montero, 2014). It is used in (Alter and Nagasawa, 2015) and (Chalmers, 2015). Some might doubt that inscrutables are really necessary. Why not just satisfy with what physics, actually or possibly, teaches us? As stated in the quoted passage from Russell, fundamental physics seems to be 'abstract' in that it provides mere structures and dynamics of the physical reality. If the physical reality is like what fundamental physics describes, our physical world itself must be abstract. It would be a strange world where only relations and dispositions exist but no relata or grounds can be found. However, it is hard to believe that our world is like that. Whatever the term 'abstract' means, our physical world is apparently not abstract. Rather, it seems to be concrete in nature. Though the opposite view has been endorsed by some philosophers, it is natural to think that relations require relata and dispositions should be grounded. In other words, there must be something that "breathes fire into the equations [of any possible grand unified theory of physics] and makes a universe for them to describe". (Hawking, 1988, p. 174) Proponents of Russellian monism emphasize that this gap between abstract physics and concrete reality provides a good reason to posit inscrutables.

constituted by or emerged from inscrutables in certain organizations.[19] This generative or grounding role of inscrutables might be the most important and distinctive feature of Russellian monism. It must be taken as the third hallmark or Rusellian monism.

Given the three hallmarks of Russellian monism, Russellian monism can be summarized in conjunction of three claims. I argue that however Russellian monism is formulated, it must commit to the three theses arranged below[20]:

Structuralism about physics: the basic properties physics describes are relational/dispositional properties.
Realism about inscrutables: there are inscrutables, the natures of which are not wholly relational/dispositional.
Foundationalism about inscrutables: at least some inscrutables ground basic physical properties as well as phenomenal properties of experience.

---

19) These commitments lead to many possible versions of Russellian monism. As already noted by many, it is hasty to judge that Russellian monism is just a sophisticated version of crazy panpsychism, which distributes experiences like ours all over the physical universe. About 'the argument from weirdness', see (Alter and Nagasawa, 2015, p. 445-446). Indeed, for there are two possible natures of inscrutables, basically two types of Russellian monism are possible: if inscrutables are phenomenal in themselves, we have a *panpsychist* Russellian monism. If they are not, we have *panprotopsychist* Russellian monism. Further, according to the two possible ways how inscrutables give rise to or ground phenomenal properties, a *constitutive* Russellian monism and *emergent* Russellian monism. Therefore, there can be at least four versions of Russellian monism: panpsychist-constitutive, panpsychist-emergent, panprotopsychist-constitutive, panprotopsychist-emergent versions. While all these versions are interesting, I will not delve into each of them. The arguments I will develop later in this section does not depend on details of versions of Russellian monism.

20) While this formulation comes from (Alter and Nagasawa, 2015, p. 425), I substitute the original formulation's '(proto)phenomenal foundationalism' with 'foundationalism about inscrutables'.

Russellian monism is indeed "hot stuff" in recent consciousness studies and philosophy of mind. This intellectual fever suggests that many of philosophers see optimistic prospects in Russellian monism.[21] There seem to be at least two virtues of Russellian monism. Once properly construed in pan(proto)psychist form, Russellian monism might explain why there is such thing as phenomenal qualities at all. If (proto)phenomenal properties are already spread in the fundamental level of the physical world and qualia are constituted or emerged from the (proto)phenomenal properties, 'why' of consciousness will be solvable in principle. Another reason for pursuing the pan(proto)psychist version of Russellian monism is that it elegantly deals with the issue of mental causation. Inscrutables realize all the relational/dispositional properties in the fundamental level of the physical. And It has been strongly argued that dispositions can be causally relevant only by inheriting their categorical grounds' causal power. (Prior, Pargetter, & Jackson, 1982) If so, inscrutables are parts of the causal implementation of our world from which all the causal powers come. Once understood in this way, pan(proto)psychist Russellian monism arises as an attractive picture of how the phenomenal can have causal influence to the physical.

Moreover, a physicalist version of Russellian monism, *Russellian physicalism*, has been developed and discussed.[22] Russellian physicalism can

---

21) Holman states "The advertising for [Russellian monism] is that it constitutes just the insight needed to break (what many see as) the current impasse on the mind-body problem." (Holman, 2008, p. 49) Alter and Nagasawa agrees with Holman by saying "Many philosophers would agree that that result is both desirable and not delivered by traditional theories in the philosophy of mind." (Alter and Nagasawa, 2015, p. 448) Russellian monism appears to properly handle three problems that never be answered by other theories of consciousness so far: the Hard problem of consciousness, the problem of mental causation, and the conceivability argument.

22) Russellian physicalism has been recently developed by (Motenro, 2014). See also (Strawson, 2006; Papineau, 2002, p. 22-23; Pereboom, 2011). Russellian physicalism is a minimal physicalism in that it holds that phenomenal facts

be immune to the old zombie argument. According to the old zombie argument, PTI&~Q is conceivable. Proponents of Russellian physicalism argue that when we conceive the zombie world, one cannot help but ignore the fact about inscrutables, because they cannot be grasped even by complete physics. P is wholly constituted by truths about microphysical structures and dynamics and totally lacks truths about inscrutables. As it is widely acknowledged that there is no *a priori* entailment between microphysical and phenomenal truths, PTI&~Q is conceivable. From this conceivability, the metaphysical possibility of PTI&~Q follows. However, what such possibility implies at most is that some phenomenal facts do not supervene on the microphysical facts. The conceivability of zombies thus does not exclude the possibility that all phenomenal facts supervene on microphysical facts *plus* inscrutables facts. If there is any way to call inscrutables physical, it can be argued that phenomenal facts are fixed by physical facts. (Chalmers, 2015; Alter and Nagasawa, 2015) Thus, even if zombies are conceivable and conceivability entails possibility, the falsity of physicalism does not follow. At least a distinctive, nonorthodox sort of physicalism can survive. In this way, Russellian physicalism can assimilate both two central premises of the zombie argument, while maintaining a form of physicalism.

Russellian monism, however, is not immune to skeptical concerns. The immediate question is how inscrutables ground the phenomenal. Doubts go both ways of grounding. If constitutive Russellian monism is the case, it is hard to see how this sort of mental composition occurs. This is what

---

supervene on the physical facts. It is also a radical form of physicalism in that it assumes physical properties that cannot be revealed even by complete physics. For this reason, Russellian physicalism would be "a highly distinctive form of physicalism that has much in common with property dualism and that many physicalists will want to reject." (Chalmers, 2010, p. 152) If "there are physicalist versions of Russellian monism, they are nontraditional physicalist theories." (Alter and Nagasawa, 2015, p. 438)

usually called *the combination problem*. (James, 1890/1950; Seager, 1995; 2010; 2016b) Constitutive Russellian monism should provide explanations of how inscrutables constitute phenomenal qualities. When such explanation is provided, we have a genuine "mental chemistry". (Mill, 1848; Coleman, 2012) Unfortunately, no "mental chemistry" has been successful yet.[23] Even if we choose emergent Russellian monism, it is doubtable that phenomenal emergence can be any explanation of how phenomenal properties are generated. All of the issues described so far can be raised against the panprotopsychist version of Russellian monism. At any rate, Russellian monism's potential is doubtable and controversial at best.[24] Both optimism and pessimism raise various issues of current studies of Russellian monism. However, they are not my concern in this chapter. What I want to show is that no matter how it is construed, Russellian monism will face some troubles.

2.4.2 The Flipping Inscrutables

At first glance, Russellian monism seems to have no problematic implications. However, I think a deeper reflection on inscrutables' grounding of qualia would reveal its own problems. In the following, I will show that Russellian monism must allow something negatively inconceivable to be

---

23) Chalmers' recent work is comprehensive as well as instructive on this matter. (Chalmers, 2015)

24) One can also wonder what the nature of inscrutables is. This question is given to both versions of Russellian monism. For the panpsychist version, the claim that basic physical entities have phenomenal properties like us sounds so weird. Even Nagel expresses his doubt on such view by stating "Presumably the components out of which a point of view is constructed would not themselves have to have a point of view". (Nagel, 1979a, p. 194). For panprotopsychist version, there is always a risk of elusive otherism, the view that whatever generates consciousness, it would always be something other than what has been thought of as phenomenal. (Bourget, 2017)

negatively conceivable and turns out to be false. If my argument is on the right track, the optimism toward Russellian monism must be seriously reconsidered.

Before we start, it is worth noting that Russellian monism is compatible with the existence of immaterial souls. The basic commitments of Russellian monism are perfectly compatible with immaterial souls and non-physical causation. However, as explained in Section 2.3.3, if those immaterial supervene on microphysical (plus indexical) facts, Russellian monism cannot be compatible with the conceivability of PTI&~Q. In order for Russellian monism to be compatible with the conceivability of zombies, souls and non-physical causation should not supervene on microphysical (plus indexical) facts. Moreover, it seems clear that inscrutables and souls or non-physical causation are conceptually distinct. One can easily conceive of intrinsic properties that ground microphysical and phenomenal properties without conceiving souls or non-physical causation. And as far as I know, all Russellian monists are thinking about inscrutables without presupposing souls or non-physical causation. If souls and non-physical causation conceptually supervene on inscrutables, then it would require that all those Russellian monists be deeply irrational. There is no reason to accept this extremely implausible idea. So, it is safe to assume that souls and non-physical causation conceptually supervene neither on facts about microphysics nor on facts about inscrutables.

Then, let us suppose that inscrutables necessarily ground phenomenal properties.[25] And start with the example of Jane introduced in Section 2.3.3. Jane is a perfect physical duplicate of Mary without soul. The only difference is that in this case, Mary and Jane share not only microphysical

25) Although it is arguable that inscrutables ground phenomenal qualities contingently, metaphysical grounding relations are often assumed to be metaphysically necessary. In this dissertation, I will follow this widely accepted assumption.

structures and dispositions but also inscrutables. As explained in the previous paragraph, since facts about souls do not supervene on facts about microphysics and facts about inscrutables, even if Mary has a soul, it can be consistent to imagine her soulless inscrutable duplicate. When Jane sees the ripe tomato for the first time, in her brain, may be somewhere in V1, a particular group of neurons is activated. However, the neuronal group is not merely a neural correlate of red qualia. It is also the partial implementation of functional organization of Jane's brain. It plays certain causal roles in Jane's brain. By that activation, she says or does whatever those who first see a red thing would say or do.

And here comes the trick. Inscrutables of the neuronal group necessarily grounds Jane's red qualia. Basic particles that compose neurons of the group instantiate inscrutables, and these inscrutables are somehow organized to ground the red qualia. Let us call that inscrutable complex I. Next, a Cartesian demon invents a neuroprosthetic device. While it functions the same as Jane's original neuronal group, the device is fundamentally different in one respect: basic particles of the neuroprosthetic device play the same microphysical roles as basic particles of the neuronal group do. The only difference is that they instantiate completely different inscrutables. For example, electrons, which compose the neuronal group, and *schlectrons*, which make up the device, perfectly share their microphysical roles. They are microphysically indistinguishable in principle. In the device, the organization of the different inscrutables grounds different phenomenal properties. It grounds blue qualia rather than red ones. Let us call such inscrutable complex I*.

When Jane visually admires the color quality of her first-seen tomato, the Cartesian demon decides to replace Jane's neuronal group with his device. By the extremely covert and sophisticated way, he succeeds to unwittingly install his device in Jane's brain. When the demon turns on the switch in

his laboratory, the neuronal group in Jane's brain is suddenly replaced by the device. Through this procedure, the inscrutable complex of the neuronal group, I, is suddenly flipped to the inscrutable complex of the device, I*. By necessary phenomenal grounding, the initial red qualia in Jane's visual field suddenly turn into blue ones. In short, when the demon turns on the switch, Jane unexpectedly sees blue.

The question is what would happen in Jane's qualia-involved cognitive processes. Can the fresh phenomenal blue in her visual experience can be attended to or noticed by Jane? Here, as I defended in Section 2.3.3, cognitive processes require varieties of information processing, and information processing needs causal chains of traces. Since Jane is supposed to be soulless, if there is any causation at all, it must be a physical one. However, although I is substituted by I*, this procedure does not make any physical difference and change. The demon's procedure only makes changes or differences in inscrutables. As long as structuralism about physics holds, there cannot be any physical differences between the neuronal group and the device. After the demon turns on the switch, in all levels of the physical, every physical causal process would be preserved. Physical causal processes in Jane's brain must be fixed, and information processing cannot be initiated. Therefore, even though the color of the ripe tomato is brutally changed from red to blue 'in front of her eye', the newly acquired blue qualia cannot be paid attention to or noticed by Jane. All these absurdities can happen even when Jane's attention is abnormally sharpened or she is fully informed and readies for the demon's procedure. This scenario can be called *the flipping inscrutables.*

The flipping inscrutables scenario is a Russellian variant of dancing qualia. (Chalmers, 1996, p. 266-273) The crucial point is that Russellian monism entails that this flipping inscrutables scenario is at least coherent. Nothing in Russellian monism is incompatible with the scenario.

Structuralism about physics, realism about inscrutables, and foundationalism about inscrutables, the conceptual distinction between inscrutables and souls are consistent with the flipping inscrutables scenario. To the extent that necessary phenomenal grounding holds, Russellian monism entails that the flipping inscrutables scenario is consistent. If the scenario is consistent, there is no way for Russellian monism to rule out the scenario *a priori*. If it cannot be ruled out *a priori*, it is at least negatively conceivable. Therefore, Russellian monism entails the negative conceivability of the flipping inscrutables. It must be noted that I am not committing to any *modal* claim. All I argue for is a weak *epistemic* claim that the flipping inscrutables scenario must be at least negatively conceivable under Russellian monism.

Even if Russellian monism entails the negative conceivability of the flipping inscrutables scenario, the scenario makes no sense. The reason is cognitive intimacy. In the scenario, no matter how Jane is rational or alert, there is no way for the new blue qualia to be attended to or noticed by her. However, cognitive intimacy enforces that in the non-defective background, the blue qualia must be potentially attended to or noticed by Jane. For the flipping inscrutables scenario implies that there can be cognitively alienated qualia, the scenario turns out to be incoherent. Under the assumption of necessary phenomenal grounding, Russellian monism must claim that the loss of cognitive intimacy of Jane's newly acquired qualia is negatively conceivable. Nonetheless, such cognitively alienated qualia are incoherent and negatively inconceivable. By *reductio*, Russellian monism is wrong.

2.4.3 Objections and Replies

There may be many possible objections against the flipping inscrutables scenario. Every step of the argument might have a corresponding objection. In what follows, I will examine possible objections and show that none of

them works.

Objection 1: the flipping scenario assumes that microphysical structures and dispositions are multiply realizable. However, proponents of Russellian monism can argue that this is inconceivable. For example, basic particles' gravitational interactions allow only one sort of inscrutables as their ground. For an electron to pull another one, it must instantiate a particular kind of inscrutables. No other inscrutable can ground the electron's disposition to pull another. At the fundamental level, functions of Jane's neuronal group must be realized by I and only by I. I* cannot ground dispositions of the neuronal group's basic particles. If so, the invention of the neuroprosthetic device by the demon would be inconceivable.

Reply: this objection contradicts with Russellian monism's first commitments. Structuralism about physics states that properties of basic physical entities are relational and dispositional. Realism about inscrutables states that inscrutables are not relational and dispositional. Following these two commitments, one must conclude that there cannot be any conceptual connection from basic particles' relations or dispositions to their inscrutables. For instance, if inscrutables' grounding is *a priori* entailed by microphysical relations/dispositions, one would be able to read off which relational/dispositional properties are grounded by which inscrutables without any empirical information. One would 'see through' what basic particles do and find out which relations/dispositions are grounded by which inscrutables. Nevertheless, structuralism about physics and realism about inscrutables guarantee that there cannot be such *a priori* entailment. And if there is no *a priori* entailment, the Cartesian demon's invention of the device is conceivable in principle. We can conceive of multiple realizations of a functional property because truths about the realized functional property do

not entail truths of realizers. Likewise, one can conceive of multiple realizations of microphysical structures or dynamics, since truths about microphysical structures or dynamics entail nothing about inscrutables.

Objection 2: you have argued that when the switch turns on, there cannot be any new cognitive process because there cannot be any new physical change or difference. There is a change in qualia, however. Flipping I into I*, the demon changes the red qualia into blue ones. It is possible that this intrinsic, phenomenal change enables the blue qualia to be attended or noticed by Jane. Russellian monists can argue that qualia-involved cognitive processes may not depend on only structural and dynamics of physics. They may depend on intrinsic, phenomenal change either. If this is the case, the intrinsic and phenomenal change between the old and new qualia may suffice to make the new blue qualia cognitively intimate.

Reply: it is hard to see how such purely intrinsic, phenomenal changes make Jane to pay attention to or notice her new qualia. *Ex hypothesi*, since all structural and dynamical properties of microphysics are fixed, all physical causal chains should remain intact. So there cannot be any physical information processing. Also, Jane is soulless. Thus, there cannot be any physical or non-physical information processing involving Jane's new qualia. If any attending to or noticing the new qualia is possible at all, therefore, it must be due to some sort of non-causal epistemic relation to those qualia. However, I have argued in Section 2.3.4 that non-causal epistemic relations cannot explain how attending to or noticing qualia is possible, because they are necessarily constituted by attention to or notice of qualia. If so, there is no way for qualia to be potentially attended or noticed. That is, qualia-involved cognitive processes cannot depend on intrinsic, phenomenal change of qualia.

Objection 3: there is still a logical space for the newly acquired qualia to be potentially paid attention to or noticed by Jane. While the old red qualia are grounded by I, the newly acquired blue qualia are grounded by I*. This difference in grounding may initiate cognitive processes involving the new blue qualia. Though the new qualia cannot trigger new information processing at all, their being differently grounded by I* can somehow affect Jane's cognition. Then, Jane's newly acquired blue qualia can be cognitive intimate without new information processing.

Reply: be that as it may, there seems to be no way for the difference in grounding to bring cognitive processes involving the new qualia. There is a good analogy for this point. If Jane can attend to or notice her new blue qualia, she must be able to do so with her neuroprosthetic device. There is a strong analogy between them: both are given by the Cartesian demon's intervention. Both are newly acquired when the switch turns on. Most of all, both are grounded by I*. Thus, if Jane can pay attention to or notice her newly acquired blue qualia in virtue of the difference in grounding, she can do so with her newly acquired device.

Then, consider whether Jane can cognitively access to the device. Obviously, she cannot. Although there is a difference in grounding between Jane's neuronal group and the device, when the switch turns on, she cannot detect anything about the device. From the perspective of Jane, whether her neuronal group in her brain is replaced or not, it would not affect her cognition in the slightest. It follows that the device's being grounded by I* does not render the device potentially attended to or noticed. This cognitive failure strongly suggests that the difference in grounding cannot make the grounded things cognitively intimate. This result straightforwardly applies to the case of qualia: newly acquired qualia's being differently grounded by I*

would not make them to be potentially paid attention to or noticed by Jane.

Objection 4: cognitive intimacy is supposed to be true *a priori*. Therefore, Russellian monism must assimilate cognitive intimacy anyway. For instance, proponents of Russellian monists may take cognitive intimacy as their fourth commitment. So even if the flipping inscrutables scenario is negatively conceivable, Russellian monists would deny that it entails the loss of cognitive intimacy. They would insist that even when inscrutables are flipped, the newly acquired blue qualia must be potentially attended or noticed by Jane and that there is a way to explain how those qualia can be cognitively intimate.

Reply: the point of my argument is that there is no way for Russellian monism to assimilate cognitive intimacy. For the sake of argument, let us suppose that Jane's new blue qualia can be attended or noticed by Jane. The question is what *underlies* this cognitive intimacy of the new qualia. Logically, there are only two candidates: 1) the grounded new blue qualia; 2) the grounding inscrutable complex I*. In replying to Objection 2, I have argued that the new qualia themselves cannot enable any qualia-involved cognitive process. And my reply to Objection 3 shows that there is a good reason to think that I* cannot make Jane's new qualia cognitively intimate. Then, Jane's new qualia *cannot* be cognitively intimate under Russellian monism. If Jane's new qualia cannot be cognitively intimate under Russellian monism, Russellian monism should be false.

Qualia epiphenomenalism would want to account for cognitive intimacy. However, if the story of dull Jane and my replies to possible objections are right, they cannot. Qualia epiphenomenalism is thus wrong. Likewise, my argument and replies are intended to show that Russellian monists cannot account for cognitive intimacy, even if they want to. Even if Russellian

monism takes cognitive intimacy as their fourth commitment, my argument would show that Russellian monism is inherently inconsistent: the first three commitments undercut the fourth.

In this section, I have argued that Russellian monism is a seemingly promising but wrong hypothesis. The optimistic prospects of Russellian monism, therefore, should be critically reconsidered. In the next section, I will deal with the last disjunct of the consequence of the conceivability of zombies, interactionist dualism.

2.5 Interactionist Dualism and Swapped Psychons

Interactionist dualism claims that nonphysical entities actually exist and causally interact with physical entities. Since it argues for a possibility that some physical events might be caused by nonphysical phenomena, Interactionist dualism is incompatible with the causal closure of the physical and the completeness of the physics. Despite its unattractive appearance, in the debate concerning the zombie argument, interactionist dualism emerges as a consistent and even decent alternative to physicalism. Chalmers argued that the conceivability of zombies might have interactionist dualism as one of its possible consequences. (Chalmers, 1999; 2010)[26] If so, the *prima facie* consistency of interactionist dualism might lend some support to the conceivability of zombies. In this section, however, I shall argue that when it comes to qualia, interactionist dualism turns out to be false. To this end, first, in Section 2.5.1, how the conceivability of the zombie might be

26) While Chalmers reserves to accept interactionist dualism, he also says that it is "elegant and appealing and not obviously false." (Chalmers, 1999, p. 493) He further thinks "there is at least room for viable interactionism to be explored and that the most common objection to interactionism has little force. […] if we have independent reason to think that consciousness is irreducible, and if we wish to retain the intuitive view that consciousness plays a causal role, then this is a view to be taken very seriously." (Chalmers, 2010, p. 129-130)

compatible with interactionism is summarized. I shall argue in Section 2.5.2 that interactionist dualism is wrong in that it cannot help but allows an essentially negatively inconceivable scenario to be negatively conceivable. Then, possible objections will be considered and rejected If the argument in this section is sound, we can have not only traditional worries or complaints about interactionism but a new argument against it. As a result, the last position entailed by the conceivability of zombies can be rejected.

2.5.1 The Conceivability of The Gappy Zombie World

How interactionist dualism is entailed by the conceivability of zombies demands some clarification. Let us presume that one believes that interactionist dualism is true of the actual world. In order to conceive of the zombie world, all that she needs to do is subtracting just one of the phenomenal qualities in the actual world, while leaving all physical events intact. This imaginary subtraction will necessarily leave a "causal gap" somewhere in the conceived situation. This "causal gap" renders some physical events causally underdetermined, so that there will be certain physical events that are unexplainable. However, there seems to no bar to conceiving of such unexplainable physical events. (Chalmers, 2004. p. 184; 2010, p. 156) Therefore, interactionist dualism is compatible with the conceivability of zombies. According to interactionist dualism, the actual world can be described by the scheme below.

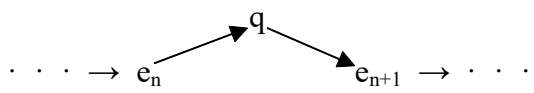

Figure 1

e_n and e_{n+1} represent physical events. q is an arbitrary phenomenal quality

of experience. The small arrows refer to physical causal chains to e_n or from e_{n+1}. The large ones represent psychophysical interactions between e_n, $e_{n+1,}$ and q. For example, e_n might be an activation of pain receptor. q might be an immaterial painfulness, and e_{n+1} might be an activation of reticular formation, which results in dopamine secretion, sending signals to prefrontal cortex, and motor planning involved in various avoidance behaviors. Dots after e_{n+1} may signify those following neural events. Given this scheme, the conceived zombie world under interactonist dualism might be described as follows.

$$\cdots \rightarrow e_n \quad (\quad) \quad e_{n+1} \rightarrow \cdots$$

Figure 2

The bracket between e_n and e_{n+1} represents the causal gap made by imaginary subtraction of q. Despite the causal gap, the full physical description of e_n and e_{n+1} and all the involved causal chains would be perfectly the same. Though in the zombie world e_n loses one of its effects and e_{n+1} loses one of its necessary causal conditions, this omission would not make any difference in the physics of the world, because the lost one, namely q, is immaterial. The remove of psychophysical interactions represented by large arrows would not bring any change to the physics, as they are psychophysical. If so, while the zombie world conceived under interactionist dualism would be causally gappy, such world is nonetheless conceivable in so far as it makes any sense.

2.5.2 The Swapped Psychons

In this section, after setting several points about the nature of interactionist dualism and causation, I will describe a scenario and argue that

interactionist dualism must allow that it is negatively conceivable.[27] However, such scenario is negatively inconceivable, so we have a *reductio* argument against interactionist dualism.

Two points must be noted. First, in order to be compatible with the conceivability of zombies, interactionist dualism must hold that psychophysical causation is *contingent.* If psychophysical causality is necessary, either the physical event in Figure 1, e_n, necessarily causes q or another physical event, e_{n+1}, must be caused by q. Either way, it would be impossible to imagine a situation in which e_n and e_{n+1} present but q is absent. For the former, since e_n is supposed to necessarily cause q, the situation where e_n presents but q_n is absent is unimaginable. For the latter, for e_{n+1} is assumed to be necessarily caused by q, imagining the situation where e_{n+1} occurs but q does not occur is impossible. As far as interactionist dualism is compatible with the conceivability of zombies, it must hold that psychophysical causation is contingent.

Second, even if interactionist dualism supposes immaterial souls, they should not supervene on microphysical facts. The reason is, again, its compatibility with the conceivability of zombies. I have explained in Section 2.3.3 that because of the 'that's-all' clause in PTI&~Q, the conceivability of zombies cannot allow any soul or non-physical causation. If interactionist dualism assumes that souls or non-physical causation supervene on microphysical facts, then it cannot assimilate the conceivability of PTI&~Q, since it cannot satisfy T. Insofar as interactionist dualism is supposed to be compatible with the conceivability of zombies, it must hold that souls or non-physical causations does not supervene on microphysical facts.

---

27) The term 'psychon' was coined by Eccles, who originally used the term to specify a special mental unit affecting neuronal activities in the brain. My use of the term in this thesis is much more liberal then Eccles'. I will use 'psychon' in order to refer to any immaterial or non-physical qualia that interactionist dualism commits to.

With these points in mind, let us picture a situation where psychons are *swapped.* The scenario goes as follows: by some quirk in prevailing laws of nature, a psychon in the actual world is swapped by another completely different psychon. Due to this swapping of psychons, whatever had been directly or indirectly caused by the psychon is instead caused by that different psychon. Also, there are no immaterial souls. The immaterial pain in the actual world, for instance, is swapped by immaterial pleasure in w_i. The situation of w_i thus can be described by the figure below.

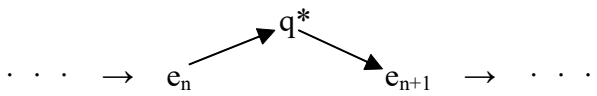

Figure 3

w_i is a *supermasochistic* world. In the actual world, the activation of pain receptors causes the intense pain q, and the pain sends signals to prefrontal cortex, which cause motor planning and finally lead to complex contractions of muscle tissues in one's limb. After the swapping, however, the same activation of pain receptors brings an extreme pleasure q*. And q* causes the signal sending to prefrontal cortex, motor planning, etc. And there is no soul or non-physical causation. This is conceivable because souls are supposed not to supervene on anything physical. A person who have felt the intense pain, say Jane, now loses her soul and feels the extreme pleasure. However, the extreme pleasure does not make any physical difference. Jane' brain operates the same, and she still shows pain behaviors. This is all that *the swapped psychons scenario* is about. It is an interactionist version of inverted qualia.

Then, can the swapped psychon, q*, be attended to or noticed by Jane? It cannot. There is no way for q* to be attended to or noticed by Jane,

since even after the swapping, there are no causal differences. *Ex hypothesi*, there are no physical differences. There cannot be any non-physical differences either, since there is no non-physical entity. Physically and non-physically, there is no difference at all. And I have shown that if there is no difference, there is no cognitive process either. This implies that even if Jane feels the extreme pleasure, she cannot pay attend to or notice that feeling. This is so even when Jane is perfectly rational or alert. In a nutshell, Jane's new pleasure is cognitively alienated from her.

Is this swapped psychons scenario negatively conceivable under interactionist dualism? In other words, is the scenario not ruled out *a priori* by interactionist dualism? Apparently, it is not. In conceiving the scenario, one should do only two things: (1) to imaginarily distort psychophysical causal chains; (2) to subtract any immaterial souls (if they exist). Since interactionist dualism must hold that psychophysical causality is contingent, (1) is clearly conceivable. (2) is also conceivable, since it is assumed that soul or anything immaterial does not supervene on the physical. Interactionist dualism is compatible with both (1) and (2). As the swapped psychons scenario is coherent under interactionist dualism, it is at least negatively conceivable. Therefore, once interactionist dualism is compatible with the conceivability of the zombie world, there is no way for interactionist dualism to deny that the inert psychons are negatively conceivable. It cannot help but entails that the scenario is at least negatively conceivable.

However, is the swapped psychons scenario really negatively conceivable? The whole point boils down to whether the swapped but cognitively alienated qualia are negatively conceivable. We have already seen that such case is not even negatively conceivable. In Section 2.3 and 2.4, I have argued that when qualia cannot make any change and difference, they cannot be informationally processed and cannot be attended to and noticed

by subjects of experience. However, cognitive intimacy does not allow qualia to be cognitively alienated, so that scenarios which entail such possibility must be rejected. The swapped psychons scenario is also exposed to the same *reductio* argument. Consider Figure 3. q* cannot be cognitively intimate. However, this cannot be the case, if cognitive intimacy is true *a priori*. Therefore, the swapped psychons scenario is incoherent and negatively inconceivable. The argument so far raises a *reductio* against interactionist dualism: interactionist dualism entails that the swapped psychons scenario is negatively conceivable. If my argument is right, it is not even negatively conceivable that psychons are swapped. By *reductio*, interactionist dualism turns out to be wrong.

2.5.3 Objections and Replies

There might be possible objections. Objections may come from largely two directions: denying the negative conceivability of the swapped psychons scenario or arguing for the cognitive intimacy of inert psychons. I will examine four of such objections and show that none of them is successful.

Objection 1: it can be argued that qualia-involved cognitive processes are so special and unique that they should be treated as exceptions of information processing theory of cognition. Proponents of interactionist dualism may claim that in attending to or noticing q*, no set of information processing is required. It might be the case that qualia can be attended or noticed through a cognitively special way. For instance, interactionist dualism can accept some sort of non-causal epistemic relation. If so, q* can be cognitively intimate under interactionist dualism.

Reply: it is not clear at all that any non-causal epistemic relation can help anything here. In Section 2.3.4 and 2.3.5, I have shown that why no

non-causal epistemic relation can make qualia cognitively intimate. The cognitive structure of non-causal epistemic relation undercuts the possibility that qualia-involved cognitive processes will be explained in terms of non-causal epistemic relation. While there are overwhelming reasons to doubt that non-causal epistemic relations enable qualia-involved cognitive processes, reasons to accept such relations are hard to find. I think one can safely assume that appealing to non-causal epistemic relation cannot help interactionist dualism.

Objection 3: if your argument is right, even physicalism would be exposed to the same *reductio*. In order to conceive the swapped psychons, the only thing one need to do is to accept that psychophysical causation contingent. And as you rightly pointed out in replying to Objection 1, interactionist dualists cannot hold that psychophysical causation is necessary. Insofar as physicalists hold that psychophysical causation is contingent, they cannot help but admit that the swapped psychons scenario is negatively conceivable at least. Then, physicalism would turn out to be wrong. In effect, if your *reductio* argument is valid, every philosophical position that assumes the contingency of psychophysical causation would turn out to be false. Of course, this is not the case.

Reply: the swapped psychons scenario is not even negatively conceivable, when physicalism is true. The minimal necessary condition for physicalism is mind-body supervenience. Therefore, if physicalism is true, psychons must supervene on something physical, regardless of whether the physical world is causally closed or not. There must be a physical base p on which q supervenes. Since q supervenes on p, e_n can cause q only by causing p. For the same reason, in order for e_n to cause q*, it must cause p* which is a physical supervenience base of q*. However, when e_n causes p* instead of

p, it cannot be e_n anymore, for physical events are individuated by their *causal profiles*. As Chalmers emphasizes, physical entities are structural and dynamical in nature: they are defined by how they interact with other properties or conditions in particular ways. For instance, what it is to have a certain mass is to stand in certain law-like relations to other entities, such as gravity. If so, physical events must be individuated by there law-like relations to other physical events. e_n is e_n because it is causally related with p. Once it causes a different effect, namely p*, e_n must be individuated differently. For instance, e_n is individuated as the activation of pain receptor because of its causal relation to C-fiber firing. If e_n is causally related to a different event, D-fiber firing, it should be individuated as something different, the activation of pleasure receptor. So once physicalism is assumed, to change psychophysical causations while leaving physical events intact is impossible. Changing the psychophysical causal chain from e_n to q entails changing the physical causal chain from e_n to p, and once the physical causal chain from e_n to p is changed, e_n cannot remain fixed as it is. Therefore, the swapped psychons scenario is inconceivable under physicalism.

2.6 The Inconceivability of Zombies

Let us take stock. The conceivability of zombies entails the disjunction of three different theses: qualia epiphenomenalism, Russellian monism, and interactionist dualism. In order to refute the conceivability of zombies, I had to show that all of the three disjuncts are false. From Section 2.3 to 2.5, I have provided my *reductio* arguments against each of them. It is easy to notice that the same pattern is repeated in all my *reductio* arguments. All three positions entailed by the conceivability of zombies entails the negative conceivability of cognitively alienated qualia. Cognitive intimacy, nonetheless, blocks this conceivability.

Why do all the positions suffer the same pattern of *reductio*? The reason seems to be that they all neglect the issue of cognitive intimacy. I think this ignorance can be partially explained by the way how the Hard problem of consciousness is raised. The Hard problem always asking why there is such thing as consciousness at all. Sometimes it asks how physical processes can generate experience. These questions are essentially about *existence* of consciousness. Accordingly, when philosophers are engaging in the debate concerning consciousness, they tend to focus on metaphysical questions: is consciousness identical to physical process? Do phenomenal facts supervene on physical facts? Can qualia be functionalized? The Hard problem of consciousness is centered on *the metaphysical nature* of consciousness. Nonetheless, there are crucial issues of consciousness other than the Hard problem. For example, is there any consciousness that cognitively insulated from subjects? Does attention supervene on to phenomenal aspects of our experience? Can phenomenal qualities of our experience be unnoticeable in principle? These are the questions of *the cognitive essence* of consciousness. Some philosophers were overly obsessed with the metaphysical nature of phenomenology that they missed the cognitive essence of it. None of qualia epiphenomenalism, Russellian monism, and interactionist dualism attempt to take account of the cognitive intimacy of consciousness. Since all of these positions neglect the cognitive essence of phenomenology, it is likely that they ignore cognitive intimacy either. Conversely, once we turn our attention to the issues of cognitive intimacy, I believe that many would seriously doubt the conceivability of zombies.

Chapter 3
Conceivability and Possibility

3.1 Chapter Introduction

In this chapter, I shall move on to the second premise of the zombie argument, which states that ideal positive primary conceivability of PTI&~Q entails the primary possibility of PTI&~Q. Whereas CP- claims that ideal negative primary conceivability entails primary possibility, CP+ states that ideal positive primary conceivability entails primary possibility. (Chalmers, 2010) The second premise of the zombie argument is an application of CP+. Recently, some type-B materialists suggest anti-zombie arguments that parody the zombie argument. Anti-zombie arguments draw a paradoxical conclusion that if CP+ is right, CP+ is wrong. However, in Section 3.2, I shall argue that all the anti-zombie arguments previously suggested fail. The failure of the previous attempts suggests that to be a precise parody, anti-zombie arguments must reflect the ideal positive primary conceivability and the possibility of Russellian monism. Taking this point into account, I will provide a Russellian version of anti-zombie arguments. If the argument succeeds, CP+ can be rejected.

3.2 The Russellian Illuminati Argument and the Conceivability-Possibility Entailment

Many philosophers have been focused on the second premise of the zombie argument: if PTI&~Q is conceivable, is primarily possible. This premise is an application of a general principle linking conceivability and modality. Chalmers(2002) argues that if a statement S is ideally positively primarily conceivable, S is primarily possible. This thesis is called CP+. (Chalmers, 2010, p.147) Despite numerous attempts to refute CP+, Chalmers believes

that all those criticisms fail. The central reason is that there seems to be no convincing counterexample against CP+. "[T]here have been many attempts at providing counterexamples to CP+, but none of these provides clear counterexamples." (ibid., p. 180) I disagree. A number of physicalists have suggested that if zombies are conceivable, a conscious creature which satisfies physicalist description is also conceivable. I shall argue that from such conceivability, one can develop a counterargument against CP+. In what follows, I will argue that mentioned cases can be considered as counterexamples against CP+. First, by examining 'the anti-zombie arguments' provided by those philosophers who prefer physicalism, I will point out that all of them share the same problems. (Section 3.2.1) Then, my versions of two-dimensionally articulated anti-zombie arguments will be suggested, and actual and possible objections will be replied. (Section 3.2.2 and 3.2.3)

3.2.1 Anti-zombie arguments

Impressed by the force of the zombie argument, a group of philosophers has been attempted to show that the same move can be made to argue for physicalism. Their arguments share common features. First, by applying CP+ or something similar to CP+, they suggest their own zombie arguments which show that the physicalism is wrong or dualism is right. Next, by applying CP+ or something like CP+ again, they parody the zombie arguments. They appeal to a conceivability of creatures that are indistinguishable from us in every physical and even conscious aspect. The crucial twist is that even though these creatures are identical with us, they have nothing over and above their physical properties. These imaginary creatures are wholly physical and do not have any non-physical consciousness. Hence, to conceive such creatures is tantamount to conceive purely physical consciousness. These creatures have many names, including

“anti-zombies” (Frankish, 2007; 2012) or “zoombies” or “shombies” (Brown, 2010; 2013) The parody arguments draw the secondary (metaphysical) possibility of such creatures. If so, physicalism is right and dualism is wrong. Combining the original and parody arguments, one can draw a contradiction: physicalism is right and wrong. Or, dualism is wrong and right. Since other premises seem OK, by *reductio*, the premise of CP+ must be rejected. This is what *the anti-zombie arguments* are about.

There have been many varieties of the anti-zombie arguments. The first one was presented by Keith Frankish. (Frankish, 2007) As a target argument, he presents a version of the zombie arguments: (1) zombies are conceivable. (2) If zombies are conceivable, then zombies are possible. (3) If zombies are possible, then consciousness is not physical. (4) So consciousness is not physical. (ibid., p. 652) Then, Frankish suggests his version of the anti-zombie argument by simply replacing ‘zombies’ with ‘anti-zombies’.

(A1) Anti-zombies are conceivable
(A2) If anti-zombies are conceivable, then anti-zombies are possible
(A3) If anti-zombies are possible, then consciousness is physical
(A4) So consciousness is physical.

The parody arguments yield a contradiction. Premise (2) or (A2) is suspicious. Both are applications of the principle that conceivability entails possibility. Thus, it seems that we have a good reason to believe that even if something is conceivable, it does not mean that it is possible.

Frankish’s argument has received much attention. In the current context, however, its problems are obvious. First, the target and parody arguments do not correctly apply CP+. CP+ states that if a statement is ideally, primarily, and positively conceivable, it is primarily possible. However, premise (2)

and (A2) state that if zombies or anti-zombies are conceivable, then they are possible. They do not reflect CP+. They are link conceivability *simpliciter* with metaphysical possibility. Frankish's anti-zombie argument cannot provide a *reductio* against CP+. His version of the anti-zombie argument is irrelevant to countering CP+.

Second, even if Frankish revises his argument by implementing CP+, the *reductio* against CP+ does not follow. To draw secondary possibility of anti-zombies, he must show that anti-zombies are ideally, positively, and primarily conceivable and that their primary possibility entails their secondary possibility. The problem is that Frankish does not provide any argument for these claims. He ignores the two-dimensional structure of the zombie argument and sticks to his own anti-zombie arguments. Without showing the positive conceivability and the secondary possibility of anti-zombies, Frankish's argument cannot conclude that physicalism is right. In turn, it fails to draw a contradiction and raise a *reductio* against CP+.

Even if Frankish sets his argument in the two-dimensional framework and argues for the positive conceivability and secondary possibility of anti-zombies, it still falls short of raising the *reductio* against CP+. Frankish's argument neglects the possibility of Russellian monism. To complete the *reductio*, the anti-zombie argument should yield a contradiction with the conclusion of the zombie argument. The conclusion of the zombie argument, however, is not merely that physicalism is wrong. It is that physicalism is wrong *or* Russellian monism is the case. To draw a contradiction, Frankish must show that physicalism is right *and* Russellian monism is not the case. Unless Russellian monism is rejected, the anti-zombie argument fails to draw a contradiction and complete the *reductio*.

There is another version of anti-zombie arguments. Recently, Richard Brown(2013) developed his own anti-zombie argument. Brown rightly points

out that like physicalism, dualism itself implies a modal claim, namely □(Q ⊃~PT). This modal claim can be interpreted as 'necessarily, if there is a certain phenomenal quality, it is not the case that everything that exists is physical'. Although Brown does not provide his target argument, we can easily reconstruct it based on the modal claim: (1*) □(Q⊃~PT) is negatively conceivable. (2*) If □(Q⊃~PT) is negatively conceivable, then □(Q→~PT) is primarily possible. (3*) If □(Q⊃~PT) is primarily possible, then □(Q⊃~PT) is secondarily possible. (4*) If □(Q⊃~PT) is secondarily possible, then dualism is true. (5*) Dualism is true. Brown's parody of this dualist conceivability argument runs as follows:

(S1) PT&Q is negatively conceivable.
(S2) If PT&Q is negatively conceivable, then PT&Q is primarily possible.
(S3) If PT&Q is primarily possible, then PT&Q is secondarily possible.
(S4) If PT&Q is secondarily possible, then dualism is false.
(S5) Dualism is false. (ibid., p. 2)

PT&Q is incompatible with □(Q⊃~PT). PT&Q can be roughly construed as 'everything that exists is physical and there is a phenomenal quality'. Brown(2013) calls creatures living in a world where PT&Q holds shombies. While shombies are purely and wholly physical, they are conscious. From the negative conceivability of PT&Q or shombies, Brown draws a contradiction between (5*) and (S5). As in Frankish's argument, here one can reject (2*) or (S2) by *reductio*. Both state that if something is negatively conceivable, it is also primarily possible.

While Brown's shombie argument reflects two-dimensional structures, it nonetheless faces similar problems with Frankish's. First, the shombie argument is irrelevant to countering CP+. The shombie argument does not involve the positive conceivability of PT&Q or shombies. It relies on the

principle that if something is negatively conceivable, it is primarily possible. This has nothing to do with CP+. Rather, such principle is more akin to CP-. That is, the shombie can be a *reductio* against CP- at best.

Moreover, the shombie argument fails to draw the secondary possibility of □(Q⊃~PT) or PT&Q, as it ignores the possibility of Russellian monism. Russellian monism and premises (3*) and (S3) are incompatible. Both premises can be justified only when primary and secondary intentions of P coincide. Nevertheless, Russellian monism claims that they are distinct. To see why, one must understand the semantics of microphysical term in the two-dimensional framework. It is usually assumed that primary and secondary intensions of microphysical term coincide. Microphysical terms, such as 'mass', 'charge', and 'spin', are *theoretical terms*. Their definitions are given by theoretical or causal roles they play. 'Charge' is defined as properties that play charge-roles. It appears intuitive that whatever satisfies the definition deserves to be called charge. Even if a certain alien property in Twin Earth occupies all the charge-roles, many would call the property charge. In this sense, primary and secondary intentions of microphysical terms are the same: in worlds considered as actual, they refer to whatever satisfies theoretical definitions. In worlds considered as counterfactual, they still pick out the same thing. When an expression's primary and secondary intentions coincide, the expression is called *semantically neutral.* (Chalmers, 2006)

Russellian monism denies the semantic neutrality of microphysical terms. According to Russellian monism, while primary intensions of microphysical terms are *a priori* given by theoretically defined roles, secondary intensions are fixed by inscrutables that actually play those roles. As explained in Section 2.4.1, microphysics captures only dispositional or structural properties of microphysical entities in the fundamental level. It does not tell us what intrinsic natures of microphysical entities ground those dispositions or

structures. Russellian monism assumes that actual inscrutables ground structures and dynamics. As the primary intention of ‘water’ is given by water-roles, Russellian monism claims that the primary intentions of microphysical terms are given by theoretical roles. However, as the secondary intension of ‘water’ is fixed by the chemical property that actually plays water-roles, the secondary intension of microphysical terms should be fixed by the inscrutables which actually play theoretical roles. That is, Russellian Monism treats microphysical terms as some sort of natural kind terms. Russellian monism’s semantic distinction of primary and secondary intensions of microphysical terms is rooted in its metaphysical distinction of intrinsic and dispositional/structural properties of microphysical entities.

This Russellian monism’s semantic of microphysical terms immediately affects the shombie argument. Microphysical terms are not semantically neutral, so that Russellian monism denies the semantic neutrality of microphysical truths P. Then, if Russellian monism is the case, even if a statement involving P is primarily possible, its secondary possibility does not follow. Indeed, in the zombie argument, from the primary possibility of PTI&~Q, Chalmers does not draw only its secondary possibility. He infers that PTI&~Q is secondarily possible or Russellian monism is true. This is the reason why the zombie argument concludes that physicalism is wrong or Russellian monism is true. (Chalmers, 2010) Thus, if P is not semantically neutral, there is no way to argue for (S3). Even if PT&Q is primarily possible, it does not follow that it is secondarily possible. This is why Brown’s parody argument fails: Russellian monism and (S3) are incompatible. Brown can argue for (S3) only when he successfully rules out the possibility of Russellian monism. As far as I can tell, however, he never provides any counterargument against Russellian monism.

Balog also provides an anti-zombie argument.[28] Her parody conceivability

argument appeals to a conceivability of illuminati. Illuminati, according to Balog, are "purely physical creatures that are our physical duplicates and enjoy phenomenal experiences". (Blog, p. 16) Illuminati are very similar to Frankish's anti-zombies and Brown's shombies. Balog knows well about other anti-zombie arguments and is clearly aware of how her illuminati argument would work as a *reductio* against CP+.[29] "The point is not to take the argument seriously as a positive argument. Rather, it is meant to be a *reductio* of Chalmers' principle connecting conceivability and modality that underlies both the CP^{pos} Principle". (ibid., p. 25) Balog's illuminati argument can be summarized as follows:

(I1) $APc^{pos}(P\&Q\&\Box(p=q))$
(I2) $AP(c^{pos}(P\&Q\&\Box(p=q))\supset\Diamond(P\&Q\&\Box(p=q)))$
(I3) $AP(\Diamond(P\&Q\&\Box(p=q))\supset\sim\Diamond(P\&\sim Q))$
(I4) $AP\sim\Diamond(P\&\sim Q)$
(I5) $AP\sim\Diamond(P\&\sim Q)\supset\sim c^{neg}(P\&\sim Q)$
(I6) $\sim c^{neg}(P\&\sim Q)$

Here, AP is an *a priori* operator, a shorthand of 'it is *a priori* that'. C^{pos} and C^{neg} are conceivability operators, respectively representing 'it is positively conceivable that' and 'it is ne gatively conceivable that'. $\Diamond$ means that 'it is metaphysically possible that'. q is an arbitrary phenomenal term, and p is a microphysical term.

The inference from (I1) to (I3) is straightforward. (I4) needs some explanations. Why is P&~Q metaphysically impossible when p=q is metaphysically possible? This is because p and q are assumed to be rigid

---

28) Three versions of Balog's manuscript are circulated online. The one used in this dissertation is from http://www.philosophy.rutgers.edu/joomlatools-files/docman-files/Balog%20paper.pdf.
29) See (Balog, p. 23, fn60)

designators. Their referents are fixed by what they refer to in the actual world. In every counterfactual world, they pick out whatever they pick out in the actual world. For instance, 'pain' would refer to painfulness in every counterfactual world where painfulness exists. 'Spin' would pick out properties playing spin-roles in all counterfactual worlds where such properties are. If so, when p=q holds in one counterfactual world, it must be true in all counterfactual worlds. Two-dimensionally put, when p=q is secondarily possible, p=q is secondarily necessary. As P includes the truth about p and Q is the truth about q, it implies that there cannot be a counterfactual world where P&~Q holds. All the inferences so far are *a priori*, so that we have (I4).

What about (I6)? It says that if it is *a priori* that P&~Q is not possible, P&~Q is negatively inconceivable. This seems to be true by the notion of negative conceivability. Remind that a statement is negatively conceivable when it cannot be ruled out *a priori*. (Chalmers, 2002) By contraposition, when a statement is ruled out by *a priori* reasoning, it is not negatively conceivable. The antecedent of (I5) shows that P&~Q is impossible on *a priori* ground, and the consequent says that it is negatively conceivable. Thus, P&~Q is negatively inconceivable. And this yields a contradiction. Obviously, P&~Q is negatively conceivable. The zombie argument supposes that P&~Q is ideally, positively, and primarily conceivable, and positive conceivability clearly entails negative conceivability. (ibid.) If so, as (I6) yields a contradictory consequence, (I2) must be rejected by *reductio*.

Although the illuminati argument is better than other anti-zombie arguments in distinguishing positive and negative conceivability, it has a number of problems. Most of all, it does not involve CP+. (I2) states that the positive conceivability of P&Q&□(p=q) entails the metaphysical possibility of it. It is not an application of CP+. Moreover, Blog's argument applies CP+ to a *wrong conceivability*. Though the argument rightly captures

the positive conceivability, it does not concern ideal or primary conceivability. If so, strictly speaking, the argument cannot apply CP+ to (I1). The argument also deals with a *wrong statement.* In (I2), the illuminati argument applies CP+ to the partially modal statement, P&Q&□(p=q). Chalmers, however, may reject this move by restricting CP+ only to non-modal statements. Indeed, he thinks that this restriction can be done without being *ad hoc* (Chalmers, 2010, p. 179). If the application of CP+ in (I2) is blocked, the illuminati argument cannot be sound.

The most serious weakness of the illuminati argument is that there is no argument for the central premise of the argument, the positive conceivability of P&Q&□(p=q). Balog merely mentions that "there is reason to think that the conceivability^{pos} of zombies and the conceivability^{pos} of illuminati are on a par. […] Both are equally *prima facie* conceivable^{pos}, due precisely to the direct and substantial grasp of phenomenal properties that phenomenal concepts afford us." (Balog, p. 23) *Prima facie* conceivability nonetheless has nothing to do with positive conceivability. Unless Balog provides any clear reason to think that P&Q&□(p=q) is positively conceivable, she cannot successfully finish the *reductio* against CP+.

All these considerations lead to the conclusion that all of the anti-zombie arguments miss the target. On the one hand, anti-zombie arguments do not concern the positive conceivability of anti-zombies. As CP+ is essentially about the relationship between positive conceivability and primary possibility, when anti-zombie arguments do not involve the positive conceivability of anti-zombies, they cannot have any bearing on CP+. They can be *reductio* arguments against CP- at best. Chalmers seems to know this problem already. Against all attempts to refute CP+, He says that they "seem to work best as challenges to CP- rather than to CP+, so that CP+, which is all that is required for the argument against materialism, is relatively unthreatened." (Chalmers, 2010, p. 160) On the other hand, all the

anti-zombie arguments do not take account of the possibility of Russellian monism. Russellian monism involves whether the primary possibility of anti-zombies entails the secondary possibility of anti-zombies. However, none of the anti-zombie arguments concern Russellian monism. They simply ignore the issue of semantic neutrality of P or Q. The anti-zombie arguments cannot do what they supposed to do, until these loopholes are fixed.

Therefore, if one wants to develop a successful anti-zombie argument, one must do two things: first and foremost, one must build his or her argument on the notion of positive conceivability. Second, the semantic neutrality and Russellian monism must be taken seriously. Anti-zombie arguments are essentially intended to provide a *reductio* against CP+ by appealing to the semantic neutrality of P and Q. In other words, one must secure the semantic neutrality of phenomenal and microphysical terms. In the next section, I will suggest my version of anti-zombie arguments that is not plagued by these problems.

3.2.2 The Russellian Illuminati Argument

My version of anti-zombie argument starts by introducing a physicalist version of Russellian monism. As explained in the previous section, Russellian monism may assume that intrinsic properties which ground microphysical dispositions and structures are (proto)phenomenal. If intrinsic properties, or inscrutables, are protophenomenal, they can be assumed to be physical. Dispositional or structural properties at the fundamental level of the physical can be grounded by physical inscrutables. Indeed, Chalmers suggests that physical properties as dispositional or structural properties might be called *narrowly physical,* but physical properties as intrinsic and inscrutable properties can be labeled *broadly physical* (Chalmers, 2015). We should allow that inscrutables can be at least broadly physical. I have

already mentioned that there are various versions of Russellian physicalism. Then, narrowly physical properties are grounded by broadly physical inscrutables. I will call this version of Russellian monism *type-B Russellian Physicalism*. In short, type-B Russellian monism is a type-B materialist version of Russellian monism which assumes broadly physical inscrutables.[30)]

Against the zombie argument, type-B materialists, physical truths do not entail *a priori* phenomenal truths. The physical and the phenomenal are conceptually distinct, so that it is possible to conceive one without another. There remains the so-called 'epistemic gap'. A group of type-B materialists, however, presupposes an *identity* between phenomenal and physical properties. This identity cannot be known *a priori* as the identity between H_2O and water cannot be known *a priori*. There is nonetheless an essential difference between these two kinds of identities: if one knows everything about H_2O and the concept of water, it seems that she can be in a position to *deduce* that H_2O is identical to water. Type-B materialists argue that such deduction is impossible in the phenomenal-physical identity. Even when one knows everything about physics and other truths, she cannot deduce that a certain physical process is consciousness. In other words, the identity is not entailed *a priori* by physical truths or PTI. For this absence of *a priori* entailment, there cannot be any *transparent* and *reductive* explanation between phenomenal qualities and physical processes. The phenomenal-physical identities are unique and *epistemically primitive* in this sense. (Chalmers, 2010; Chalmers and Jackson, 2001)

Type-B Russellian physucalism inherits many features of type-B materialism. It claims that phenomenal properties are conceptually distinct from complexes of broadly physical inscrutables. There is no *a priori* entailment from truths about broadly physical inscruatables to truths about

30) Chalmers already drew a similar distinction between "type-A constitutive panpsychism" and "type-B constitutive panpsychism". (Chalmers, 2015, p. 25)

phenomenal properties. There would be the epistemic gap between them. Further, type-B Russellian physicalism argues that although phenomenal properties are identical to a complex of broadly physical inscrutables, the identity is not entailed *a priori* by truths about broadly physical inscrutables. If so, why and how phenomenal properties are complexes of broadly physical inscrutables cannot be explained transparently and reductively. Like phenomenal-physical identities in type-B materialism, identities between phenomenal qualities and complexes of broadly physical inscrutables are epistemically primitive.

Then, my anti-zombie argument can be formalized as follows:

(R1) $c^{ipp}(P_iQTI\&(p_i=q))$
(R2) $c^{ipp}(P_iQTI\&(p_i=q))\supset\Diamond^1(P_iQTI\&(p_i=q))$
(R3) $\Diamond^1(P_iQTI\&(p_i=q)\supset\Diamond^2(P_iQTI\&(p_i=q))$
(R4) $\Diamond^2(P_iQTI\&(p_i=q))\supset\sim\Diamond^2(P_iQTI\&\sim Q)$
(R5) $\sim\Diamond^2(P_iTI\&\sim Q)$
(R6) $c^{ipp}(P_iTI\&\sim Q)$
(R7) $c^{ipp}(P_iTI\&\sim Q)\supset\Diamond^1(P_iTI\&\sim Q)$
(R8) $\Diamond^1(P_iTI\&\sim Q)\supset\Diamond^2(P_iTI\&\sim Q)$
(R9) $\Diamond^2(P_iTI\&\sim Q)$
(R10) (R2) or (R7) is wrong. Either way, CP+ is false.

The argument above can be called *the Russellian illuminati argument*. I introduced new terminology: c^{ipp} in (R1) and (R2) is a conceivability operator, which means 'It is ideally, positively, primarily conceivable that'. The term p and P in Balog's illuminati argument are replaced by the term p_i and P_i. p_i is an inscrutable term for an arbitrary complex of physical inscruatables. Let us call it a *complex physical inscrutable*. On the other hand, whereas P in the illuminati argument refers to the conjunction of all

physical truths, P_i in the Russellian illuminati argument represents a conjunction of all inscrutable truths. When every microphysical term in P is replaced by its corresponding inscrutable term, we have P_i. As P includes the truth about p, P_i includes a truth about p_i as one of its conjuncts. For insrutables are broadly physical in type-B Russellian physicalism, P_i can be considered as a conjunction of all *broadly physical truths*. In Chalmers' terms, P is the conjunction of narrowly physical truths, but P_i is that of broadly physical truths. $\Diamond^1$ and $\Diamond^2$ are two-dimensional modal operators that respectively represent 'It is primarily possible that' and 'It is secondarily possible that'.

Whereas the original argument identifies a phenomenal property with a microphysical property, my argument makes the same identification with a complex physical inscrutable. As all versions of anti-zombie arguments assume purely physical creatures with consciousness, type-B Russellian physicalism supposes that consciousness itself is somehow identical to a complex physical property. Both allow purely physical consciousness and deny non-physical consciousness. (R1) claims that $P_iQTI\&(p_i=q)$ is ideally positively primarily conceivable. (R2) is an application of CP+. (R3) states that $P_iQTI\&(p_i=q)$ is primarily possible, it is secondarily possible. The inference from (R4) to (R5) can be justified analogously to the inference from (I4) to (I5) in the illuminati argument. (R6) says that $P_iTI\&\sim Q$ is ideally, positively, primarily conceivable. From CP+ and the semantic neutrality of P_i and Q, the inference from (R6) to (R9) is straightforward. (R9) contradicts with (R5). Since there seem to be no problems in other premises, it is either (R2) or (R7) that should be rejected. Both are applications of CP+. Therefore, CP+ must be false. Given CP+ and the semantic neutrality of the terms, inferences from (R4) to (R5) and from (R7) to (R9) seem to be safe. What should be justified is (R1), (R2), (R3), and (R6). In what follows, I will argue for each of them in turn.

For (R1), the first thing to notice is that it does not involve any modal claim. This makes (R1) safe from the objection that the illuminati argument faces. As explained in the previous section, the illuminati argument involves the partially modal claim, namely P&Q&□(p=q). It applies CP+ to P&Q&□(p=q), so that it cannot avoid the objection that CP+ may be restricted only to non-modal statements. The Russellian illuminati argument does not have such problem, because (R1) includes no modal operator. Instead of □(p=q), (R1) has p_i=q. It is purely a non-modal statement. Therefore, even if the applicability of CP+ is narrowed down to non-modal claims, it does not affect the Russellian illuminati argument.

(R1) is a conceivability claim. It claims that $P_iQTI\&(p_i=q)$ is ideally, positively, and primarily conceivable. It is easy to see that $P_iQTI\&(p_i=q)$ is ideally and primarily conceivable. There seems to be no better reasoning to refute the statement. I cannot find any inconsistency or contradiction in it. Further, it is conceivable that the actual world turns out to be the world where $P_iQTI\&(p_i=q)$ is the case. Hence, (R1) is ideally as well as primarily conceivable. Justifying (R1) thus hinges on showing that $P_iQTI\&(p_i=q)$ is positively conceivable.

As explained in Section 1.3.2, to see if a certain statement is positively conceivable, one should check whether a situation is coherently modally imagined. On the one hand, there must be a psychological process of having an intuition of a situation. On the other hand, there must be a rational process of interpreting or reflecting upon the intuited situation. Interpreting or reflecting upon such situation, a conceiver investigates whether there will be any contradiction in the situation. In the end of the intertwined two processes, if the conceiver finds that the fully detailed situation verifies the statement in question, she can be said to modally imagine the statement. If so, in order to argue for the positive conceivability of $P_iQTI\&(p_i=q)$, one must answer the following question: is the situation

that verifies $P_iQTI\&(p_i=q)$ coherently modally imaginable?

I think the answer is simply yes. One can have an intuition about a fully detailed situation, and one's reflection upon such situation would tell that it verifies $P_iQTI\&(p_i=q)$. Chalmers says "I can detect no internal incoherence; I have a clear picture of what I am conceiving when I conceive of a zombie." (Chalmers, 1996, p. 99) Likewise, I have a very clear picture of the situation I am coherently modally imagining. There is a psychological process of having an intuition of a situation. There is also a rational process of reflecting upon the imagined situation. Finally, my rational processes lead me to think that such fully detailed situation verifies $P_iQTI\&(p_i=q)$. The situation should be treated as *the Russellian illuminati world*, the world where type-B Russellian physicalism is the case.

One can even illustrate what the Russellian illuminati world would be like. For instance, in such world, a complex physical inscrutable C is a migraine. C has every property that migraine has, and *vice versa.* Migraine also satisfies every inscrutable description that C satisfies. Both share everything. Nonetheless, there are no transparent and reductive explanations for their identity. The identity between migraine and C is not entailed *a priori* by truths about C or even by P_iTI. Even if one knows everything about C, she cannot be in a position to deduce any truth about migraine. The identity is supposed to be epistemically primitive. Such identity would appear as *epistemically primitive regularities*: whenever and wherever C is instantiated in my forehead, I suffer from migraine, and *vice versa*. Conversely, anytime and anywhere C is not instantiated in my forehead, I do not suffer from migraine, and the reverse holds. All the theoretical advances and empirical findings lead to the conclusion that there cannot be any law of nature or metaphysical principle, or even God's action or miraculous coincidence connecting C and migraine as two distinct phenomena. This is only a very rough picture of the situation, but one can

easily expect that a complete world can be constructed by adding all the details required. Thus, the psychological process for coherent modal imagination of the situation will go on. What about the rational process? What would our interpretation or reflection tell us about such situation? I think the answer is clear: there is no reason to think that they are distinct. Then, the rational reflection upon such situation will conclude that the situation verifies the identity statement 'C=migraine'. Clearly, the Russellian illuminati world can be exhaustively detailed without contradiction. Therefore, one can coherently modally imagine the world where type-B Russellian physicalism is true.

(R2) is an application of CP+ to (R1). The first thing to notice is that (R1) captures every aspect of conceivability that should be considered. Whereas (I1) focuses only on the positive conceivability, (R1) deals with the ideal positive primary conceivability. Moreover, unlike the illuminati argument, the statement to be conceived is free of any modal operator. Due to this non-modality, one can safely apply CP+ to $P_iQTI\&(p_i=q)$ even when CP+ is restricted to non-modal statements. This is how my Russellian version of the illuminati argument can be safe from the charge of focusing on wrong conceivability and applying CP+ to wrong statements. As Russellian argument rightly reflects three aspects of conceivability and only involves non-modal statements, there is nothing to worry about applying CP+ to (R1). Then, we can have (R2).

In (R3), we can see how the Russellian illuminati argument solves one of the two problems for anti-zombie argument. Contrast to other versions, my argument secures the semantic neutrality of $P_iQTI\&(p_i=q)$. The primary and secondary intentions of inscrutable terms must coincide. As 'H_2O' picks out H_2O in all worlds no matter how worlds are considered, in all worlds considered as actual or counterfactual, inscrutable terms would refer to physical inscrutables they actually pick out. Accordingly, the conjunction of

all inscrutable truths P_i also should be neutral. And I temporarily take a widely accepted view that the primary and secondary intensions of phenomenal terms coincide. If so, q and Q are semantically neutral and one can safely draw the Russellian illuminati world's secondary possibility from its primary possibility. The Russellian illuminati argument is not vulnerable to the possibility of Russellian monism, since it is built on inscrutable terms rather than microphysical ones: even if Russellian monism is true, primary and secondary intentions of inscrutable terms must coincide. Regardless of the truth of Russellian monism, with the assumption that q and Q are semantically neutral, (R3) holds.

(R6) argues that $P_iTI\&\sim Q$ is ideally positively primarily conceivable. For the ideal conceivability, I cannot find any rational reasoning that falsifies $P_iTI\&\sim Q$. Remind that P_i is the conjunction of inscrutable truths. These inscrutable truths must be broadly physical truth. And according to type-B Russellian physicalism, there cannot be any *a priori* entailment between broadly physical truths and an arbitrary phenomenal truth. Then, there is no barrier to ideally conceive P_iTI without Q. For positive conceivability, I have a clear positive conception about a situation where $P_iTI\&\sim Q$ is true. The situation might be considered as *the Russellian zombie world*, which shares every broadly physical inscrutables with the Russellian illuminati world but lacks q. For Q is a truth about q, Q cannot hold in the Russellian zombie world. In such world, p_i still realizes microphysical dispositions and structures but does not have any property that q has. Neither does it satisfy any phenomenological description which q would satisfy. No matter how arbitrary details are fleshed out, my rational reflection upon such world does not find any contradiction. The Russellian zombie world is simply a zombie world whose microphysical structures and dynamics are realized by physical inscrutables. This Russellian version of zombie worlds appears to be as positively conceivable as the original

version. Last, for primary conceivability, I can easily conceive of the case where the actual world turns out to be the Russellian zombie world. There seems to be no *a priori* reasoning to rule out the epistemic hypothesis that the Russellian zombie world is actually the case. Thus, $P_iTI\&\sim Q$ is conceivable, in an ideal, positive, and primary sense. Let us call (R6) the conceivability of Russellian zombies.

For the sake of argument, I assumed that phenomenal terms are semantically neutral thus far. While this assumption has been widely endorsed, one may doubt it in principle. Primary and secondary intensions of phenomenal terms might not coincide in such a way that primary intensions are determined by what phenomenal properties are *like*, while secondary intensions are fixed by what phenomenal properties really *are*. In the case of the phenomenal term pain, for instance, the primary intension would pick out the painfulness of pain, while the secondary intension would refer to a property that is actually painful. This distinction of primary and secondary intensions of 'pain' revives a modal illusion for pain. That is, one can legitimately conceive of a world where a certain property that is painful but not pain is not identical with the actually painful property. This modal illusion directly defies the Kripkean intuition that what is qualitatively identical with pain just is pain. It is widely believed that introspection upon experience tells us what experience is like, and what experience is like is what it is. However, introspection may reveal only what it is like, not what it really is. Once the appearance-reality gap is restored in phenomenal properties, there seems to be no way to deny such counterintuitive conceivability. Except the Kripkean intuition, there seems to be no principled reason to preclude the appearance-reality gap in phenomenal properties. If so, one can have a right to doubt the semantic neutrality of phenomenal terms.

If phenomenal terms are not semantically neutral, the Russellian illuminati

argument cannot be sound. Premise (R3) and (R8) fail to be justified. I think, however, another argument that is equivalent to the Russellian illuminati argument can be constructed. Even when primary and secondary intensions of phenomenal terms are distinct, one can build a new Russellian illuminati argument by utilizing *expressions describing properties that are essentially reflected by the primary intensions of phenomenal terms*. Let me explain. If type-B Russellian physicalism is right, a phenomenal term would behave like natural kind terms. For instance, as the primary and second intensions of 'water' are respectively <the watery stuff> and <H_2O>, those of 'pain' would be like <the painful feeling> and <complex physical inscrutable>. Here, painfulness can be considered as a property that is closely associated with the primary intension of 'pain'. The primary intension of 'pain' must be determined by what can be known by *a priori* understanding of the term. Painfulness is the only property we can know *a priori* about 'pain'. Thus, the primary intension of 'pain' must reflect painfulness. Without reflecting painfulness, it cannot be the primary intension of 'pain'. In this sense, painfulness is the property essentially reflected by the primary intension of 'pain'. Concerning the other phenomenal term, 'phenomenal redness', phenomenal redness-likeness would be such property. The only property that we can know *a priori* about 'phenomenal redness' is that it is phenomenal redness-like, so that the primary intension of 'phenomenal redness' must reflect phenomenal redness-likeness. Let us call such properties *intensional essences.*

Now, we can think of an expression that describes painfulness, 'painfulness' for example. The important step is analyzing this expression in two-dimensional way. What is the primary intension of 'painfulness'? Clearly, 'painfulness' would pick out painfulness in all worlds considered as actual. What is the secondary intension of the term? Definitely, the term would refer to painfulness in every world considered as counterfactual. The

primary and secondary intensions of ‘painfulness’ coincide. They both rigidly designate painfulness in all worlds no matter how worlds are considered. Likewise, the primary and secondary intensions of ‘phenomenal redness-likeness’ would rigidly designate phenomenal redness-likeness in all world. These expressions can be called *intensional essence terms.* It can be said that intensional essence terms for consciousness always capture intensional essences of consciousness, which are appearances of consciousness revealed by introspection.

Now, we can see how a new version of the Russellian illuminati argument can be constructed even when phenomenal terms are semantically non-neutral. If phenomenal terms are not semantically neutral, it is impossible to build the Russellian illuminati argument with phenomenal terms. However, there is more than one way to skin a cat. Instead of phenomenal terms, one can use intensional essence terms for phenomenal properties. To illustrate, suppose that the primary and secondary intensions of ‘pain’ are distinct. Then, we must give up the original version of the Russellian illuminati arguments. Nonetheless, the phenomenal term leaves an intensional essence term behind. As a semantic substitute for ‘pain’, one can use ‘painfulness’. ‘Painfulness’ is a perfect substitute for ‘pain’ in that it provides everything required for constructing a new Russellian illuminati argument: almost everybody knows *a priori* what ‘painfulness’ means. Moreover, with inscrutable terms and truths, it can yield various semantically neutral truths, whose primary possibilities entail secondary possibilities. Therefore, if ‘pain’ is semantically non-neutral, one can use ‘painfulness’ instead. If ‘phenomenal redness’ is semantically non-neutral, one can use ‘phenomenal redness-likeness’ as an ersatz phenomenal term. Formalizing a new Russellian illuminati argument is simple. All we need to do is slightly revising the terminology of the original version: replace q with e_i, which is an intensional essence term for an arbitrary phenomenal property.

To sum up, regardless of whether phenomenal terms are semantically neutral or not, the Russellian illuminati argument holds. If the primary and secondary intensions of phenomenal terms coincide, one can have the original Russellian illuminati argument. Even if they do not coincide, then one can construct a variant of the original argument, replacing phenomenal terms with intensional essence terms. The situation also can be put in terms of consciousness: if there is no appearance-reality gap in consciousness, there would be both Russellian illuminati worlds and Russellian zombie worlds, and one can raise a *reductio* argument against CP+. Even if consciousness allows the appearance-reality gap, then there would be a world where the appearance of consciousness is broadly physical and its broadly physical duplicate without the appearance of consciousness, so that CP+ faces a *reductio* again. In one way or another, there can be a version of the Russellian illuminati arguments against CP+.

3.2.3 Objections and Replies

Possible objections to my argument are expected. As other premises of the Russellian illuminati argument seem acceptable, objections can be raised against the two central premises, (R1) and (R6). I anticipate most of objections will be concentrated on the conceivability of Russellian illuminati and Russellian zombies. In this section, I will consider a number of possible objections and reply in turn.

Objection 1: you claim that we can positively conceive both Russellian illuminati and Russellian zombies. The problem is, however, that we do not know much about the central notions involved in these statements yet. Type-B Russellian physicalism's minimal and rough description falls short of providing positive conceptions about what inscrutables are. Further, while inscrutables are assumed to be broadly physical in type-B Russellian

physicalism, we do not have any idea about what these broadly physical inscrutables would be like. Our understanding about the nature of the physical has been governed by the standard fundamental physics. The physics, however, does not, and cannot, tell us anything about being broadly physical. If the descriptions of type-B Russellian physicalism are not positive and specific enough, it is not clear at all that we can positively conceive Rusellian illuminati or Russellian zombies.

Reply: though we do not know everything about inscrutables, we know *enough.* We know what properties should be counted as inscrutables. Properties that ground microphysical dispositions and structures and constitutively contribute to phenomenal properties are inscrutables. This is all we need to know when we are positively conceiving Russellian illuminati and Russellian zombies. The rest is a matter of arbitrary details. Even under rough and schematic descriptions, one can still have an intuition of a situation, if she can fill in all the details without contradiction. For instance, even when only approximate and abstract descriptions about a flying pig is given, it is still possible to have an intuition of such imaginary creature. Thus, while type-B Russellian physicalism's descriptions about inscrutables are not specific, it cannot be a reason to doubt the positive conceivability of a statement about inscrutables. Even though we do not know any specific description about being broadly physical, we do know what should be considered as broadly physical. If something is an inscrutable but neither phenomenal nor neutral, it can be counted as broadly physical and one can name it 'broadly physical'. The lack of specific and positive ideas about broadly physical inscrutables does not matter. What such neither-phenomenal-nor-neutral inscrutables would be like is, again, a matter of details. Anything can be counted as broadly physical, insofar as it does not contradict with the descriptions provided by type-B Russellian

physicalism. Indeed, nothing prevents one from positively conceiving a statement involving broadly physical inscrutables.

It seems that the objection conflates *inconceivability* with *infinite conceivability*. From the fact that central notions involved in a statement are not specific, the objection argues that such statement cannot be positively conceivable. The exact opposite seems true, however. It is not that there is no way to positively conceive a statement involving rough notions. Rather, there are infinite ways to positively conceive the statement. There is no limit in filling in arbitrary details of a verifying situation. This explains why some find Russellian illuminati or Russellian zombies positively inconceivable. Non-specific, negative descriptions about broadly physical inscrutables allow too much: there are too many ways of detailing, so that coherently modally imagining Russellian illuminati or Russellian zombies tends to go beyond the limit of ordinary psychological and rational abilities. For this overwhelming cognitive overload, we are compelled to think that the statements are positively inconceivable.[31)]

Objection 2: $P_iQTI\&(p_i=q)$ may be negatively conceivable, but not positively. Some may object that the identity claim $p_i=q$ is positively inconceivable, since we cannot form a positive conception of a situation that verifies the claim. There can be several reasons for this objection. First, some may point out that the identity between p_i and q is unexplainable in principle and one cannot have a positive conception of something unexplainable.

---

31) In this respect, positively conceiving such statements is similar to imagining what would happen inside a black hole. Many would say that we 'cannot' imagine what would happen inside of a black hole, since everything is possible in there. Yet, if everything is really possible, it implies that we can imagine as much as possible. There would be almost infinite ways to provide a consistent and detailed story about the heart of the black hole. This infinity is so dazzling that one might feel that she 'cannot' think up anything about it. Likewise, one might conflate positively inconceivable statements with infinitely positively conceivable ones, when verifying situations can be detailed in almost unlimited ways.

Second, in the Russellian illuminati world, there are no properties, relations, laws, or principles that can explain the identity. This implies that in order to form a positive conception of the Russellian illuminati world, one should have an intuition of the absence and reflect upon it. This intuition of or reflection upon the absence seems psychologically and rationally problematic. At best, it would provide a certain negative conception.

Reply: first, there are many situations that are unexplainable but positively conceivable. In fact, Chalmers provides one example. Arguing that interactionist dualism is compatible with the conceivability of zombies, he says "physically identical beings without consciousness will presumably have large causal gaps in their functioning (or else will have some new element to fill those gaps), but there is nothing obviously inconceivable about such causal gaps." (Chalmers, 2010, p. 156) One can form a positive conception of zombies' unexplainable functioning. The fact about zombies' functioning cannot be explained by any fact about realizers of functioning, because there are no such facts. Likewise, the fact about the identity between p_i and q cannot be explained by certain grounding facts, since there no such grounding facts. Thus, if one can have a positive conception of such brutal functioning, I think she can also have a positive conception of the brutal identity in the Russellian illuminati world.

Second, there seems to be no problem in having an intuition of or reflecting upon a situation in which there is no ectoplasm or psychophysical law. Philosophers always engage in such thought experiments. The objection appears to assume that intuition of or reflection upon absences is cognitively impossible. Or, it might be thought that intuition of or reflection upon absences enforces us to think about something rather than nothing. Further, one might even think that we can have an intuition of or reflect upon an absence of something only when we already know well about it. The

examples mentioned above clearly show that all these claims are plainly false. We can form a positive conception of an absence of something that we do not know exactly what it is.

Last, the history of debates concerning the zombie argument suggest that if there are debates about a positive conceivability, the burden of proof must be on the side of those who are *against* the conceivability. I think this dialectic should be applied to the debates concerning the Russellian illuminati argument. Once there is a claim of the positive conceivability of $P_iQTI\&(p_i=q)$, opponents must provide a certain counterargument or counterevidence. I think the opponent cannot shoulder this burden, however. In order to deny the positive conceivability of Russellian illuminati, opponents must argue against the coherent modal imagination of the Russellian illuminati world. There are only two ways: denying that one can have a detailed intuition of the Russellian illuminati world or denying that the imagined Russellian illuminati is coherent. Both seems unlikely. Intuition is not something that can be simply denied. As I have argued in Section 1.3.4, one cannot have an intuition from the scratch. Intuition is largely relative to subjects' psychological, philosophical or ideological bias. Even if a situation does not seem intuitive to dualists or orthodox physicalists, it may appear intuitive to type-B Russellian physicalists. It is not obviously at all that how this disagreement in intuitions can be resolved. One cannot merely pit one intuition against another. Further, there seems to be no apparent contradiction or inconsistency in the imagined Russellian illuminati world. Unless it turns out to be incoherent, the imagined Russellian illuminati world should be taken as coherent. So it seems hard to *argue* against the conceivability of Russellian illuminati.

Objection 3: type-B Russellian physicalism seems to involve a very problematic sort of identity. The identity between phenomenal qualities and

complex physical inscrutables is assumed to be epistemically primitive. This identity is so different from paradigmatic cases that one may claim that there cannot be such identity. Indeed, Chalmers and Jackson claim "Identities are ontologically primitive, but they are not epistemically primitive." (Chalmers and Jackson, 2001, p. 354) Almost in all cases, identities are entailed by underlying truths that are irrelevant to identities. Water=H_2O, for instance, is implied by the underlying truths in chemical or microphysical truths. When one exhaustively knows all the truths about H_2O, she will be able to deduce *a priori* water=H_2O. To epistemically primitive identity, however, there cannot be any *a priori* deduction: even if all the truths about C is known, a subject will not be in a position to deduce *a priori* C=migraine, since there are no underlying truths that entail C=migraine. Such identity is too exceptional to be considered as identity. There seems to be no reason to believe that there is such epistemically primitive sort of identities.

Reply: the absence of epistemically primitive identities cannot be the reason to think that they are *inconceivable.* The Russellian illuminati argument is grounded by the ideal positive primary *conceivability* of epistemically primitive identities. It does not require the actual truth or even the possibility of such identities. Pointing out that identities are generally implied by underlying truths that do not involve any identity is thus irrelevant to my claim that $P_iQTI\&(p_i=q)$ and C=migraine are conceivable. There seems to be no better reasoning to defeat the belief about epistemically primitive identities. And there seems to no reason not to form a positive conception of epistemically primitive identities. We can argue that the Russellian illuminati world is conceivable in every relevant sense.[32)]

32) Chalmers and Jackson says "A type-B materialist might bite the bullet on these things and hold that psychophysical identities are sui generis. In response, one can argue that identities between natural phenomena *cannot* be epistemically primitive.

Interestingly, it is Chalmers and Jackson's discussion that strongly suggests epistemically primitive identities are conceivable. Although Chalmers and Jackson(2001) do not believe that there are epistemically primitive identities, they spend a considerable amount of time discussing what such identities would be like. "We think that this sort of case cannot occur, but we will set that worry aside for the moment, and will *pretend* that it can occur." (Chalmers and Jackson, 2001, p. 353, italics added) What does this 'pretending' mean? It simply means that the situation where epistemically primitive identities hold is intuitive enough to be discussed and analyzed. If

---

The point where one finds objective (nonindexical) epistemically primitive regularities among natural phenomena is precisely the point at which one finds fundamental natural laws. And one can argue that what it is to be a fundamental law of nature is precisely to be an objective, epistemically primitive counterfactual-supporting regularity. If this is right, then if there are epistemically primitive psychophysical regularities, they must be regarded as fundamental natural laws." (Chalmers and Jackson, 2001, p. 357) They seem to think that epistemically primitive regularities between natural phenomena entail fundamental natural laws. If there are such fundamental laws, identities would be explained by those laws and cannot be epistemically primitive. However, this response begs the question of *conceivability* of epistemically primitive identities. If it is conceivable that epistemically primitive regularities are *sui generis*, it will be also conceivable that such regularities are supported by epistemically primitive identities. Moreover, even if to be a fundamental law is nothing but to be an epistemically primitive regularity, it would be still conceivable that epistemically primitive regularities are grounded by epistemically primitive identities. There is not reason not to conceive these cases.

Chalmers also claims "Indeed, it is often held that this sort of primitiveness—the inability to be deduced from more basic principles—is the mark of a fundamental law of nature. In effect, the type-B materialist recognizes a principle that has the epistemic status of a fundamental law but gives it the ontological status of an identity. An opponent will hold that this move is more akin to theft than to honest toil. Elsewhere, identifications are grounded in explanations, and primitive principles are acknowledged as fundamental laws." (Chalmers, 2010, p. 116) Again, the issue is not whether assuming epistemically primitive identity is "theft" or not. What identifications in elsewhere are like does not matter. The issue is that such "theft" is ideally positively and primarily conceivable. No matter how it is akin to theft, if an epistemically primitive identity is conceivable, the Russellian illuminati argument works.

the situation were not intuitive, Chalmers and Jackson would not even attempt to pretend that such situation can occur. However, under the pretense that epistemically primitive identities hold, they describe what such identities would be like and what kind of explanatory work they can do. Their analysis clearly shows that one can have an intuition of and reason about epistemically primitive identities. Chalmers and Jackson's discussion does not provide any reason to doubt that statements about epistemically primitive identity are positively conceivable. Rather, it lends a strong support for their positive conceivability.

Objection 4: replacing $\Box(p=q)$ with $p_i=q$ in (R1), you claim that the Russellian illuminati argument is safe from the objection that CP+ should be applied only to non-modal claims. This is a mistake, however. Even though there is no explicit modal operator, $p_i=q$ is an identity statement anyway. An identity statement already smuggles in necessity. There would be no difference between conceiving an identity statement and conceiving necessary statement. Either way, we are to conceive a necessary identity statement. If so, conceiving $p_i=q$ would be equivalent to conceiving $\Box(p_i=q)$. Then, the apparent difference between the original illuminati argument and the Russellian illuminati argument disappears, and the Russellian illuminati argument will suffer from the same objection as the illuminati argument.

Reply: at least for the positive conceivability, there is a significant difference between conceiving non-modal claims and conceiving modal claims. The crucial point is that when we are positively conceiving a non-modal claim, it is enough to have an intuition of and reflect upon a non-modal situation. In order to positively conceiving a modal claim, however, one must have an intuition of and reflect upon a modal situation. For instance, when one positively conceives water=H_2O, it is enough for her

to coherently modally imagine a situation where H_2O has all the non-modal properties that water has and *vice versa*. In this case, we imagine a situation where H_2O is watery and water is H_2O-like. No modal properties are involved. Water and H_2O share all their non-modal properties and that is all. On the other hand, in positively conceiving $\Box$(water=H_2O), one must coherently modally imagine a situation where H_2O is necessarily identical to water and water is necessarily identical to H_2O. At least two modal properties, being necessarily identical to water and being necessarily identical to H_2O, must be involved in the imagined situation. If there are no such modal properties, it is clear that after the reflection upon the imagined situation, one would find that the imagined situation does not verify $\Box$(water=H_2O). That is, positively conceiving non-modal and modal claims involve different imagined situations.

If this is the case, the same goes with p_i=q and $\Box$(p=q). Positively conceiving p_i=q and positively conceiving $\Box(p_i$=q) are not equivalent, since imagined situations that verify them are different. The imagined situation that verifies p_i=q involves only non-modal properties, while the imagined situation that verifies $\Box$(p=q) must be partially constituted by necessary properties. To verify $\Box(p_i$=q), the situation must contain at least two modal properties: being necessarily identical to q and being necessarily identical to p_i. Thus, one can distinguish the claim that p_i=q is positively conceivable from the claim that $\Box(p_i$=q) is so. The Russellian illuminati argument applies CP+ only to the former, not the latter. Then, the Russellian illuminati argument would not suffer the same weakness as the illuminati argument does.

Objection 5: Even if $P_iQTI\&(p_i$=q) is ideally positively primarily conceivable, it is still harder to conceive than PTI&~Q or $P_iTI\&\sim Q$. Russellian illuminati are more difficult to conceive than zombies or

Russellian zombies. Chalmers says "Many people have noted that it is very hard to imagine that consciousness is a physical process. I do not think this unimaginability is so obvious that it should be used as a premise in an argument against materialism, but likewise, the imaginability claim cannot be used as a premise either." (Chalmers, 2010, p. 180) This also can be said to Russellian illuminati. Though Russellian illuminati are conceivable in some relevant sense, such conceivability is not strong enough to be used as a premise of the Russellian illuminati argument against CP+.

Reply: conceivability is not a matter of degree. As I explained in Section 1.3.4, conceivability is *digital*. If a statement is conceivable, it is conceivable. Saying a certain statement is harder to conceive than others is misleading. I have argued in Section 1.3.4 that such view conflates probability of actual conceiving with conceivability *per se*. There is no threshold of degrees of conceivability to be used as a premise of an argument. Frankish(2007) provides helpful comments on this issue.

> It is true that there is some imaginative resistance to the idea that consciousness might be physical. 'How could this', people sometimes ask, mentally indicating some experience, 'be just a neurological state?'. Difficulty is irrelevant here, however. Conceivability is all or nothing, and one state of affairs may be harder to imagine than another without being less conceivable. (It is, for example, much harder to imagine Ronald Reagan and Freddie Mercury being the same person than to imagine their being distinct, but the two scenarios are on a par with respect to primary conceivability.) (ibid., p. 660)

Objection 6: the conceivability of Russellian zombies is questionable. As mentioned in objection 1, we have no idea what P_iTI would be like. If there is an *a priori* entailment from P_iTI to Q, $P_iTI\&\sim Q$ would not be

conceivable in any relevant sense. At the current stage of discussion, there is no guarantee that there is no such *a priori* entailment. In this respect, P_iTI is essentially different from PTI. There is a principled reason to deny that PTI entails *a priori* any phenomenal truth. No matter what the complete physics turns out to be, it would only deliver structures and dynamics of the world. It seems that structures and dynamics, whatever they are, can be instantiated without experience. This is the crucial point that Chalmer(2010; 2015) repeatedly emphasizes.[33] Nonetheless, we cannot find any such principled reason in the relationship between P_iTI and Q. The only thing we know about P_iTI is that it is a conjunction of all broadly physical truths plus indexical truths. No one knows at this point that what the broadly physical truths plus indexical truths would entail *a priori*. If one cannot rule out the possibility that $P_iTI \supset Q$ holds, it is possible to reject (R6), and the Russellian illuminati argument fails.

Reply: first thing to notice is that this objection is dialectically weak. To argue against a conceivability claim, one must explicitly point out a hidden contradiction in the claim. Likewise, in arguing against the claim that $P_iTI\&\sim Q$ is conceivable, merely pointing out that P_iTI is poorly understood or $P_iTI \supset Q$ might hold does not work. Those who argue against the

33) Chalmers says "I have occasionally heard it said that panprotopsychism can be dismissed out of hand for the same reason as materialism. According to this objection, the epistemic arguments against materialism all turn on there being a fundamental epistemic (and therefore ontological) gap between the nonphenomenal and the phenomenal: There is no *a priori* entailment from nonphenomenal truths to phenomenal truths. If this were right, the gap would also refute panprotopsychism. I do not think that this is right, however. The epistemic arguments all turn on a more specific gap between the physical and the phenomenal, ultimately arising from a gap between the structural (or the structural/dynamical) and the phenomenal. We have principled reasons to think that phenomenal truths cannot be wholly grounded in structural truths. But we have no correspondingly good reason to think that phenomenal truths cannot be wholly grounded in nonphenomenal (and nonstructural) truths, as panprotopsychism suggests." (Chalmers, 2015, p. 81)

conceivability of $P_iTI\&\sim Q$ must point out $P_iTI\&\sim Q$ is inconsistent or prove that $P_iTI \supset Q$ holds. As I mentioned in reply to Objection 2, this feature reflects general dialectics of debates concerning conceivability arguments: once there is a debate about a certain conceivability claim, the burden of proof is always on the side of those who argue against the claim. If there is any moral from the two decades of 'the zombie war', it would be that it is always physicalists who must show why zombies are inconceivable. If someone rejects the conceivability of zombies by pointing out that we do not have PTI yet or $PTI \supset Q$ might hold, we would dismiss such reaction as dialectically inappropriate. This dialectic of the zombie argument should be applied to the debate concerning the Russellian illuminati argument. The burden of showing that $P_iTI\&\sim Q$ is inconceivable is on the side of those who argue against it. In order to reject the conceivability of $P_iTI\&\sim Q$, what is needed is an actual refutation, not a mere skepticism.

Further, I think there is no essential difference between PTI and P_iTI regarding *a priori* entailment. There is a principled reason to deny that $P_iTI \supset Q$ holds. Whatever it is, P_i should consist of truths about physical inscrutables. Nonetheless, both truths about physical inscrutables can be accessed neither by perception nor by science. We cannot even introspect what they are. If there is any truth about physical inscrutables, it must be acquired through theoretical and even speculative considerations. Indeed, many Russellian monists claim that they pursue the best explanation for most of data. All things considered, it is clear that P_i must be *public truth*: truth that can be shared with others and communicable through public language. If Russellian monists find some our inscrutable truths, we would be very glad to hear what they are and (hopefully) be able to understand them. Q, on the other hand, cannot be public truth. Qualia are usually characterized as *ineffable,* and such ineffability leads to qualia's another essential property, *privacy*. Even though we can indirectly infer or describe

phenomenal truths, there cannot be any public language to capture phenomenal truths. In this sense, Q must be *private truth*. There is a sort of epistemic gap between these public and private truths. Chalmers often says "the structure and dynamics of physical processes yield only more structure and dynamics" (Chalmers, 2010, p. 15). It seems plausible that the public and objective truths yield only more public and objective truths. How can something ineffable in nature come from something essentially effable? More precisely, *how can publicly communicable broadly physical truths (plus indexical truths) entail publicly incommunicable phenomenal truths?* This gap between the public and the private provides a good reason to doubt that there can be *a priori* entailment from P_iTI to Q. Thus, there is a principled reason to believe that phenomenal truths cannot be *a priori* entailed by inscrutable truths.

In this chapter, I have argued that there is a counterargument against CP+. According to the Russellian illuminati argument, insofar as both Russellian illuminati and Russellian zombies are equally ideally positively primarily conceivable, CP+ yields a contradiction. If the argument is sound, the second central premise of the zombie argument is false. That is, the ideal positive primary conceivability of zombies does not entail the metaphysical possibility of zombies.

Chapter 4
Phenomenal Concept Strategy and the Master Argument

4.1 Chapter Introduction

There are many ways for type-B materialists to argue against the zombie argument. Some type-B materialists take the Phenomenal Concept Strategy(PCS). PCS appeals to the special nature of phenomenal concepts to explain why there is the explanatory gap or why zombies are conceivable. While PCS has been thought of as denying the second premise of the zombie argument, close examinations reveal that PCS is irrelevant to the second premise. Rather, even when the first and the second premises of the zombie argument are well defended, PCS enables physicalists to argue against the zombie argument. Chalmers, however, provides the master argument against all possible forms of PCS. In Section 4.2, I will show how PCS can be 'the third way' for physicalists and why the master argument fails. Moreover, I shall present a dilemma against the master argument in Section 4.3. This will complete my fourfold argument against the zombie argument.

4.2 Epistemic Equilibrium and the Anti-Master Argument

Type-B materialists claim that although there is the epistemic gap between physical processes and experience, there is no ontological gap. If so, even if the zombies are conceivable, they are not possible. One of the ways to argue for this claim is so-called the Phenomenal Concept Strategy(PCS). Proponents of PCS often argue that why there is the epistemic gap can be explained by the special feature of phenomenal concepts. Several versions of

PCS already have been suggested and developed by type-B materialists. Chalmers, however, argues that PCS is inherently doomed to fail. According to his master argument, PCS faces a dilemma: no matter how the strategy is developed, either phenomenal concepts cannot be physically explained or our epistemic situation cannot be explained by phenomenal concepts. Either way PCS cannot succeed. Although the master argument seems plausible at first sight, I shall argue that it is not PCS that is stuck in a dilemma. Rather, it is the master argument that faces its dilemma. When the notion of epistemic situation is rightly understood, all the Chalmers argues turn out to be misleading because they ignore the special constraint Chalmers himself puts on the epistemic situation. First, in Section 4.2.1, I will briefly introduce PCS and how exactly it can reply to the zombie argument. Chalmers' analysis of PCS and the master argument is summarized in Section 4.2.2. In Section 4.2.3, it will be argued that, contrary to what Chalmers claims, our and our zombie twins' epistemic situation are epistemically equal. Then, PCS can hold one of the two horns of the master argument.

4.2.1 Phenomenal Concept Strategy and the Zombie Argument

Some physicalists admit that zombies are conceivable in some sense. They accept that there is a certain epistemic or conceptual gap between microphysical truths and phenomenal truths. What they deny is that there is a metaphysical or ontological gap between physical properties and phenomenal properties. According to these physicalists, although zombies are conceivable in some sense, this conceivability tells us nothing about the metaphysical possibility of zombies. These physicalists are called type-B materialists.

How can type-B materialists accept the epistemic gap but deny the ontological gap? Some of type-B materialists claim that while phenomenal

properties are identical to physical properties, phenomenal concepts are distinct from physical concepts. That is, even though there is only physical kind of properties, there are two sorts of concepts, phenomenal and physical concepts. According to type-B materialists, phenomenal concepts have some conceptually, cognitively, or psychologically special features. These special features enable us to explain why zombies are conceivable and why there is the epistemic gap between the phenomenal and the physical, without appealing any non-physical, phenomenal properties. Also, type-B materialists hold that the special features themselves can be explained in physical terms. If this is the case, while there cannot be a physical explanation for consciousness, why there cannot be a physical explanation for consciousness can be physically explained. In this way, conceptual dualism and ontological monism can be reconciled. Further, physicalism can be compatible with the conceivability of zombies and epistemic gap. Stoljar(2005) named this interesting and attractive move *the phenomenal concept strategy (PCS).*

One can distinguish at least four versions of PCS. First, some philosophers claim that phenomenal concepts are nonstandard *recognitional concepts* that pick out their referents via essential modes of presentation. (Loar, 1990/1997; Carruthers, 2004; Tye, 2003a; Levin, 2007) According to these philosophers, some physical properties essentially have *special modes of presentation*, which can explain the explanatory gap. Others argue that phenomenal concepts and physical concepts play very *different conceptual roles.* (Hill, 1997; Hill and McLaughlin, 1999) They claim that special cognitive roles played by phenomenal concepts can explain the explanatory gap. A number of philosophers think that phenomenal concepts are *indexical concepts.* (Ismael, 1999; O'Dea, 2002; Perry, 2001) They suggest that the explanatory gap can be considered as a gap between indexical concepts and physical concepts. Finally, other philosophers suggest that phenomenal concepts are *quotational concepts*, which 'quote' physical properties in such

a way that some expressions quote other expressions. (Papineau, 2002; 2007; Block, 2007) Despite their differences, all four versions show the general structure of PCS: first, PCS endows certain special features to phenomenal concepts. Then, it explains our epistemic situation with the special features of phenomenal concepts.

It is crucial to understand what type-B materialists must do with PCS. Dialectically, type-B materialists are not obligated to justify their physicalism. What they must do is showing how our epistemic situation with regard to consciousness can be explained by some special features of phenomenal concepts, *if* physicalism is right. In PCS, physicalism is presupposed or assumed, not argued. Once PCS succeeds in explaining why there is the explanatory gap or why zombies are conceivable, it completes its mission. In other words, what PCS should do is to reconcile ontological monism (physicalism) and conceptual dualism. It is not PCS's duty to justify ontological monism itself.

Further, it is worth noting that PCS has a semantically significant implication. According to PCS, physical properties can be picked out by two concepts. In Fregean term, a physical property can have two modes of presentation: a phenomenal mode and a physical mode. Fregean senses of phenomenal concepts can be phenomenal modes of presentation, but referents of phenomenal concepts must be physical properties. In the two-dimensional framework, this Fregean distinction between sense and referent in phenomenal concepts can be interpreted as a distinction between the primary and secondary intensions. That is, primary and secondary intensions of phenomenal concepts do not coincide: whereas primary intensions of phenomenal concepts refer to phenomenal properties, secondary intensions pick out physical properties. On the other hand, primary and secondary intensions of physical concepts refer to the same physical properties. For example, both primary and secondary intensions of the physical concept

<H_2O> pick out H_2O. Simply put, PCS entails that while phenomenal concepts are semantically non-neutral, physical concepts are semantically neutral.

Confronted with the zombie argument, many type-B materialists have adopted PCS to strike back. However, exactly how PCS can be used to argue against the zombie argument is not clear. PCS enables type-B materialists to accept the conceivability of zombies but denies the metaphysical possibility of zombies. To the 'old' zombie argument, this can be a proper reply: the second premise of the old zombie argument states that if PTI&~Q is conceivable, PTI&~Q is metaphysically possible. The problem is that the zombie argument physicalists must deal with is not the old zombie argument. Criticizing the old zombie argument is toothless, since there is a more developed and articulated, official version of the zombie argument.[34] Only this official version deserves to be called the zombie argument against physicalism. The zombie argument does not merely suppose that conceivability entails metaphysical possibility. As we have seen in the previous chapter, the zombie argument's second premise is an application of CP+, which connects ideal positive primary conceivability to primary possibility. Primary possibility is not metaphysical possibility. Thus, even if type-B materialists argue that zombies are conceivable but not metaphysically possible, it does not have any bearing on the zombie argument. Then, it is not obvious that PCS can be an effective strategy for type-B materialists.

I think there is a good reason for type-B materialists to take PCS: it can deny the third premise of the zombie argument. The third premise states that if PTI&~Q is primarily possible, then PTI&~Q is secondarily possible or Russellian monism is true. How can PCS reject the third premise? Remind that PCS entails the semantic non-neutrality of phenomenal concepts.

34) This was the main point of Section 1.2.

Since phenomenal concepts' primary and secondary intensions differ, Q's primary and secondary intensions must differ. If so, even if PTI&~Q is primary possible, it may not be secondarily possible. Further, PCS entails the semantic neutrality of physical concepts. As physical concepts' primary and secondary intensions coincide, P's primary and secondary intensions should coincide. Then, Russellian monism cannot be true, because Russellian monism supposes that primary and secondary intensions of microphysical concepts do not coincide. In short, if PCS is successful, one can accept the primary possibility of zombies, while denying both the secondary possibility of zombies and Russellian monism.

The analysis so far suggests that PCS can be 'the third way' for physicalists. If PCS works, even if zombies are conceivable and conceivability entails primary possibility, the secondary (metaphysical) possibility of zombies does not follow. The key to refute physicalism is the metaphysical possibility of zombies, and PCS can effectively reject the metaphysical possibility of zombies. Indeed, PCS is in itself extremely interesting and also can be an attractive strategy for defending physicalism.

4.2.2. The Master Argument

Chalmers knows well about PCS. He presents the general structure of PCS. (Chalmers, 2010) According to Chalmers, proponents of PCS hold a thesis C that attributes special psychological features of phenomenal concepts. These special psychological features can be called "the key features". (ibid., p. 310) As explained in the previous section, PCS must show two things: first, it must show that how C explains our special epistemic situations with regard to consciousness, such as the conceivability of zombies or the epistemic gap. Second, PCS must show how C can be explained in physical terms. If PCS can do both, as Chalmers says, we may have "the next best thing". (ibid., p. 311) While PCS would not provide a direct physical

explanation of consciousness itself, it would provide a physical explanation of our epistemic situations. For instance, PCS never answers to the Hard problem of consciousness. Rather, it will explain away why we cannot answer the Hard problem. This is undoubtedly considerable progress.

Chalmers concedes that PCS is a powerful strategy for physicalists. However, he nonetheless thinks that PCS is doomed to fail. Chalmers argues, "[W]e can see that no account of phenomenal concepts is both powerful enough to explain our epistemic situation with regard to consciousness and tame enough to be explained in physical terms." (ibid., p. 306) According to Chalmers, PCS cannot claim C explains our epistemic situation and can be physically explained at the same time. In other words, either C cannot explain our epistemic situation or C cannot be physically explicable. For any kinds of C, Chalmers provides the following argument:

(M1) If P&~C is conceivable, then C is not physically explicable.
(M2) If P&~C is not conceivable, then zombies satisfy C.
(M3) Zombies do not share our epistemic situation.
(M4) If zombies satisfy C but do not share our epistemic situation, then C cannot explain our epistemic situation.
(M5) If P&~C is not conceivable, then C cannot explain our epistemic situation.

(M6) Either C is not physically explicable, or C cannot explain our epistemic situation. (ibid., p. 313-315)

Chalmers calls this argument *the master argument.* (Chalmers, 2010, p. 312-320) The master argument is valid. It takes the form of dilemma. The first horn is Premise (M1). The inference from Premise (M2) to (M4) yields (M5), which is the second horn of the dilemma. I will clarify Premise (M1)

and the inference in turn.

Premise (M1) can be justified as follows: if P&~C is conceivable, then P cannot entail C *a priori*. If P cannot entail C *a priori*, P cannot explain C. In other words, even if one knows everything about microphysics of our world, she would not be in a position to deduce C. This means that there is no transparent physical explanation of why C holds. Therefore, between P and C, there will be an explanatory gap. For example, if our physical duplicates that lack a key feature attributed by C are conceivable, then we would wonder why we have the key feature. As our physical duplicates share everything physical with us, actual physics will not answer the question. As the conceivability of P&~Q is enough to make the explanatory gap between P and Q, the conceivability of P&~C is enough to do the same thing. If so, we can say that if P&~C is conceivable, C is not physically explicable.

Here, it must be noted that C must be cast in *topic neutral terms.* (Chalmers, 2010, p. 314) C must not require the existence of non-physical phenomenal properties or concepts that refer to them. I will call this requirement *the constraint of topic-neutrality.* The constraint of topic-neutrality is justified by the following consideration: suppose that C explicitly requires non-physical phenomenal properties or concepts refer to such properties. Then, zombies would fail to satisfy C or acquire phenomenal concepts. Conceiving P&~Q would be tantamount to conceiving P&~C. As type-B materialists admit P&~Q is conceivable, they must admit that P&~C is conceivable. Once the conceivability of P&~C is accepted, as the conceivability of P&~Q makes the explanatory gap between P and Q, it makes the explanatory gap between P and C. A transparent physical explanation of the truth of C would be automatically ruled out. If C is not constrained to be topic-neutral, type-B materialists must take the first horn of the master argument. This is not fair to type-B materialists and PCS,

however. In PCS, the thesis C is *supposed* to be physically explained. Casting C as a thesis explicitly about non-physical phenomenal properties or concepts that refer to them begs the question of why it should be. Type-B materialist can legitimately reject such formulation of C. Therefore, C must be cast in topic-neutral terms.

How should C be formulated? According to Chalmers, although C can include psychological or epistemological vocabulary, phenomenal vocabulary must be barred. Once C is cast in this way, strictly speaking, it cannot be a thesis about phenomenal concepts. Rather, C is a thesis about *quasi-phenomenal concepts.* (Chalmers, 2010, p. 314) Chalmers claims that quasi phenomenal concepts "can be understood as concepts deployed in certain circumstances that are associated with certain sorts of perceptual and introspective processes and so on." (ibid., p. 314) Under the constraint of topic-neutrality, phenomenal concepts become quasi-phenomenal concepts. And it is clear that zombies can have such quasi-phenomenal concepts.

The argument for Premise (M2) is straightforward. Premise (M2) states that if P&~C is not conceivable, then zombies satisfy C. If P&~C is not conceivable, there can be only one reason for such inconceivability. It is because P entails C *a priori*. And once P entails C *a priori*, zombies necessarily satisfy C. Hence, the inconceivability of P&~C entails that zombies must have key features as we do.

In Premise (M3), Chalmers introduces the notion of *epistemic situation.* About epistemic situation, he says

> I will take it that the epistemic situation of an individual includes the truth values of their beliefs and the epistemic status of their beliefs (as justified or unjustified and as cognitively significant or insignificant). As before, an epistemic situation (and a sentence E characterizing it) should be understood in *topic-neutral* terms, so that it does not build in claims about the presence of phenomenal states or phenomenal concepts. We can say that two

> individuals share their epistemic situation when they have corresponding beliefs, all of which have corresponding truth values and epistemic status. (Chalmers, 2010, p. 314, emphasis original)

We and our zombie twins can be said to share their epistemic situation if we and our zombie twins have corresponding beliefs, all of which have corresponding truth values and epistemic status. Here, Chalmers assumes that if two utterances correspond to each other, they express corresponding beliefs.

It is crucial to note that as C is under the constraint of topic-neutrality, epistemic situations are under the same constraint. Again, construing epistemic situation in phenomenal terms would beg the question against PCS. If our epistemic situation is, at least partially, portrayed in non-physical phenomenal properties or phenomenal concepts referring to them, then there cannot be any way for type-B materialists to explain our epistemic situation *in principle*. Type-B materialists cannot accept such non-physical phenomenal properties and phenomenal concepts because of their metaphysics. And PCS is tailor-made for type-B materialists to explain our epistemic situations without resorting to anything existing outside the domains of the functional, physical, or psychological. However, once our epistemic situation is understood in anything phenomenal, the whole project of PCS cannot get off the ground. So, epistemic situations must not require any non-physical phenomenal properties or concepts that refer to them. In other words, beliefs in an epistemic situation must not refer to non-physical phenomenal properties or be consist of phenomenal concepts. Let us call such beliefs *quasi-phenomenal beliefs.* Under the constraint of topic-neutrality, epistemic situations must include only quasi-phenomenal beliefs and preclude all phenomenal beliefs. These epistemic situations under the constraint of topic-neutrality deserve to be called *phenomenally neutral situations.*[35)]

Chalmers presents two reasons for Premise (M3). The first reason is that zombies are seemingly less accurate in their self-conception than we are. I have many beliefs about my conscious experience, such as <I am conscious> or <I have this such-and-such experience that insistently resists any functional or physical explanation>. While my zombie twin has certain corresponding beliefs, it is intuitively appealing that its beliefs are different from my beliefs in their truth value and/or epistemic status. It seems likely that my zombie twin's beliefs are false or at least that less justified than my beliefs. The second reason is that our zombie twins' knowledge seems essentially different from our knowledge. For instance, when Mary sees red for the first time, the knowledge she gains is cognitively significant. It teaches what it is like to see red and cannot be inferred from complete physical knowledge. On the other hand, when Mary's zombie twin, Zombie Mary, is released from the black-and-white room and encounters the ripe tomato, it is plausible that she does not learn anything cognitively significant. While Zombie Mary may gain certain abilities to classify, imagine, and recognize red things or certain indexical knowledge such as <I am having this experience now>, these are by no means analogous to the cognitively significant knowledge that Mary gains. In this sense, Zombie Mary does not share Mary's epistemic situation. If so, zombies fail to share our epistemic situation.

Premise (M4) is obvious. If zombies satisfy C but do not share our

---

35) Chalmers does not require corresponding beliefs have the same content. He says, "It is plausible that a nonconscious being such as a zombie cannot have beliefs with exactly the same content as our beliefs about consciousness." (Chalmers, 2010, p. 316) It seems that while we have phenomenal beliefs that refer to non-physical phenomenal properties, our zombie twins cannot. Then, if sharing epistemic situations requires sharing contents of beliefs, it is impossible for us and our zombie twins to share their epistemic situations in principle. Again, this begs the question. Why should sharing epistemic situations require sharing contents of beliefs in the first place? Type-B materialists would deny this assumption. Thus, corresponding beliefs do not have to have the same contents.

epistemic situation, it means that there is no *a priori* entailment from C to E, which is the full characterization of our epistemic situation. And if E is not entailed by C *a priori*, even if one knows everything about C, she would not be in a position to know why E holds. Therefore, C cannot transparently explain our epistemic situation. From (M2), (M3), and (M4), (M5) follows. Premise (M1) and (M5) completes the dilemma. Then, we have (M6): either C is not physically explicable, or C cannot explain our epistemic situation.

4.2.3 Epistemic Equilibrium between Us and Zombies

The master argument is compelling. It is designed to refute all possible forms of PCS. If the argument succeeds, no PCS can succeed. In this section, I shall argue that the central premise of the master argument, (M3), is false. I think both two reasons for (M3) provided by Chalmers cannot support (M3) because of the constraint of topic-neutrality. If so, our zombie twins share our epistemic situation, and the master argument fails.

First, once our epistemic situation is cast in topic-neutral terms, there cannot be any discrepancy of beliefs between zombies and us. Under the constraint of topic-neutrality, our epistemic situation must contain quasi-phenomenal concepts and beliefs only. For the same reason, under the constraint of topic-neutrality, whatever their corresponding beliefs are, zombies' epistemic situation must be phenomenally neutral too. As a result, in the phenomenally neutral situation, we and our zombie twins only have quasi-phenomenal, topic-neutrally explicable beliefs. The crucial point is that, except non-physical phenomenal properties and phenomenal concepts, we and our zombie twins share everything. We and our zombie twins are internally as well as externally identical: physical and functional properties, environments, histories, contexts, pieces of evidence, cognitive and inferential abilities, and so on. If so, we and zombies would share every

quasi-phenomenal belief, including perceptual, introspective, indexical, abstract, and high-order beliefs. For instance, if I have a belief <I am conscious> or <I am having this experience now>, this belief must be construed in topic-neutral language. That is, it must be a quasi-phenomenal belief. Since my zombie twin shares every topic-neutral, non-phenomenal aspect with me, it must have the same quasi-phenomenal belief. If we have a certain belief, so do zombies, and *vice versa*. By contraposition, if zombies do not have a certain belief, neither do we, and the reverse holds. It short, under the constraint of topic-neutrality, we and our zombie twins are *doxastically symmetric.*

As we and zombies share everything in phenomenally neutral situations, all quasi-phenomenal beliefs shared by us and our zombie twins would share their truth values. Suppose that Mary has a true perceptual belief <The sky is blue>. Her zombie twin, Zombie Mary, would share that belief, for she shares all brain processes and cognitive abilities with Mary. Moreover, Zombie Mary's belief would be true too, since Mary and Zombie Mary shares their environments. If Mary has a wrong belief, for the same reason, Zombie Mary will share the same wrong belief. Likewise, all other sorts of shared beliefs would share their truth values. If I have a true indexical belief <I am here>, so does my zombie twin. If I falsely believe that 4 is an odd number, my zombie twin will share that false belief. The reverse also holds. In this sense, we and our zombie twins are *veridically symmetric.*

Further, all the quasi-phenomenal beliefs shared by us and our zombie twins should share their justifications. Roughly, there are two main sources of justification: causation and inference. Again, the point is that we and our zombie twins share all causal and inferential relations and pieces of evidence. I and my zombie twin share every background condition, environment, and causal law. Thus, if a certain causal process justifies my

perceptual belief <The sky is blue>, the same causal process would justify the same perceptual belief of my zombie twin. If my belief is not causally justified, my zombie twin would also fail to causally justify the same belief, and *vice versa*. On the other hand, if Mary infers her belief <I am in my room> from her indexical beliefs <I am here> and <Here is my room>, such inference justifies her belief. Due to the doxastic symmetry between Mary and Zombie Mary, Zombie Mary has the same indexical beliefs and inferential capacities. She would justify the same belief with the same inference. If Mary fails to infer that belief, so does Zombie Mary. Simply put, causally or inferentially, if we rationally hold some quasi-phenomenal beliefs, zombies would rationally hold the same beliefs. The reverse is also true. Therefore, we and zombies are *rationally symmetric.*

There cannot be any discrepancy of cognitive significance in quasi-phenomenal beliefs shared by us and our zombie twins. A belief is cognitively significant when it contains new information. Information is formed, transmitted, stored, and retrieved by cognitive systems. And we and our zombie twins share every information processing. When we have a certain belief with new information, our zombie twins would have the same belief with the same new information. Conversely, if our zombie twins have cognitively significant beliefs, such as <Hesperus=Phosphorus>, we would have those beliefs too. As far as we and our zombie twins have the same brains, the same cognitive processes, and the same environments and histories, when we have cognitively significant beliefs, zombies have the same cognitively significant beliefs, and *vice versa*. Further, zombies do not have a certain cognitively significant beliefs, neither do we, and the reverse is also true. We and our zombie twins are *cognitively symmetric.*

When two or more epistemic subjects are rationally as well as cognitively symmetric, they deserve to be called *epistemically symmetric.* Moreover, if multiple epistemic subjects are doxastically, veridically, and epistemically

symmetric, we can say that they reach an *epistemic equilibrium*. Once two or more epistemic subjects are in an epistemic equilibrium, they must share the same beliefs with the same truth values and epistemic statuses. It is easy to see that when multiple epistemic subjects are in their epistemic equilibrium, by definition, they must share their epistemic situation. I have shown thus far that once their epistemic situation is constrained to be topic-neutral, we and our zombie twins achieve an epistemic equilibrium. Then, we and our zombie twins must share their epistemic situation. In other words, when our epistemic situation is supposed to be phenomenally neutral, zombies must share our epistemic situation. This results from the constraint of topic-neutrality and the definition of zombies. If so, Premise (M3) is false, so that the master argument fails.

The whole point of my argument boils down to thins: *we and our zombie twins share every topic-neutral aspect by definition.* Of course, there can be topic-neutrally construable properties other than those mentioned so far. Whatever such properties are, unless they are phenomenal, our zombie twins must share them with us. If so, once our epistemic situation is understood in topic-neutral terms, there cannot be difference between us and our zombie twins: once beliefs, truth making, and epistemic statuses are construed in topic-neutral language, everything of our epistemic situation is explicable by our topic-neutral aspects. And our zombie twins share all topic-neutral aspects with us. Therefore, epistemically, they must have what we have, and we must have what they have. They cannot have what we do not have, and we cannot have what they do not have. The epistemic equilibrium between us and our zombie twins goes both ways.

Under the constraint of topic-neutrality, our epistemic situation is phenomenally 'neutralized'. Once our epistemic situation is phenomenally neutralized, one would not be able to find any difference between the epistemic situation of us and that of our zombie twins. Indeed, the only

thing that makes epistemic differences between us and zombies is phenomenal consciousness and phenomenal beliefs. When they are neutralized by the constraint of topic-neutrality, both our epistemic situation and our zombie twins' epistemic situation will only contain quasi-phenomenal beliefs. Then, there will be no reason for zombies not to share our epistemic situation. Rather, our zombie twins *must* share our epistemic situation by their definition. Under the constraint of topic-neutrality, we and our zombie twins become not only physically but also epistemically doppelgangers. All our quasi-phenomenal beliefs and zombies' quasi-phenomenal beliefs should perfectly mirror each other. In phenomenally neutral situations, we epistemically become zombies and zombies epistemically become us. This is the whole point of epistemic equilibrium.

What about the two reasons for (M3)? As Chalmers points out, it is strongly intuitive that zombies have wrong or at least less justified beliefs about themselves than our beliefs. Also, unlike Mary, Zombie Mary appears not to gain any knowledge even when she gets out of the room. Though she might gain a certain new belief, there is a strong intuition that such belief is different from Mary's knowledge both in justification and cognitive significance. Given the intuition, it seems that we and our zombie twins cannot reach any epistemic equilibrium.

Although these intuitions seem strong at first glance, if one reminds that epistemic situation must be characterized in topic-neutral terms, I think the intuitions in question will disappear or at least lose its force. The intuition results from ignoring the phenomenally neutral nature of epistemic situation. Why do we easily think that zombies do not have true beliefs that we have or Zombie Mary does not gain knowledge that Mary gains? What is the source of such strong intuition? The only reason I can think of is that we implicitly suppose that Mary's epistemic situation is not phenomenally

neutral. In other words, we unwittingly ignore the fact that our epistemic situation must be cast in topic-neutral terms. If our epistemic situation is not restricted to quasi-phenomenal concepts and beliefs, implicitly or explicitly, phenomenal concepts and beliefs which require the presence of non-physical phenomenal properties will be presupposed. This is tantamount to presuppose non-physical phenomenal properties. Then, one cannot help but think that we and our zombie twins or Mary and Zombie Mary are epistemically unequal. It is quite natural to believe that zombies falsely believe that they are conscious while we rightly believe that we are conscious. It is even inevitable that Mary gains a certain justified and cognitively significant knowledge but Zombie Mary does not. Once the constraint of topic-neutrality is restored, however, the initially strong intuition seems to stop being appealing and even disappear. We cannot have true phenomenal beliefs that correspond to our zombie twins' false beliefs. Mary cannot acquire a new, justified, and cognitively significant knowledge that Zombie Mary do not have.

It is worth noting that under the constraint of topic-neutrality, the newly acquired knowledge of Mary cannot be the one that is usually thought of. When we imagine what Mary would know when she gets out of the achromatic room, we naturally think that Mary gains a piece of new knowledge about a phenomenal redness of the ripe tomato. Further, some might think that such knowledge is justified by seeing the phenomenal redness. The constraint of topic-neutrality, however, undercuts these thought processes. Whatever it is, everything Mary can acquire when she is released must be quasi-phenomenal knowledge. And there is no reason for Zombie Mary not to share Mary's quasi-phenomenal knowledge.

The observation thus far reveals that Chalmers commits the fallacy of equivocation. From the start, Chalmers explicitly emphasizes that our epistemic situation must be conceptualized in topic-neutral terms and not to

build in non-physical properties or phenomenal concepts. But when he is wanting to force us to think that our zombie twins do not share our epistemic situation, he implicitly assumes that our epistemic situation must be understood under phenomenal characterization. That is, Chalmers is illicitly violating the rule of topic-neutrality. He is unwittingly changing a topic-neutral understanding of our epistemic situation with a phenomenal understanding in an attempt to argue that our zombie twins do not share our epistemic situation. Due to the constraint of topic-neutrality, our epistemic situation must not involve anything phenomenal. However, we can understand why our zombie twins do not share our epistemic situation *only* if we already assume that our epistemic situation must involve something phenomenal. Without equivocation, Chalmers cannot make these two claims at the same time. A topic-neutral understanding applies to both of us and our zombie twins. Saying that our zombie twins do not obtain cognitively significant knowledge because they lack phenomenal consciousness is to forget the fact that such cognitive significant knowledge is supposed to be construe without invoking phenomenal consciousness. Ergo, as far as our epistemic situation and our zombie twins' epistemic situation are understood *fully* or *exhaustively* in topic-neutral terms, there cannot be any difference between us and our zombie twins. In other word, they reach a perfect epistemic equilibrium.

This diagnosis helps us to deal with Chalmers' response to the claim that our zombie twins share our epistemic situation. Responding possible and actual reactions to the master argument, Chalmers writes

> This proposal might be developed in two different ways: either by deflating the phenomenal knowledge of conscious beings or by inflating the corresponding knowledge of zombies. That is, a proponent may argue either that Mary gains *less* new knowledge than I suggested earlier or that Zombie Mary gains *more* new knowledge than I suggested earlier. Earlier, I argued

> that Mary gains new cognitively significant non-indexical knowledge, whereas Zombie Mary does not. The deflationary strategy proposes that Mary gains no such knowledge; the inflationary strategy proposes that Zombie Mary gains such knowledge, too. (Chalmers, 2010, p. 327)

Chalmers rejects both ways in that neither the deflationary nor the inflationary strategy can succeed. To the deflationary strategy, he argues that if we deflate Mary's epistemic progress, then we are committed to holding that what Mary learns when she sees red for the first time is just an indexical knowledge. (Chalmers, 2010, p. 327-329) But this is implausible. It is widely held that there is more than the mere indexical knowledge in Mary's epistemic progress. On the other hand, if one inflates Zombie Mary's epistemic progress, then one is committed to holding that as Mary acquires cognitively significant phenomenal knowledge, Zombie Mary acquire analogous cognitively significant knowledge. Chalmers rightly points out that this is not what happens when we conceive zombies. When we imaginarily subtract phenomenal consciousness from ourselves, we do not put something instead to where our phenomenal consciousness has been. We just take our phenomenal consciousness away, so that our zombie twins' inner life should be poorer than ours. Even if it is conceivable that our zombie twins have something analogous to our phenomenal consciousness, it is still conceivable that they do not. Then, PCS still in trouble, since even if our zombie twins have C, it cannot reductively explain their analogous cognitively significant knowledge. There arises another epistemic gap between our zombie twins' C and their analogous cognitive significant knowledge.

I will not delve into details of Chalmers' responses because they are irrelevant to my argument. The problem of his objection is that it relies on the same intuition I have already dissolved: there is or can be a difference between our epistemic situation and our zombie twins' epistemic situation. In dealing with the case of Mary and Zombie Mary, I noticed that while

the intuition seems strong at first glance, its force would be weakened or even neutralized when one bears in mind the constraint of topic-neutrality. Due to the constraint, both Mary and Zombie Mary cannot keep their epistemic situation topic-neutral without sharing it with her counterparts. They must reach epistemic equilibrium. Arguing for their epistemic equilibrium, I make no claims about the *amount* of knowledge that Mary and Zombie Mary gains when they exit the black-and-white room. My argument only involves the *language* characterizing their epistemic situations. It is also worth noting that my argument does not preclude that Mary acquires cognitively significant non-indexical knowledge when she exits her room. Mary acquires such knowledge. In this sense, my argument is not deflationary. Instead, all that I argue is *conditional*: *when* epistemic situations are characterized in topic-neutral terms, *if* Mary gains such cognitively significant knowledge, it must be shared by Zombie Mary. Further, I am not arguing that Zombie Mary acquires a certain cognitively significant non-indexical knowledge after seeing red. Thus, I am not committed to the inflationary strategy either. What I have to argue is that under the topic-neutral understanding of epistemic situations, if Zombie Mary does not acquire some cognitive significant non-indexical knowledge, Mary would not acquire such knowledge too. Therefore, my argument for epistemic equilibrium is neither inflationary nor deflationary. If epistemic situations assumed to be given an appropriate characterization in topic-neutral terms, then there is really no need to deflate Mary's knowledge or inflate Zombie Mary's knowledge. Thus my argument is not susceptible to Chalmers' criticisms against the deflationary and inflationary strategy. I see no need to deflate what Mary would know, or inflate what Zombie Mary would know. I simply claim that under the topic-neutral understanding of their epistemic situations, there cannot be any gap to be bridged by deflation or inflation of knowledge. I think this is a virtue of my argument.

Last, one might ask whether Chalmers can argue that our zombie twins do not share our epistemic situation and hold that our epistemic situation can be fully or exhaustively understood in topic-neutral terms. I do not see how this is even possible. *Ex hypothesi*, the *only* difference between us and our zombie twins is the fact that we do have phenomenal consciousness while our zombie twins do not. This means that their epistemic situations can differ is only if a full characterization of epistemic situations involves phenomenal consciousness. By assumption, however, this is simply impossible as far as the constraint of topic-neutrality holds. Under the topic-neutral characterization of epistemic situations, one cannot argue that our epistemic situations differ from our zombie twins, and *vice versa*. As a result, there can be no difference between our epistemic situation and that of our zombie twins. Unless Chalmers gives up the constraint of topic-neutrality, he cannot insist that zombies do not share our epistemic situation.

To conclude, PCS does not lose its explanatory potential. If we and our zombie twins share their full epistemic situation, by the epistemic equilibrium, PCS can physically account for our epistemic situation. Therefore, type-B materialists can hold the second horn of the master argument.

4.3 The Anti-Master Argument

Given the arguments so far, now I provide an argument against the master argument. The argument can be formalized as follows:

(A1) If our epistemic situation is not cast in topic-neutral terms, the master argument begs the question of whether PCS can explain our epistemic situation.

(A2) If our epistemic situation is cast in topic-neutral terms, the master

argument fails to show that PCS cannot explain our epistemic situation.

(A3) Therefore, either the master argument begs the question of whether PCS can explain our epistemic situation, or it fails to show that PCS is doomed to fail.

The argument has the form of a dilemma and seems valid. Each premise constitutes one of its horns. If both premises are successfully defended, the anti-master argument will undermine the master argument. Accordingly, PCS would be safe from the master argument and still be worth to pursue.

Premise (A1) is the first horn of the dilemma. It says that if our epistemic situation is not constrained to be topic-neutral, the master argument begs the question of whether PCS can explain our epistemic situation. If our epistemic situation allows phenomenal beliefs that referentially, causally, or inferentially require non-physical phenomenal properties, our epistemic situation itself builds in claims of the presence of non-physical phenomenal properties. If so, there is no way for PCS to explain our epistemic situation. PCS is originally designed for type-B materialism to explain our epistemic situation without appealing to any non-physical phenomenal properties. Insofar as PCS denies phenomenal properties as such, it cannot explain our epistemic situation including phenomenal beliefs. In this sense, once our epistemic situation is not cast in topic-neutral terms, the master argument begs the question against PCS: it must presuppose the failure of PCS in explaining our epistemic situation.

Premise (A2) is the second horn. Under the topic-neutrality, the master argument cannot go through. I have argued in the previous section that there is no epistemic difference between us and our zombie twins. There are doxastic, veridical, and epistemic symmetries between us and zombies. By this threefold symmetry, we and zombies reach an epistemic equilibrium.

This equilibrium implies that we and zombies perfectly share every component of their epistemic situations. Zombies' sharing of our epistemic situation directly falsifies the central premise of the master argument, namely (M3). Therefore, if our epistemic situation is constrained to be phenomenally neutral, the master argument fails to show PCS can explain our epistemic situation, as one of its premises turns out to be wrong.

Overall, Chalmers' master argument fails. The master argument is a dilemma for all possible forms of PCS. Nonetheless, if my argument in this section is on the right track, the table would be turned. It is the master argument that faces a dilemma from the constraint of topic-neutrality on our epistemic situation. On the one hand, if our epistemic situation is not characterized topic-neutrally, Chalmers must presuppose the phenomenal concepts or beliefs that necessarily refer to non-physical phenomenal properties. Since PCS only admits physical properties and denies non-physical properties, this begs the question against PCS. On the other hand, if our epistemic situation is characterized topic-neutrality, PCS can explain our epistemic situation and type-B materialists can hold the second horn of the dilemma. In one way or another, the master argument against PCS cannot succeed. Therefore, type-B materialists still can have the third way to reply to the zombie argument.

4.4 Conclusion

From Chapter 1 to 4, I have examined the central notions and premises of the argument. If they are successful, all the works in those chapters will lead to a fourfold argument against the zombie argument: 1) the zombie argument is based on the problematic notion of conceivability. 2) Even if the notion of conceivability is accepted, the first premise of the zombie argument, the conceivability of zombies, is wrong. 3) Even if the first premise of the zombie argument is right, the second premise, namely CP+,

is wrong. 4) Even if the second premise is right, it does not guarantee that zombies are metaphysically possible. Therefore, the zombie argument fails.

First, the notion of conceivability supposed by the zombie argument is problematic. The zombie argument supposes that zombies are ideally, positively, and primarily conceivable. In Chapter 1, I have shown that ideal conceivability is problematic in that its two possible formulations face their own problems. Positive conceivability is questionable in that it is too intuition-sensitive. Knowing whether zombies are positively conceivable requires a complete theory of qualia, which is not given yet. As the notion of conceivability plays essential roles in the zombie argument, these problems shake the argument to its ground. The zombie argument may not be able to get off the ground.

Second, even if the notion of conceivability is clarified and well defended, the first premise of the zombie argument is false. Zombies are not ideally positively primarily conceivable. All the sections in Chapter 2 cover my long and hard *reductio* against the conceivability of zombies. The real consequence of the conceivability of zombies is the disjunction of Qualia epiphenomenalism, Russellian monism, and interactionist dualism. I have argued that all of the disjuncts are wrong, as they necessarily commit to the negative conceivability of negatively inconceivable scenarios. So qualia epiphenomenalism, Russellian monism, and interactionist dualism are wrong. Therefore, the conceivability of zombies is wrong.

Third, even if the first premise of the zombie argument is true, the second premise is false. That is, even if zombies are conceivable, they are not primarily possible. The second premise is justified by CP+. In Section 3.2.2, I have provided the Russellian illuminati argument against CP+. The Russellian illuminati argument is a *reductio* against CP+. It draws a contradictory conclusion that Russellian zombies are secondarily (metaphysically) possible and impossible. This yields a paradoxical result

that if CP+ is right, it is wrong. If my Russellian version of anti-zombie argument is sound, CP+ would be self-defeating. In other words, ideal positive primary conceivability cannot be a guide to primary possibility. If so, even if zombies are ideally positively primarily conceivable, there is no guarantee that they are primarily possible.

Finally, even if the second premise of the zombie argument is right, zombies can be metaphysically impossible. In Section 4.2.1, I have argued that PCS can be the third way for physicalists to argue against the zombie argument. According to Chalmers' master argument, however, as far as phenomenal concepts are physically explainable, they cannot explain our epistemic situation. I have argued that the constraint of topic-neutrality, both we and our zombie twins share their epistemic situation. Since PCS can explain our zombie twins' epistemic situation, it can explain our epistemic situation either. Thus, the master argument fails.

Bibliography

Alter, T. and Nagasawa, Y. 2015, *Consciousness in the Physical World: Perspectives on Russellian Monism*. Oxford: Oxford University Press.

———————————. 2015, What is Russellian Monism? In Alter, T. and Nagasawa, Y. (eds.), *Consciousness in the Physical World: Perspectives on Russellian Monism*, pp. 422-452, Oxford: Oxford University Press.

Alter, T. and Walter, S. 2007, *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. New York: Oxford University Press.

Balog, K. Illuminati, zombies and metaphysical gridlock (unpublished manuscript) http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.593.6410&rep=rep1&type=pdf Retrieved June 9, 2019

Benacerraf, P. and Putnam, H. 1983, *Philosophy of Mathematics: Selected Readings.* Cambridge: Cambridge University Press.

Bernays, P. 1983, On Platonism in Mathematics, In Benacerraf, P. and Putnam, H. (eds.) *Philosophy of Mathematics: Selected Readings,* pp. 258-271, Cambridge: Cambridge University Press.

Block, N. 2007, Consciousness, Accessibility, and the Mesh between Psychology and Neuroscience. *Behavioral and Brain Sciences* 30: 481-548.

Block, N., Flanagan, O. and Güzeldere, G. 1997, *The Nature of*

Consciousness: Philosophical Debates, Cambridge. Mass: MIT Press.

Bourget, D. 2017, Is Emergent Anomalous Panpsychism Viable? (manuscript), http://www.dbourget.com/papers/anomalous_panpsychism.pdf Retrieved June 9, 2019

Brouwer, L. E. J. 1975, *Collected Works, Vol. 1: Philosophy and Intuitionistic Mathematics*. Amsterdam: North-Holland.

Brown, R. 2010, Deprioritizing the A Priori Arguments Against Physicalism, *Journal of Consciousness Studies* 17: 47-69.

_____. 2013, The Two-Dimensional Argument Against Dualism, https://philarchive.org/archive/BROTTA-6 Retrieved June 9, 2019

Brüntrup, G. and Jaskolla, L. 2017, *Panpsychism: Contemporary Perspectives.* Oxford: Oxford University Press.

Carruthers, P. 2004, Phenomenal Concepts and Higher-order Experiences. *Philosophy and Phenomenological Research* 68(2): 316-36.

Carruthers, P. and Veillet, B. 2007, The phenomenal concept strategy. *Journal of Consciousness Studies* 14(9-10): 212-236.

Cartwright, N. 1983, *How the Laws of Physics Lies?* Oxford: Oxford University Press.

Chalmers, D. J. 1996, *The Conscious Mind*. Oxford University Press.

————————. 1999, Materialism and the Metaphysics of Modality. *Philosophy and Phenomenological Research* 59(2): 473-496.

_____. 2002, Does Conceivability Entail Possibility? In Gendler, T. and Hawthorne, J. (eds), *Conceivability and Possibility*, pp. 145-200, Oxford University Press.

———————. 2003, The Content and Epistemology of Phenomenal Belief, In Smith, Q. and Jokic, A. (eds.), *Consciousness: New Philosophical Essays*, pp. 220-272, Oxford: Oxford University Press.

———————. 2004, Imagination, Indexicality, and Intension. *Philosophy and Phenomenological Research* 68(1): 182-190.

_____. 2006, The foundations of two-dimensional semantics, In Garcia Carpintero, M. and Macia, J. (eds.) *Two-Dimensional Semantics: Foundations and Applications*, pp. 55-140, Oxford: Oxford University Press.

———————. 2010, *The Characters of Consciousness*. Oxford: Oxford University Press.

———————. 2015, Panpsychism and Panprotopsychism, In Alter, T and Nagasawa, Y. (eds.), *Consciousness in the Physical World: Perspectives on Russellian Monism*, pp. 247-276, Oxford: Oxford University Press.

———————. 2016, The Combination Problem of Panpsychism, In Jaskolla, L. and Brüntrup, G., *Panpsychism: Contemporary Perspectives,* pp. 179-214, Oxford: Oxford University Press.

_____. 2017, Panpsychism and Panprotopsychism, In Brüntrup, G. and Jaskolla, L. (eds.), *Panpsychism: Contemporary Perspectives,* pp. 19-47, Oxford: Oxford University Press.

Chalmers, C. and Jackson, F. 2001, Conceptual analysis and reductive explanation. *Philosophical Review* 110(3): 315-61

Chisholm. R. 1989, *Theory of Knowledge*. 3rd edition. New Jersey: Printice-Hall.

Coleman, S. 2012, Mental Chemistry: Combination for Panpsychists. *Dialectica* 66: 137-166.

Dantas, D. F. 2017, Ideal Reasoners don't believe in zombies. *Principia* 21(1): 41-59.

Davies, M. and Humphreys, G. W. 1993, *Consciousness: Psychological and Philosophical Essays*. Malden: Blackwell Publishing.

Dennett, D. 1988, Quining Qualia, In Marcel, A. and Bisiach, E. (eds.), *Consciousness in Contemporary Science*, pp. 43-77, Oxford: Oxford University Press. Reprinted in Lycan, W. and Prinz, J. (eds.) 2008, *Mind and Cognition: An Anthology*, 3rd ed. Oxford: Blackwell.

De Paul, M. 2001, *Resurrecting Old-Fashioned Foundationalism,* pp. 3-20, Lanham, MD: Rowman & Littlefield.

Elpidorou, A. 2013, *Having it Both Ways: Consciousness, Unique Not Otherworldly*. *Philosophia* 41(4): 1181-1203.

Eccles, J. C. 1986. Do mental events cause neural events analogously to the probability fields of quantum mechanics? *Proceedings of the Royal Society of London* B227: 411-428.

Frankish, K. 2007, The Anti-Zombie Argument. *The Philosophical Quarterly* 57(229): 650-666

_____. 2012, Quining diet qualia. *Consciousness and Cognition* 21(2): 667–676.

Fumerton, R. 1985, *Metaphysical and Epistemological Problems of Perception.* Lincoln: University of Nebraska Press.

———————. 1995, *Metaepistemology and Skepticism*. Lanham, MD: Rowman and Littlefield.

———————. 2001, Classical Foundationalism, In De Paul, M. (ed.), *Resurrecting Old-Fashioned Foundationalism,* pp. 3-20, Lanham, MD: Rowman & Littlefield.

Garcia Carpintero, M. and J. Macia, J. 2006, *Two-Dimensional Semantics: Foundations and Applications.* Oxford: Oxford University Press.

Gendler, T. and Hawthorne, J. 2002, *Conceivability and Possibility.* Oxford: Oxford University Press.

Gertler, B. 2001, Introspecting Mental States. *Philosophy and Phenomenological Research* 63: 305-328.

————. 2011, *Self-Knowledge*. New York: Routledge.

————. 2012, Renewed Acquaintance, In Smithies, D. and Stoljar, D. (eds.), *Introspection and Consciousness*, pp. 93-128, Oxford: Oxford University Press.

Gödel, K. 1983, What is Cantor's Continuum Problem? In Benacerraf, P. and Putnam, H. (eds.) *Philosophy of Mathematics: Selected Readings*, pp. 470-485, Cambridge: Cambridge University Press.

Harman, G. 1973, *Thought.* Princeton, NJ: Princeton University Press.

Hasan, A. 2011, Classical Foundationalism and Bergmann's Dilemma for Internalism, *Journal of Philosophical Research* 36: 391–410.

————. 2013, Phenomenal Conservatism, Classical Foundationalism, and Internalist Justification. *Philosophical Studies* 162: 119–141.

Hawking, S. 1988, *A Brief History of Time.* New York: Bantam.

Hilbert, D. 1983, On the Infinite, In Benacerraf, P. and Putnam, H. (eds.) *Philosophy of Mathematics: Selected Readings*, pp. 183-201, Cambridge: Cambridge University Press.

Hill, C. 1997, Imaginability, Conceivability, Possibility, and the Mind-body Problem. *Philosophical Studies* 87: 61-85.

Hill, C. and McLaughlin, B. P. 1999, There Are Fewer Things in Reality Than Are Dreamt of in Chalmers's Philosophy. *Philosophy and Phenomenological Research* 59: 445-54.

Holman, E. 2008, Panpsychism, Physicalism, Neutral Monism and the Russellian Theory of Mind, *Journal of Consciousness Studies* 15: 48-67.

Ismael, J. 1999, Science and the Phenomenal. *Philosophy of Science* 66: 351-69.

James, W. 1890/1950, *The Principles of Psychology*, Vols. 1&2. New York: Dover Publications.

Jaskolla, L. and Brüntrup, G. 2016, *Panpsychism: Contemporary Perspectives.* Oxford: Oxford University Press.

Kim, J. 1998, *Mind in a Physical World*. A Bradford Book.

_____. 2008, *Physicalism or Something Near Enough*. Princeton University Press.

Klein, P. 1971, A proposed definition of Propositional Knowledge. *The Journal of Philosophy*, 68: 471-82

_____. 1976, Knowledge, Causality, and Defeasibility. *The Journal of Philosophy*, 73: 792-812

Kriegel, U. 2009. *Subjective Consciousness: A Self-representational Theory.* New York: Oxford University Press.

Lehrer, K. and Paxon, T. 1969, Knowledge: Undefeated Justified True Belief. *The Journal of philosophy* 66: 225-237

Levin, J. 2007, What Is a Phenomenal Concept? In Alter, T. and Walter, S. (eds.), *Phenomenal*

Concepts and Phenomenal Knowledge: Essays on Consciousness and Physicalism, pp. 87-110, New York: Oxford University Press.

Levine, J. 2011, Review of The Character of Consciousness, https://ndpr.nd.edu/news/the-character-of-consciousness/ Retrieved June 9 2019

Loar, B. 1990/1997, Phenomenal States. *Philosophical Perspectives* 4: 81-108. Revised in Block, N., Flanagan, O. and G. Güzeldere, G. (eds.), *The Nature of Consciousness: Philosophical Debates*, pp. 597-615 Cambridge, Mass: MIT Press.

Ludlow, P., Yujin Nagasawa, Y. and Stoljar, D. 2004, *There's*

Something about Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument. Cambridge: MIT Press.

Lycan, W. and Prinz, J. 2008, *Mind and Cognition: An Anthology*, 3rd ed. Oxford: Blackwell.

Mill, J. S. 1848, *A System of Logic, Ratiocinative and Inductive; being a connected view of the principles of evidence and the methods of scientific investigation*. New York: Harper & Brothers.

Montero, B. G. 2014, Russellian Physicalism, In Alter, T. and Nagasawa, Y. (eds.), *Consciousness in the Physical World: Perspectives on Russellian Monism*, pp. 209-223, Oxford: Oxford University Press.

Nagel, T. 1974, What Is It Like to Be a Bat? *The Philosophical Review* 83(4): 435-450

————. 1979a, Panpsychism, In Nagel, T. *Mortal Questions*, pp. 181-185, Cambridge: Cambridge University Press.

————. 1979b, *Mortal Questions*. Cambridge: Cambridge University Press.

Newell, A. and Simon, H. A. 1976, Computer Science as Empirical Inquiry: Symbols and Search. *Communications of the ACM* 19(3):113-126.

O'Dea, J. 2002, The Indexical Nature of Sensory Concepts. *Philosophical Papers* 31: 169–81.

O'Hear, A. 2003, *Minds and Persons*. New York: Cambridge

University Press.

Papineau, D. 2002, *Thinking about Consciousness.* New York: Oxford University Press.

———— . 2007. Phenomenal and Perceptual Concepts, In Alter, T. and Walter, S. (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism,* pp. 111-144, New York: Oxford University Press.

Pereboom, D. 2011, *Consciousness and the Prospects of Physicalism*. New York: Oxford University Press.

Perry, J. 2001, *Knowledge, Possibility, and Consciousness*. MIT Press.

Pollock, J. L. 1987, Defeasible Reasoning. *Cognitive Science* 11: 481-518

_____. 1990, *Nomic Probability and the Foundations of Induction.* Oxford: Oxford University Press.

_____. 1995. *Cognitive Carpentry: a blueprint for how to build a person*. MIT Press.

Prior, E. W. Pargetter, R. and Jackson, F. 1982, Three Theses about Dispositions. *American Philosophical Quarterly* 19(3): 251-257.

Pust, J. 2017, Intuition. Stanford Encyclopedia of Philosophy https://plato.stanford.edu/entries/intuition/ Retrieved June 9, 2019

Rosenthal, D. 1986, Two concepts of consciousness. *Philosophical Studies* 49: 329–359.

_____. 1993, Thinking that one thinks, In Davies, M. and

Humphreys, G. W. (eds.), *Consciousness: Psychological and Philosophical Essays*, pp. 197-223, Malden: Blackwell Publishing.

———————. 2005, *Consciousness and Mind*. Oxford: Oxford University Press.

Russell, B. 1910–11, Knowledge by Acquaintance and Knowledge by Description. *Proceedings of the Aristotelian Society* 11: 108–128.

—————. 1912, *The Problems of Philosophy*. Oxford: Oxford University Press.

—————. 1914, On the Nature of Acquaintance. *Monist* 24:161–187.

—————. 1959, *My Philosophical Development*. London: George Allen & Unwin. (References to the London: Unwin Books 1975 edition).

Schroer, R. 2013, Do the Primary and Secondary Intensions of Phenomenal Concepts Coincide in all Worlds? *Dialectica* 67(4): 561-577

Seager, W. E. 1995, Consciousness, information, and panpsychism. *Journal of Consciousness Studies* 2: 272-288.

———————. 2010, Panpsychism, aggregation and combinatorial infusion. *Mind and Matter* 8:167-184.

———————. 2016a, *Theories of consciousness: an introduction and assessment*, 2nd Edition. Routledge.

———————. 2016b, Panpsychist infusion, In Jaskolla, L. and Brüntrup, G. (eds.), *Panpsychism: Contemporary Perspectives*, pp.

229-248, Oxford: Oxford University Press.

Smith, Q. and Jokic, A. 2003, *Consciousness: New Philosophical Essays.* Oxford: Oxford University Press.

Smithies, D. and Stoljar, D. 2012, *Introspection and Consciousness*. Oxford: Oxford University Press.

Stapp, H. P. 1993, *Mind, Matter, and Quantum Mechanics*. Berlin: Springer-Verlag.

Steiner, M. 1975, *Mathematical Knowledge.* Ithaca, New York: Cornell University Press.

Stoljar, D. 2005, Physicalism and Phenomenal Concepts. *Mind and Language* 20: 296-302.

Strawson, G. 1989, Red and 'red'. *Synthese* 78: 193-232

————————. 2006, Realistic Monism: Why Physicalism Entails Panpsychism. *Journal of Consciousness Studies* 13: 3-31.

Swain, M. 1974, Epistemic Defeasibility. *The American Philosophical Quarterly* 11(1): 15-25

Tieszen, R. L. 1989, *Mathematical Intuition*. Springer.

Tye, M. 2003a, A Theory of Phenomenal Concepts, In O'Hear, A. (ed.), *Minds and Persons*, pp. 91-106, New York: Cambridge University Press.

_____. 2003b. Blurry Images, Double Vision, and Other Oddities: New Problems for Representationalism? In Smith, Q. and Jokic, A. (eds.), *Consciousness: New Philosophical*

Perspectives, pp. 7-32, New York: Oxford University Press.

van Gelder, T. J. 1999, Dynamic Approaches to Cognition, In Wilson, R. and Keil, F. (eds.), *The MIT Encyclopedia of Cognitive Sciences*, pp. 244-246, Cambridge MA: MIT Press.

van Heijenoort, J. 1967, *From Frege to Gödel: A Source Book in Mathematical Logic.* Cambridge: Harvard University Press.

Wilson, R. and Keil, F. 1999, *The MIT Encyclopedia of Cognitive Sciences*. Cambridge MA: MIT Press

Yablo, S. 1993, Is conceivability a guide to possibility? *Philosophy and Phenomenological Research* 53:1-42.

국문초록

본 논문의 목적은 철학자 데이비드 차머스에 의해 개발된 좀비 논변을 재검토하는 것이다. 좀비 논변은 물리주의에 반대하는 상상가능성 논증으로 알려져 있다. 논변은 우리와 물리적으론 동일하지만 우리 경험의 현상적 느낌, 또는 감각질을 결여한 존재의 상상가능성을 주장하면서 시작한다. 차머스는 좀비의 상상가능성으로부터 그것의 가능성을 끌어낸다. 좀비가 가능하다면 감각질은 물리적 사실들에 수반하지 않는다. 물리주의가 심신 수반을 함축한다는 것은 널리 수용되고 있기에, 좀비의 가능성은 모든 가능한 형태의 물리주의를 논박하게 된다. 그 크나큰 함축으로 인해, 좀비 논변은 의식의 본성과 물리주의에 대한 격렬한 논쟁을 불러일으켰다. 실로 그것이 처음 제시된 이후로 좀비 논변은 논쟁적이기를 그친 적이 없다. 좀비 논변에는 의미론, 형이상학, 인식론과 관련된 까다로운 문제들이 뒤얽혀있다. 본 논문에서, 나는 좀비 논변의 핵심 개념들과 전제들을 비판적으로 검토하고 관련된 문제들을 탐색해 볼 것이다.

본고는 4장으로 구성된다. 1장은 좀비 논변에서 사용되는 상상가능성 개념을 다룬다. 좀비 논변에서 상정된 상상가능성 개념은 문제적이다. 좀비의 상상가능성과 관련하여 차머스는 몇 가지 구분을 제시한다: 일견적/이상적, 적극적/소극적, 일차적/이차적 상상가능성이 그것이다. 이상적 상상가능성의 개념이 문제적인 것으로 밝혀진다. 그것이 잘 작동할지 의심스럽기 때문이다. 좀비의 적극적 상상가능성 또한 의심스럽다. 그것은 지나치게 직관에 민감하고 또한 감각질에 대한 완전한 이론을 요구하는데, 그런 이론은 아직 주어지지 않았다. 만약 상상가능성의 개념이 문제적이라면 좀비 논변은 시작조차 할 수 없게 된다.

2장은 좀비 논변의 첫 번째 전제, 즉 좀비의 이상적 적극적 일차적 상상가능성에 대한 나의 귀류 논변을 다룬다. 좀비의 상상가능성은 감각질 부수현상론, 러셀일원론, 그리고 상호작용론적 이원론의 선언을 함축한다. 이 모든 선언지들이 틀렸음을 보이기 위해, 나는 감각질의 인지적 친밀성을 논증했다: 의식적 경험의 현상적 질은 훼손되지 않은 배경조건 하에서, 경험의 주체에 의해 반드시 잠재적으로 주의를 받거나 알아차려질 수 있어야만 한다. 인지적 친밀성은 선험적으로 참이며, 따라서 인지적으로 친밀하지 않은, 소외된 감각질이란 소극적으로조차 상상불가능하다. 그러나 모든 선언지들은 인지적으로 소외된 감각질의 소극적 상상가능성에 개입한다. 귀류에 의해, 좀비는 이상적 적극적 일차적으로 상상불가능해진다. 만약 이러한 귀류논변이 통한다면, 설사 이상적 적극적 일차적 상상가능성 개념이 옹호될 수 있더라도, 좀비 논변의 첫째 전제는 거짓이 된다.

3장에서는 좀비 논변의 두 번째 전제가 검토된다. 두 번째 전제는 이상적 적극적 일차적 상상가능성은 일차적 상상가능성을 함축한다는 원리(CP+)의 한 적용이다. CP+에 맞서, 몇몇 철학자들은 좀비 논변을 패러디하려 했다. 그러나 이러한 반-좀비 논변들은 그들 나름의 문제가 있다. 러셀 일루미나티 논변은 내 버전의 반-좀비 논변으로 그러한 문제들을 피할 수 있다. 러셀 일루미나티 논변은 CP+를 전제하고 그로부터 모순을 끌어낸다. 만약 이 논변이 건전하다면, 이상적 적극적 일차적 상상가능성은 일차적 가능성에 대한 가이드가 될 수 없을 것이다. 따라서, 설사 좀비가 이상적 일차적 적극적으로 상상가능하더라도, 그것이 일차적으로 가능하리라는 보장은 없다. 좀비 논변의 첫째 전제가 참이라도, 두 번째 전제는 거짓인 것이다.

4장은 좀비 논변에 대한 물리주의자들이 또다른 대응, 현상적 개념 전략(PCS)을 다룬다. 어떤 철학자들은 현상적 개념의 특수한 본성에 호소하면서 설명적 간극이나 좀비의 상상가능성 등을 포함한, 의식과 관련된 우리의 문제적인 인식적 상황들을 설명해치워 버리려 시도한 바 있다. PCS에 의존하여, 물리주의자들은 좀비 논변의 핵심 전제들을 받아들이면서도 그 결론은 부정할 수 있다. 그러나 차머스의 만능 논변에 따르면, 현상적 개념이 물리적으로 해명가능한 한 그것은 우리의 인식적 상황을 설명할 수 없다. 만능 논변에 맞서, 나는 인식적 상황이 주제-중립적 용어에 의해 특성화되어야 하는 한 PCS는 그 설명적 잠재력을 유지할 수 있음을 논증했다. 그러므로 만능 논변은 실패하며 오히려 그 나름의 딜레마에 봉착하게 된다. PCS는 여전히 물리주의자들에게 살아있는 선택지인 것이다. 즉 설사 좀비 논변의 첫째 전제와 둘째 전제가 옳다고 하더라도, 물리주의자들에게는 좀비 논변을 거부할 수 있는 '제3의 길'이 있는 셈이다.

만약 본고에서 제시된 논변들이 성공적이라면, 그 모든 작업들은 좀비 논변에 대한 4중의 비판으로 귀결될 것이다: 1) 좀비 논변은 상상가능성에 대한 문제적인 개념에 기반을 두고 있다. 2) 설사 상상가능성의 개념이 수용되더라도, 좀비 논변의 첫째 전제, 즉 좀비의 상상가능성은 틀렸다. 3) 설사 좀비 논변의 첫째 전제가 옳다고 하더라도, 둘째 전제가 틀렸다. 4) 설사 첫째와 둘째 전제가 모두 옳다고 해도, 이것이 좀비가 형이상학적으로 가능하다는 것을 보장해주진 않는다. 따라서, 좀비 논변은 실패한다.

주요어 : 의식, 감각질, 상상가능성, 가능성, 데이비드 차머스, 좀비 논변

학 번 : 2012-30825