【Paper (論文)】

The Meta-Modal Argument Against Dualism

Reijiro YAUTA (矢歌 礼次郎)[1]

Abstract

This paper critically examines the "***meta-modal argument***," a physicalist conceivability argument that seeks to refute the anti-physicalist "***zombie argument***" in the metaphysics of consciousness (Marton 1998, Sturgeon 2000, etc.). While the meta-modal argument presupposes the conceivability of physicalism to counter the zombie argument, this paper argues that its responses to key objections remain insufficient. The analysis advances three central claims: (1) the conceivability of zombies concerns zombies themselves, not merely their possibility; (2) a key model employed to undermine the zombie argument within S5 framework is metaphysically inappropriate; and (3) the ideal conceivability of physicalism remains indeterminate due to the persistence of the hard problem of consciousness.

Keywords: Conceivability argument, The meta-modal argument, The zombie argument, The hard problem of consciousness, Physicalism, Dualism

1. Introduction

The primary aim of this paper is to criticize the meta-modal argument within the metaphysics of consciousness. The meta-modal argument is a physicalist conceivability argument that presupposes, implicitly or explicitly, the premise that the modal claim

1 The University of Tokyo, Graduate School of Arts and Sciences, Master's Program. (東京大学総合文化研究科修士課程。)

of physicalism is conceivable and attempts to attack the zombie argument. If the meta-modal argument is successful, it would undermine the zombie argument, as both rely on the same underlying principle, the *Conceivability-Possibility Thesis* (CP thesis). This criticism thus seeks to evaluate whether the meta-modal argument can withstand scrutiny and whether it effectively challenges the zombie argument.

This paper is organized as follows. In Section 2, I outline the theoretical background of the meta-modal argument, beginning with an explanation of the zombie argument and its reliance on the CP thesis, followed by a detailed exposition of the meta-modal argument as formulated by Peter Marton (1998). Section 3 discusses major criticisms of the meta-modal argument, while Section 4 examines physicalist responses to these objections. In Section 5, I criticize these responses and present my original argument against the meta-modal argument. At this point, it should be noted that the argument remains neutral in the conflict between dualism and physicalism. This paper merely highlights that some physicalist responses to dualistic objections are not justified. Finally, in Section 6, I propose directions for future research based on the findings of this paper.

2. Background and Prior Studies

2-1. The Zombie Argument

The debate between the zombie argument and the meta-modal argument concerns the truth or falsity of *physicalism* about consciousness. Generally, physicalism is the idea that “any world which is a *minimal* physical duplicate of our world is a duplicate *simpliciter* of our world” (Jackson 1998, p. 12). Thus, according to physicalism, because phenomenal facts——e.g. “someone has consciousness “ or “someone feels pain”——hold in the minimal physical duplicate as well, physical facts necessitate the existence of phenomenal consciousness. In contrast, *dualism* posits that

physical facts do not necessitate the existence of phenomenal consciousness. Formally, if "P" stands for the totality of microphysical truths, "Q" stands for any phenomenal truth, the two positions can be defined as follows:

- **Physicalism**: $\Box (P \supset Q)$
- **Dualism**: $\neg \Box (P \supset Q)$

Dualism often resorts to the *zombie argument* to refute physicalism. Let us now introduce the *zombie argument* proposed by Chalmers (1996, 2010). Using " $\blacklozenge$ " as the conceivability operator, the argument can be formalized as follows:

(1) $\blacklozenge (P \wedge \neg Q)$	Conceivability of zombies
(2) $\blacklozenge (P \wedge \neg Q) \supset \Diamond (P \wedge \neg Q)$	An instance of the CP thesis
(3) $\underline{\Diamond (P \wedge \neg Q) \supset \neg \Box (P \supset Q)}$	Logical entailment
$\therefore \quad \neg \Box (P \supset Q)$	Falsity of physicalism

Here, "$P \wedge \neg Q$" is the *zombie* claim that expresses the scenario where all microphysical truths hold, yet a particular phenomenal truth does not.[2] Therefore, Premise (1) asserts the conceivability of zombies. Premise (2) states that the conceivability of zombies implies their metaphysical possibility. This premise relies on the *Conceivability-Possibility Thesis (CP thesis)*, proposed by Chalmers (2002, 2010), which can be formulated as follows:

- **CP**: For any proposition S, $\blacklozenge S \supset \Diamond S$

Here, according to Chalmers, the conceivability of S must be *ideal*. Accordingly,

2 Here, "zombie" refers to *philosophical zombies*, not Hollywood-style zombies. Generally, a philosophical zombie is a being (or a scenario, or a world) that has the same physical properties as us (or our world) but lacks consciousness.

A proposition S is ideally conceivable if it cannot be ruled out a priori under conditions of rational reflection free from cognitive limitations. This is because, if a proposition is merely prima facie conceivable, the proposition might not entail the possibility of the proposition. For example, it is apparently conceivable that Fermat's last theorem is false, but since the theorem is in fact true, the conceivability of its falsity does not entail possibility of its falsity. To exclude such cases, the conceivability that entails possibility must be ideal, meaning it must be *undefeatable*. By contrast, a proposition is prima facie conceivable if it is conceivable at first impression. Based on these notions, the falsity of Fermat's Last Theorem is not ideally conceivable but only prima facie conceivable, because the theorem is provable through further reasoning. In this paper, unless explicitly stated otherwise, "conceivability" refers to ideal conceivability.

Finally, Premise (3) asserts that if a zombie is possible, then physicalism is false. This is because the possibility of a zombie implies that there is at least one possible world where phenomenal facts are not necessitated by physical facts, which in turn means that physicalism is false. Thus, Premise (3) is true. Accordingly, from Premises (1)-(3), we can logically infer the falsity of physicalism through recursive applications of syllogism.

2-2. The Meta-Modal Argument

The zombie argument has faced various criticisms, among which is a strategy that employs the same principle on which the zombie argument relies—the CP thesis—to criticize dualism (Yablo 1999, Marton 1998; Sturgeon 2000; Frankish 2007; Brown 2010, 2013; Balog 2012; Campbell et al. 2017; Heikinheimo & Vaaja 2013; Piccinini 2017, etc.). Arguments constructed using this strategy are referred to as *mirror arguments* or *parody arguments*. Among the several types of *mirror arguments*, this

paper focuses on the meta-modal argument, which deals with modal conceivability. [34]

This paper centers its discussion on the meta-modal argument proposed by Peter Marton (1998). Abstracting Marton's reasoning, it can be formally represented as follows:

(1) $\blacklozenge (P \wedge \neg Q)$	Conceivability of zombies
(2) $\blacklozenge (P \wedge \neg Q) \supset \Diamond (P \wedge \neg Q)$	An instance of the CP thesis
(3) $\Diamond (P \wedge \neg Q) \supset \neg \Box (P \supset Q)$	Logical entailment
(4) $\neg \Box (P \supset Q) \supset \Box \neg \Box (P \supset Q)$	Theorem of S5
(5) $\Box \neg \Box (P \supset Q) \supset \neg \Diamond \Box (P \supset Q)$	Definition of necessity
(6) $\underline{\neg \Diamond \Box (P \supset Q) \supset \neg \blacklozenge \Box (P \supset Q)}$	Contrapositive instance of the CP thesis
$\therefore \neg \blacklozenge \Box (P \supset Q)$	Inconceivability of physicalism

Let us confirm the validity of each premise one by one. Premises (1), (2), and (3) correspond to those of the zombie argument. Premise (4) is a theorem of S5 modal logic. S5 is an axiomatic system of modal logic in which the proposition $\neg \Box A \supset \Box \neg \Box A$ holds as a theorem for any proposition A. This can be proven as follows: In modal logic, it is tautological that $\neg \Box A \Leftrightarrow \Diamond \neg A$. Moreover, in S5, the axiom $\Diamond A \supset \Box \Diamond A$ holds for any A (Axiom 5). Again, since $\neg \Box A \Leftrightarrow \Diamond \neg A$, it follows that

3 There are mirror arguments other than meta-modal argument, such as the *anti-zombie* argument (Frankish 2007, Brown 2010, 2013, Campbell et al. 2017), the *zoombie argument* (Brown 2010) and the *reductio argument* (Yablo 1999, Howell 2008, Mizrahi & Morrow 2015). Among these arguments, although Chalmers (2010) mainly refers to Yablo's reductio argument as the meta-modal argument, this paper primarily takes Marton's argument as the meta-modal argument.

4 The meta-modal argument is typically formulated as follows (Chalmers 2010):

(1) $\blacklozenge \Box (P \supset Q)$
(2) $\blacklozenge \Box (P \supset Q) \supset \Diamond \Box (P \supset Q)$
(3) $\underline{\Diamond \Box (P \supset Q) \supset \Box (P \supset Q)}$
(4) $\Box (P \supset Q)$

In this paper, however, let us formulate the meta-modal argument in accordance with Marton's original argument. This is because objections to the meta-modal argument are primarily directed at Marton's argument. It should be noted that both formulations remain relevant to the discussion in this paper.

¬ □ A ⊃ □ ¬ □ A. Finally, by substituting A with (P ⊃ Q), we obtain ¬ □ (P ⊃ Q) ⊃ □ ¬ □ (P ⊃ Q). Thus, Premise (4) is provable within the framework of S5.

Consequently, rejecting Premise (4) would require rejecting the framework of S5 itself, particularly its axioms or accessibility relations. However, this is not a promising strategy for dualists. This is because, since the accessibility relation in S5 is equivalence relation[5], denying S5 would allow for the following model: A zombie world exists but is inaccessible from the actual world, while P ⊃ Q holds true in all accessible worlds from the actual world. In such a model, the existence of a zombie world would not refute physicalism, which would be disadvantageous for dualists. To exclude this model from the outset, dualists must assume a model that relies on a universal accessibility relation. A universal relation is an accessibility relation in which all worlds in a set of possible worlds are connected to each other. To guarantee this type of accessibility relation, dualists have no choice but to accept S5. This is because, to refute physicalism using the zombie argument, one must assume that all models are universal. Since a universal model is an equivalence model, any proposition that holds in an arbitrary universal model must also hold in any equivalence model, i.e., it must be provable in S5. Therefore, dualists who wish to challenge physicalism using the zombie argument must accept all propositions that are provable in S5. Thus, dualists must accept Premise (4) as well.

Premise (5) asserts that if physicalism is necessarily false, then it is impossible for physicalism to be true. This follows logically from the definition of necessity ($\Box A =_{def} \neg \Diamond \neg A$) and cannot be rejected. Premise (6) claims that if physicalism is impossible, then it is inconceivable, which is an instance of the contrapositive of the CP thesis. Therefore, as long as the CP thesis is endorsed, Premise (6) cannot be rejected either.

From these premises—each of which dualists are compelled to accept—

5 An accessibility relation R is equivalence relation iff R is reflexive, transitive, and symmetric. That is, R is equivalence relation iff for all w, w', w'' wRw (reflexive), and, if wRw' and w'Rw'' then wRw'' (transitive), and if wRw' then w'Rw (symmetric).

it follows that physicalism is inconceivable. However, according to Marton, this conclusion is implausible. Recall that the basis of (ideal) conceivability is the absence of an explicit or explicable contradiction. Since the proposition of physicalism ($\Box$ (P $\supset$ Q)) does not appear to contain any logical or conceptual contradiction, *physicalism should be conceivable*. Regardless of its truth value, physicalism is at least conceivable. Hence, Marton argues that the zombie argument, which leads to the inconceivability of physicalism, ultimately fails.[6]

3. Objections against the Meta-Modal Argument

Several criticisms have been raised against the meta-modal argument described above. In this section, we will discuss some of these criticisms and present counterarguments in response.

3-1. Objection 1: Applicability of the CP Thesis to Modal Claims

One strategy (Frankish 2007; Chalmers 2010) argues that the CP thesis cannot be applied to modal claims. The basis of this objection lies in the asymmetry between the conceivability of zombies and the conceivability of physicalism. Specifically, in the case of the conceivability of zombie worlds, one only needs to conceive of a single world that satisfies P $\wedge$ $\neg$Q. In contrast, since physicalism is a necessary modal claim ($\Box$ (P $\supset$ Q)), conceiving of physicalism requires conceiving of all possible worlds belonging to a set of worlds that satisfy P $\supset$ Q, meaning one must conceive of the entire space of possible worlds satisfying P $\supset$ Q. However, it is unclear whether such thought is coherently possible. Since the meta-modal argument presupposes the conceivability of physicalism, proponents of the meta-modal argument must explain why such an act of conceiving is indeed feasible. With this in mind, dualists can claim that the meta-modal argument is no longer a mirror argument because there is an

6 The same point has been ponited out by VandenHombergh (2017).

asymmetry between the zombie and physicalism with respect to the contents of their conceivability.

3-2. Objection 2: Dispensability of S5 for the Zombie Argument

Another strategy (VandenHombergh 2020) attempts to demonstrate that the zombie argument does not necessarily rely on S5. If it can be shown that the zombie argument does not require the S5 framework, since the meta-modal argument obviously relies on S5 (cf. Premise (4)), the meta-modal argument loses its symmetry with the zombie argument and can be criticized for smuggling in assumptions favorable to physicalism. For example, consider the following model $M=<W,R,V>$:

- The set of possible worlds: $W=\{w_@, w_1, w_2, w_3\}$
- Accessibility relation: $R=\{<w_@, w_@>, <w_@, w_1>, <w_@, w_2>, <w_@, w_3>, <w_1, w_2>, <w_1, w_3>\}$
- Valuation: $V=\{<P, \{w_@, w_1, w_2, w_3\}>, <Q, \{w_@, w_1, w_2\}>\}$

Let $w_@$ represent the actual world. Here, w_2 and w_3 are accessible from w_1, but w_2 and w_3 are not mutually accessible. Thus, the accessibility relation is non-Euclidean[7] in this model, and Axiom 5 does not hold, so it is not a model of S5. Therefore, Premise (4) does not hold in this model, and the meta-modal argument fails. However, since all worlds are accessible from the actual world $w_@$, and w_3 (a zombie world) exists among those accessible worlds, the zombie argument remains valid. This demonstrates that the zombie argument does not necessarily depend on S5 just because all worlds must be accessible from the actual world.

3-3. Objection 3: Inconceivability of Physicalism

7 R is Euclidean iff for all w, w', w'', if wRw' and wRw'', then w'Rw''.

A third strategy asserts that physicalism is inherently inconceivable. For example, Chalmers (2010) claims that physicalism is, at best, prima facie conceivable, but not ideally conceivable. In other words, physicalism appears to be conceivable only at first glance, lacking deeper coherence upon closer examination. While this argument relies on intuition, the difficulty widely acknowledged in imagining the identity of phenomenal and physical properties lends some credibility to this intuition-based inconceivability.

Additionally, Cutter (2023) argues that if physicalism is understood as a claim about grounding, it is not ideally and positively conceivable. Here, a proposition is positively conceivable if one can imagine a scenario in which the proposition is true. Additionally, grounding is defined as a binary relation: A grounds B if and only if B exists wholly in virtue of A. Based on these notions, according to Cutter, for example, it is unimaginable that an experience of pinkness ultimately exists wholly in virtue of the motion of atoms. Here, it should be noted that a grounding relation is different from a causal relation. Perhaps, it is conceivable that the motion of atoms invokes the experience of pinkness in terms of causality. However, according to Cutter, it is inconceivable that the motion of atoms *constitutes* the experience by itself, in terms of grounding. In this sense, at the very least, if physicalism is a grounding claim, then physicalism is not ideally and positively conceivable. If these arguments are correct, it follows naturally that physicalism is inconceivable, a conclusion that proponents of the meta-modal argument find counterintuitive. Thus, the conclusion of the meta-modal argument no longer poses a problem.

3-4. Objection 4: Overreach of the Meta-Modal Argument

Finally, a strategy argues that the meta-modal argument is too strong, rendering the majority of philosophical arguments invalid. Phillips (1998) adopts this approach to criticize Marton's meta-modal argument. According to Phillips, Marton's reasoning can be applied to any putative counterexample against a theory that asserts necessary

conditions for a concept. For example, in epistemology, *the appropriate function theory of warrant* claims that a belief is warranted only if it is formed by properly functioning cognitive faculties. A common counterexample to this theory is the conceivability of a swampman.[8] However, if Marton's reasoning is correct, then the conceivability of a *swampman* would lead to consequences analogous to the conceivability of zombies. Since, according to proponents of the meta-modal argument, the conceivability of a swampman ultimately entails the inconceivability of the appropriate function theory of warrant, which is counterintuitive, the swampman argument is therefore undermined. Yet, a swampman is widely regarded as a legitimate conceptual device for critiquing the appropriate function theory. Hence, Phillips argues that Marton's reasoning is unpromising because it would invalidate many other meaningful philosophical arguments.[9]

4. Responses to the objections against the Meta-Modal Argument

In response to some of the criticisms raised in the previous section, how might proponents of the meta-modal argument reply? In this section, I will sort out some responses available to supporters of the meta-modal argument.

4-1. Response to Objection 1: Modality of Zombie Claims

First, in addressing the Objection 1 (Sec. 3-1), Marton argues as follows. According to the Objection 1, the CP thesis cannot be applied to modal claims, but if so, the same point applies, in fact, to the conceivability of the zombie world. As

8 A swampman is a physical duplicate of a human that arises purely by chance. Since the swampman lacks a history, it is often cited as a counterexample to theories that attribute the justification of beliefs to historical or causal factors, as well as to the theories that claim mental content is determined historically.

9 However, Phillips does not necessarily agree with the validity of the zombie argument. Phillips primarily criticizes Marton's methodology and maintains a skeptical stance toward the zombie argument, particularly the CP thesis.

Chalmers (1996) acknowledges, zombies are *naturally* impossible. According to Marton, that is, given the natural laws that hold in the actual world, zombies cannot exist in the actual world. Therefore, we cannot conceive of the *actual* existence of zombies. If this is the case, what we are actually conceiving is not the zombies *per se* but rather the (counterfactual) *possibility* of zombies. If what we are conceiving is the possibility of zombies, then the zombie claim is a modal claim, and the CP thesis is rendered inapplicable to their conceivability.

Here, even if dualists circumvent this response by claiming that CP thesis is applicable to the conceivability of *contingent* propositions but not to the conceivability of *necessary* propositions—and thus argue that CP thesis does not apply to the conceivability of physicalism—the same reasoning equally undermines the conceivability of zombies. Marton argues that if the possibility of zombies is conceivable, the necessary possibility of zombies must also be conceivable. This is because, again, in S5, for any proposition A, $\Diamond A \supset \Box\Diamond A$ is an axiom (Axiom 5). Therefore, to conceive of zombies is to conceive of a necessary proposition. Thus, even if the conceivability of physicalism's applicability to CP thesis is rejected on the grounds that it cannot apply to necessary propositions, the same reasoning would also undermine the conceivability of zombies.

4-2. Response to Objection 2: The Unsoundness of the Zombie Argument without S5

Turning to the second criticism, Marton (2023) offers the following response. According to Objection 2, the zombie argument does not have to rely on S5 to be valid. However, while it is true that models can be constructed within non-S5 modal logics in which the zombie argument is valid, abandoning S5 undermines the *soundness* of the zombie argument. Specifically, the second premise of the zombie argument—the entailment of the possibility of a zombie from its conceivability—cannot be in fact true. Again, let's consider the non-S5 model M used in 3-2:

- $W=\{w_@, w_1, w_2, w_3\}$
- $R=\{<w_@, w_@>, <w_@, w_1>, <w_@, w_2>, <w_@, w_3>, <w_1, w_2>, <w_1, w_3>\}$
- $V=\{<P, \{w_@, w_1, w_2, w_3\}>, <Q, \{w_@, w_1, w_2\}>\}$

Assume that all worlds are conceivable from any world. Then in this model, while the conceivability of a zombie entails its possibility at $w_@$, it does not at w_2. Generalizing from this, although CP thesis holds at the actual world, it does not hold at w_2. In such a model, there would be worlds where CP thesis holds and others where it does not, and therefore CP thesis is rendered as a contingent thesis. If CP thesis becomes contingent, then the zombie argument itself becomes contingent. That is, the validity of the zombie argument would depend on the particular world in which it is proposed—it may hold in the actual world but fail in others. However, this is an undesirable conclusion for proponents of the zombie argument. Thus, proponents of the zombie argument have no choice but to adopt S5 because once S5 is abandoned, there is no longer a guarantee that conceivability implies possibility, as there is no guarantee that all the possible worlds in the set are mutually accessible.

Furthermore, it should be noted that if the accessibility relation is not universal, even adopting S5 does not guarantee the validity of the zombie argument (Marton 1998, 2023; Piccinini 2017). For instance, it is possible to construct the following model $M'=<W', R', V'>$ that satisfies S5 while rendering the zombie argument invalid.

- $W'=\{w_@, w_1, v_1, v_2\}$
- $R'=\{<w_@, w_@>, <w_@, w_1>, <w_1, w_@>, <w_1, w_1>, <v_1, v_1>, <v_1, v_2>, <v_2, v_1>, <v_2, v_2>\}$
- $V'=\{<P,\{w_@, w_1, v_1, v_2\}>, <Q,\{w_@, w_1, v_2\}>\}$

There are two distinct equivalence classes in this model: $\{w_@, w_1\}$ and $\{v_1, v_2\}$. In this model, $v_1 \models_{M'} P \wedge \neg Q$, which identifies v_1 as a zombie world. If $w_@$ represents

the actual world, this model includes a zombie world which is inaccessible from the actual world. Furthermore, in this model, $P \supset Q$ holds in all worlds accessible from $w_{@}$, so $w_{@} \models_{M'} \Box (P \supset Q)$. Accordingly, this model establishes that physicalism is true in the actual world. Hence, even within S5, if the accessibility relation is not universal, the zombie argument could fail in some worlds. Moreover, assuming that all propositions in this model are conceivable from any world, the entailment of possibility of a zombie from its conceivability, as well as CP thesis, does not hold in the actual world. These outcomes place a burden of proof on proponents of the zombie argument, particularly dualists, who must now account for why zombie worlds should belong to the same equivalence class as the actual world, thereby rendering them accessible from it.

4-3. Response to Objection 3: Question-Begging and the Problem of Redundancy

With regard to the third criticism, the response is relatively straightforward. According to Objection 3, physicalism is not ideally conceivable. However, so long as no explicit contradiction is identified within physicalism, the inconceivability of physicalism rests entirely on the proponent's intuitions (cf. Heikinheimo & Vaaja 2013; Piccinini 2017, etc.). In other words, only dualists possess the intuition that zombie worlds are conceivable, and just as dualists claim that zombie worlds are conceivable, physicalists are equally entitled to claim that physicalism is conceivable. Thus, if the intuition supporting the conceivability of zombie worlds is to override the intuition supporting the conceivability of physicalism, dualists must provide a justification for this claim (Marton 1998).

Moreover, even if physicalism is inconceivable, this would lead to an undesirable outcome for proponents of the zombie argument (Marton 1998; Frankish 2007; Campbell et al. 2017). For instance, suppose that physicalism is inconceivable due to some intrinsic logical inconsistency. In this case, physicalism is false. However, if the falsity of physicalism can be established through its internal inconsistency,

then the zombie argument becomes logically *redundant*. This would be problematic for dualists who aim to refute physicalism by deploying the zombie argument. This is because, in this case, the zombie argument itself cannot refute physicalism completely; rather, it can do so only if it uses the inconsistency of physicalism as a premise. However, demonstrating the inconsistency of physicalism alone is sufficient to reject it. If this is the case, then the zombie argument itself is both insufficient and unnecessary for rejecting physicalism and is therefore redundant. While dualists may choose to abandon the zombie argument in light of this redundancy, doing so would be inconvenient for those who consider the zombie argument indispensable for refuting physicalism.

4-4. Response to Objection 4: Argumentum ad Verecundiam

Finally, in response to the fourth criticism, Marton (2000) offers a concise rebuttal. According to Objection 4, meta-modal argument will yield undesirable consequence for other philosophical arguments such as the swampman argument. However, Phillips, in rejecting Marton's position, relies on the fact that a swampman is widely regarded as a meaningful conceptual device. However, this amounts to an *argument from authority* (argumentum ad verecundiam). As such, Phillips's criticism does not identify any intrinsic inconsistency within the meta-modal argument but merely expresses dissatisfaction with its conclusions. This is insufficient to invalidate Marton's argument.

5. Further Objections against Responses

In this section, I argue that the Responses offered by proponents of the meta-modal argument fail.

5-1. Non-Modality of Zombie Claims

According to the Response to Objection 1, when we think of zombies, we are in fact thinking of the possibility of zombies. However, this insight is mistaken. Specifically, what we are thinking of is $P \wedge \neg Q$, not $\Diamond (P \wedge \neg Q)$.[10] This is because we can literally conceive of $P \wedge \neg Q$ being true in the actual world. For example, let Q here represent the phenomenal truth "Tom has consciousness." Then $P \wedge \neg Q$ expresses the proposition "all microphysical truths hold, and yet Tom does not have consciousness." It is conceivable that Tom *in fact* lacks consciousness, and its conceivability is supported by the existence of *the problem of other minds* (cf. Chalmers 1996, p. 74). Traditionally, the problem of other minds raises the epistemological question: How can I know that others have minds? This issue typically revolves around the argument that the physical features of others do not suffice to determine whether they possess consciousness. If this argument is correct, one can have a non-zero degree of belief that Tom lacks consciousness. Therefore, it is plausible that it is conceivable that for all physical truths to hold while some phenomenal truth does not *in the actual world*. From this observation, it follows that when we think of zombies, we are not thinking of the possibility of zombies but rather of zombies themselves. In other words, we are not thinking that $P \wedge \neg Q$ holds at some possible world other than the actual world, but rather at the actual world.

5-2. The Justification Problem in Modal Models of the Zombie Argument

According to the discussions so far, each philosopher constructs various models to counter their opponents' arguments. For example, VandenHombergh has argued that it is possible to construct a model in which the zombie argument remains valid even without assuming S5. In contrast, Marton contends that the zombie argument necessarily relies on S5 to ensure the soundness of its premises. Furthermore, according

10 Chalmers (2010, p.158) calls such conceivability of possibility of a proposition "meta-conceivability".

to Marton and Piccinini, even if a model is based on S5, the zombie argument does not hold unless all possible worlds within the model are mutually accessible. In this way, various philosophers have developed their arguments by constructing different models, whether based on S5 or not. However, they have not provided sufficient justification for *why they assume these specific models*. In other words, the question of *whether these models are appropriate for the metaphysics of consciousness* has not been addressed.

For instance, Marton and Piccinini use model M' to argue that even if S5 is adopted, there is no compelling reason to believe that zombie worlds are conceivable yet inaccessible from the actual world. However, the model M' presented in Section 4-3 can be deemed inappropriate. This is because, assuming that all worlds in the set are conceivable from any world, as we did in Section 4-3, positing the conceivability of worlds that are inaccessible from the actual world under S5 effectively entails endorsing modal dualism. However, modal dualism is implausible for the following reason (Chalmers 2010; Prelević 2017). Modal dualism entails treating the space of conceivable worlds and the space of metaphysically possible worlds as independent. This, in turn, raises an epistemological problem: How can we access or know metaphysical modalities? Since epistemic modalities, such as conceivability and logical possibility, are generally used as guides for investigating metaphysical modality, a gap between epistemic and metaphysical modalities would undermine this guiding role. Consequently, a theory that relies on such a gap is epistemologically uneconomical. To avoid these issues, the space of epistemic modalities and the space of metaphysical modalities should be considered as part of the same unified space (modal monism). Thus, there is no sufficient reason to claim that zombie worlds are conceivable yet do not belong to the set of worlds accessible from the actual world. Therefore, we can at least conclude that there is no compelling reason to resort to model M' in order to undermine the zombie argument.

5-3. Indeterminacy of the Conceivability of Physicalism

According to Response to Objection 3, the negation of conceivability of physicalism rests entirely on the proponent's intuitions, so it is not justified. However, although the proponents of the meta-modal argument assume the conceivability of physicalism, strictly speaking, whether physicalism is ideally conceivable remains indeterminate. This is because the very existence of the hard problem of consciousness demonstrates that it is uncertain whether physicalism is ideally conceivable. The hard problem of consciousness, as described by Chalmers (1995, 1996), concerns the question: Why does conscious experience arise from mere physical processes in the brain? The hard problem arises and remains unsolved precisely because we currently lack an understanding of why the physical gives rise to the phenomenal. That is, we cannot comprehend the necessity of the physical giving rise to the phenomenal. Thus, the existence of the hard problem of consciousness might serves as evidence for the indeterminacy of the ideal conceivability of physicalism. Conversely, if physicalism were ideally conceivable as true, the hard problem of consciousness would not arise in the first place. In this sense, *as long as the hard problem of consciousness exists*, whether physicalism is ideally conceivable remains indeterminate, and, as Chalmers (2010) argues, physicalism is at most prima facie conceivable.[11] Therefore, the assumption of the meta-modal argument—that physicalism is ideally conceivable—is not justified.

It should be noted, however, that while the hard problem of consciousness undermines the conceivability of physicalism, *the existence of the hard problem itself does not entail the falsity of physicalism*. This is because the existence of the hard problem does not imply an intrinsic contradiction of physicalism but merely shows that it remains indeterminate whether physicalism is ideally conceivable. Therefore,

11 In contrast to this claim, Sepetyi (2019) claims that the existence of the hard problem of consciousness supports the conceivability of zombies. However, it is unclear whether this is the case. This is because it is possible to consider that the hard problem merely supports the prima facie conceivability of zombies. To use the hard problem to support the conceivability of zombies, we need to show why the hard problem leads to the *ideal* conceivability of zombies. However, it seems quite difficult to show this without begging the question.

what follows from the existence of the hard problem is simply the indeterminacy of the ideal conceivability of physicalism, and this indeterminacy itself does not imply the impossibility or falsity of physicalism. Consequently, since the existence of the hard problem of consciousness does not entail the falsity of physicalism, then the argument that relies on the hard problem avoids the problem of redundancy discussed in the previous section while still attacking the meta-modal argument.

6. Future Prospects

This paper has reviewed the meta-modal argument, its criticisms, and the subsequent responses, and developed the author's position in response to these debates. Specifically, the following claims were advanced: (1) with respect to the conceivability of zombies, we are not conceiving of the possibility of zombies but rather of zombies themselves; (2) some model employed by physicalists to undermine the zombie argument within S5 framework is metaphysically inappropriate; and (3) the ideal conceivability of physicalism remains indeterminate due to the persistence of the hard problem of consciousness.

What prospects, then, emerge from the arguments presented in this paper? First, it would be worthwhile to investigate how widely the conceivability of physicalism is accepted. As discussed in Section 3, Chalmers and Cutter themselves do not appear to share the intuition that physicalism is conceivable, whereas Marton just accepts this intuition. Thus, it would be valuable to explore whether the conceivability of physicalism is broadly recognized, at least at the intuitive level.

Second, if certain models are inappropriate for use, we must clarify what kind of model should be employed. In other words, we should seek a model that is both plausible and suitable for metaphysical reasons. For example, if every model constructed in the metaphysics of consciousness must adopt S5 for some reason, then although the zombie argument itself may remain valid, proponents of the zombie argument would be unable to construct a non-S5 model that satisfies its conditions

while simultaneously rebutting the meta-modal argument. By pursuing such investigations, it may become possible to identify a model that serves as common ground in the debate between physicalism and dualism.

Third, it is also worthwhile to examine how widely the problem of other minds is acknowledged. This is because the argument in this paper presupposes that the problem is meaningful. However, if this problem is merely a pseudo-problem, it can no longer be used to support the claim that we can conceive of a zombie itself rather than merely its possibility.

Finally, it is also meaningful to reexamine whether the hard problem of consciousness itself is a well-defined problem. For example, some have suggested that the hard problem may be a pseudo-problem (Frankish 2016), while others have explored attempts to dissolve the hard problem by elucidating the belief-forming processes that lead us to think the hard problem exists (Chalmers 2018). If this line of research ultimately reveals that the hard problem is indeed a pseudo-problem, it would no longer be viable to rely on its existence to support the indeterminacy of the conceivability of physicalism. Consequently, such meta-level considerations regarding the hard problem of consciousness are worth pursuing.

Bibliography:

Brown, R. (2010) “Deprioritizing the a priori arguments against physicalism” in *Journal Consciousness Studies* 17, 3-4: 47-69.

Balog, K. (2012) “In Defense of Phenomenal Concept Strategy” in *Philosophy and Phenomenological Research* 84 (1): 1-23.

Brown, R. (2013) “The Two-Dimensional Argument Against Dualism” in PhilArchive. URL: https://philarchive.org/archive/BROTTA-6

Campbell, D., Copeland, J., & Deng, Z. R., (2017) “The inconceivable popularity of conceivability arguments” in *The Philosophical Quarterly* 67, 267: 223-240.

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

Chalmers, D. (2002) “Does conceivability entail possibility?” in Gendler, T. S, & Hawthrone, J. (eds.) *Conceivability and Possibility*, Oxford: 145-200.

Chalmers, D. (2010) *The Character of Consciousness*, Oxford.

Chalmers, D. (2018) “The Meta-Problem of Consciousness” in *Journal of Consciousness Studies* 25 (9-10): 6-61.

Cutter, B. (2023) “The Inconceivable Argument” in *Ergo an Open Access Journal of Philosophy* 9.

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

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

Heikinheimo, A. & Vajja, T.（2013）”Redundancy of the Zombie Argument in *The Conscious Mind*” in Journal of Consciousness Studies 20, (5-)6: 6-26.

Howell, R. (2008) “The Two-Dimensionalist Reductio” in *Pacific Philosophical Quarterly* 89: 348-358.

Jackson, F. (1998) *From Metaphysics to Ethics*: A Defense of Conceptual Analysis, Oxford.

Marton, P. (1998) “Zombies versus materialists: The battle for conceivability” in *Southwest Philosophy Review* 14. 1 (1998): 131-138.

Marton, P. (2000) “The murderer returns: A reply on zombies to Jamie Phillips” in *Southwest Philosophy Review* 16. (2): 195-200.

Marton, P. (2023) “Conceivability, Kripkean Identity, and S5: A Reply to Jonathon VandenHombergh” in *Erkenntnis* : 1-10.

Mizrahi, M. & Morrow, D. (2015) “Does Conceivability Entail Metaphysical Possibility?” in *Ratio* 28 (1): 1-13

Piccinini, G. (2017) “Access Denied to Zombies” in *Topoi* 36. (1): 81-93.

Prelević, D. (2017) “Access Granted to Zombies” in *Theoria: Beograd* 60 (1): 58-68.

Sepetyi, D. (2019) “Why phenomenal zombies are conceivable whereas anti-zombies are not” in *Actual Problems of Mind* 20: 18-27.

Sturgeon, S. (2000) *Matters of Mind: Consciousness, reason and nature*, Routledge.

VandenHombergh, J. (2017) “Inconceivable physicalism” in *Analysis* 77. (1), 116-125.

VandenHombergh, J. (2020) “Consciousness, conceivability, and intrinsic reduction” in *Erkenntnis* 85: 1129-1151.

Yablo, S. (1999) “Concepts and Consciousness” in *Philosophy and Phenomenological Research* 59 (2): 455-463.