Abstract:
Since 1996, when David Chalmers introduced the zombie argument against physicalism in *The Conscious Mind*, numerous works of ever-increasing technical complexity and nuanced argumentation have been written on the conceivability and possibility of zombies. In this paper, I focus on the main points of the argument. First, I discuss the conceivability of zombies. I briefly outline three other thought-experiments in order to determine what is expected of a good thought-experiment and its workings. Next, I turn to Chalmers' defense of their conceivability, where the key consideration is to present conceivability as a credible a priori method that can entail metaphysical possibility. I conclude that Chalmers does not manage to create a credible link between conceivability and possibility, thus failing to show that zombies are not only conceptually possible, but also metaphysically possible. The most problematic idea is the identification of logically possible worlds with metaphysically possible worlds. Chalmers' main aim is to defend conceivability as an a priori method of acquiring modal knowledge, but by limiting it to the rational domain, the acquired knowledge is not knowledge of the objective reality, but of the content of our thoughts.

Key words: modal knowledge, conceivability, a priori, materialism, Chalmers.

The zombie argument was introduced by David Chalmers in his *The Conscious Mind* (1996) as his main argument against physicalism. If physicalism (or materialism) is true, then all positive facts about the world are supervenient on physical facts, including facts about consciousness. However, if we can conceive a world, a physical duplicate of ours, in which there are no conscious experiences at all, and if we can, furthermore, show that such a world is not just a figment of our imagination, but a “real” possibility—that is, a “true” alternative to our world—then consciousness does not supervene on the physical, and physicalism is therefore false. Such a world would be a zombie world, inhabited by zombies who are our physical duplicates and function and behave exactly like us, but who are entirely devoid of conscious experience.

The main question here is whether zombies are conceivable as well as possible in such a way that their possibility refutes physicalism, a metaphysical thesis concerning the nature of our world. Chalmers answers affirmatively to both questions, and in his detailed and comprehensive argumentation focuses especially on the once-taken-for-granted step from conceivability to possibility which came under fire after Kripke's persuasive defense of a posteriori necessities (Kripke 1980). To disarm the danger that necessities which are knowable only through sense experience pose to conceivability as an a priori method of acquiring modal knowledge, Chalmers developed epistemic two-dimensional semantics, which he further improved over the years (Chalmers 2004; Chalmers 2010).

Numerous works have been written as a response to the zombie argument, works which are often very detailed and technical. Critics usually attack the thesis that conceivability entails possibility as Chalmers predicted, but there are some who question whether or not zombies are really conceivable. Chalmers has more clearly formulated the argument and provided answers to the main objections in *The Character of Consciousness* (2010: Ch. 6). The topic is obviously very popular, and it is literally impossible to read everything that has been written about it. It is also quite an expansive since it concerns the nature of the world, the nature of consciousness and its connection with the physical. The metaphysical aspect is closely

---

1 Here I will follow Chalmers in understanding the terms “physicalism” and “materialism” as being interchangeable.
interwoven with the epistemological, conceptual, and methodological aspects (e.g., the nature of explanation). On top of this, the formulation of the argument within the two-dimensional semantic framework makes it very technical. In this paper I will try not to get entangled in technicalities and details—I will only present the main points of the argument. First, I discuss the conceivability of zombies. I consider Chalmers' thought-experiment in which he invites us to imagine our zombie twins and compare it with three other well-known thought-experiments. Next, I turn to Chalmers' defense of the conceivability of zombies, which relies on demonstrating that conceivability can be a respectable method of acquiring knowledge. The key issue here is that the method must be natural, but it must also be a guide to the right kind of possibility—metaphysical possibility. In the end, I assess how successful Chalmers is in this, and whether he manages to show that the conceivability of zombies entails their metaphysical possibility.

1. The Conceivability of Zombies

A zombie world is a physical duplicate of our world in which there are no conscious experiences. While physicalism claims that everything globally supervenes on the physical, i.e., that all the microphysical facts in the world (including fundamental microphysical laws) entail all other facts, it is reasonable to assume that consciousness is supposed to supervene locally. Thus, we can characterize the zombie world as a world which is inhabited by zombies, and in order to determine its conceivability we need to determine whether we can conceive a zombie: “someone or something physically identical to me (or to any other conscious being), but lacking conscious experiences altogether” (Chalmers 1996: 94).

To make the notion more vivid and compelling, Chalmers proposes a thought experiment: whatever one is doing at the moment, they should try to imagine their zombie twin doing exactly the same thing but lacking conscious experience. For example, my twin zombie is at this point typing, listening to music, but unlike me is not feeling any aching sensation in her back. Everything will be the same, says Chalmers, imagining his zombie twin, only “[t]here will be no phenomenal feel. There is nothing it is like to be a zombie” (ibid: 95).

Let me try to imagine my zombie twin, then. By definition my zombie twin is my physical duplicate, embedded in an identical environment, functionally and psychologically identical to me. As it is a habit when conducting a philosophical thought experiment, I start with a pretty general idea—I imagine an environment identical to the one in which I am situated and put in my place a creature that is identical to me molecule for molecule, looks exactly like me and is doing exactly the same things as me, but lacks conscious experiences. I have no problem imagining such a situation, so I agree with Chalmers that the zombie hypothesis is prima facie coherent and imaginable (Chalmers 2010: 154), but does it persuade me that physicalism is false? No, it turns out that I am not quite sure how the fact that I can imagine a zombie relates to the falsity of physicalism. How does something I can imagine refer to what is the case in the world? Some further information is needed for the thought-experiment to work, but before I turn to Chalmers’ exposition, here is a short reminder of how thought-experiments are being used in philosophy.

2. Comparison with Other Thought-Experiments

Testing theories against imaginary examples is an old and respected philosophical method of research. Philosophers develop them to support theories, but also to refute them. Probably the most famous thought-experiment in support of a theory is found in Meditations (1641/1996), in which Descartes, sitting by the fire, asks himself what he can be sure of, and by abstracting away things that do not pass the test, ends with the only thing he cannot doubt—that he is a thinking thing (res cogitans). Since this is distinct of his body (res extensa), he proceeds by developing an argument for substance dualism. In contemporary analytic philosophy, Gettier's imaginary counterexamples to the analysis of knowledge as justified true belief undoubtedly present the most successful refutation of a theory by use of thought-experiments (Gettier 1963). Also very influential was Kripke's use of modal arguments in his criticism concerning the descriptivist theories of proper names and natural kind terms (Kripke 1980).

Gettier and Kripke proceed in such a way that

---

2 Chalmers adds that to disprove physicalism we actually only need conceivability of a world with one zombie amidst other normal conscious beings, or even a world with one invert, an individual who has a slightly different experience from the corresponding individual in our world (Chalmers 1996: 124 and 2010: 142). However, these details are irrelevant for the purposes of this paper.
they first present the theory they want to criticize, then they develop examples that the theory wrongly classifies. Gettier describes a situation in which a subject infers a proposition from another proposition. While the subject is justified in believing that the latter proposition is true, it is in fact false. However, the inferred proposition which the subject believes on this basis happens to be true, and the subject possesses justified true belief. According to the traditional analysis of knowledge, this means that the subject possesses knowledge as well, but Gettier's intuition, shared by most of philosophers today, is that it should not count as knowledge. From this it follows that justified true belief is not sufficient for—let alone equivalent to—knowledge. Similarly, Kripke, criticizing the descriptivist theory of proper names, invites us to imagine a situation in which the referent of a proper name does not satisfy the definite description by which he was picked out as a referent in our world. Would we in such a case think that he is the referent of the name, or would we pick out another individual who happens to satisfy the definite description in the counterfactual situation as a referent? Intuitively, argues Kripke, we would choose the first option, which shows that we referent? Intuitively, argues Kripke, we would choose the first option, which shows that we understand proper names as rigid designators that pick the same objects in all situations. Since the descriptivist account suggests the second, counterintuitive option, it must be false. Admittedly, Kripke was not as successful as Gettier, but he persuaded quite a lot of philosophers.

What do these two examples have that the zombie example lacks? All are trying to show that some theory is wrong, but the theories in question are very different. Gettier is criticizing the analysis of knowledge that is expressed in one very short sentence, and it is thus very easy to comprehend. Kripke is assessing a semantic theory, namely a theory of naming, which he manages to expound on one page. It is true that some object that he misrepresents it, but this is a topic for some other time. What matters for this discussion is that the criticized theory is clearly stated, and we know precisely what the target is. In the case of zombies, criticism is directed towards physicalism, a metaphysical theory about the nature of our world. The main idea can be put very succinctly: all positive facts in our world supervene on physical facts. But when we start to fill in the details, matters quickly become complicated. What does the physical basis on which everything supervenes consist of, and how does it exactly relate to all the rest? What kind of relation is supervenience? Why must there be a conceptual link between the two classes? Maybe, because the theory is so complex, it is more difficult to decide whether or not the presented counterexample refutes it. But an even more important difference between the first two cases and the case against physicalism is their subject matter. Physicalism is a theory about a reality independent of mind and language, while in the case of knowledge and naming we are concerned with our dealings with this reality, so they are in part also dependent on us. As a result, we are more competent in judging whether the proposed thought-experiments reveal difficulties for the particular theory. What would I count as knowledge? How would I use a proper name? In the case of physicalism we are not so sure—is the zombie twin I am imagining really a physical duplicate of me? Of course I can imagine something along these lines, but is it just a figment of my imagination, or does the fact that I can imagine it really say something about the nature of our world? What is the nature of our world? Is everything in our world physical or is there also another ingredient—the phenomenal? Can the answer really come from the zombie argument?

Related to this is also the issue of how far-fetched the situations we are asked to imagine are. Chalmers admits that the idea of zombies is a strange one, and that it is unlikely that zombies are naturally possible, so the situation we are supposed to imagine is quite removed from our actual world (Chalmers 1996: 96). Accordingly, we are less sure about our imaginative abilities. Kripke's examples ask us to imagine counterfactual situations, some quite ordinary, like a situation in which Aristotle did not become a philosopher, or a situation in which Gödel did not prove the incompleteness of arithmetic. However, later in the book, when he discusses essentialism or develops the account of natural kind terms, counterfactual situations become less familiar. We are asked, for example, to judge whether Queen Elizabeth could have been born of different parents than the ones from whom she actually came, or whether water could have been composed of something other than H2O. Judging these examples is more difficult because it is not enough to rely on our naming practices. We also need to consult our beliefs on the nature of things, what we consider as a legitimate change in a thing and what would bring about its destruction. Gettier's examples are on the opposite end of this scale of
closeness to the actuality. We are not asked to imagine anything that could not have happened in our world, and the situations are very easy to imagine as a result.

The faculty of imagination is often accused of being too closely connected with our perceptual faculties, which are directed toward detecting happenings in the world, mostly limited to the level of middle-sized objects. In this way it is restricted in its abilities. Already Descartes has pointed out that it is unable to visualize extremely complex geometrical shapes (like chiliagons), while in assessing metaphysical issues it is completely useless. Descartes' solution is to distinguish between conceiving and imagining. Intellect, not the imagination, grasps the natures of things that according to him are the primary subject matter of metaphysics. These are what the thoughts of possibility and necessity are about, and something is judged to be possible if the mind clearly and distinctly conveys it.\(^3\)

Thought-experiments, however, do not rely only on imagination. In constructing scenarios, words and concepts are used in addition to (sensory) images. In fact, a scenario often just means a verbal description of a situation. So, in conducting a thought-experiment we use imagination in a rationally constrained way. In fact, philosophers of a more rationalist bent treat imagination more like an occasionally helpful illustration, while the real work is done by conceiving a scenario and determining whether it is conceptually coherent. In this manner, Chalmers first invites us to imagine a zombie to “fix ideas,” but then points out that it is crucial to determine whether the notion of a zombie is conceptually coherent or not (Chalmers 1996: 94–96).\(^4\) So my ability to conjure up an image of my zombie twin serves as an illustration, while the majority of work is done through the analysis of concepts which are used in presenting the situation. It is conceivability which is the true guide to metaphysical possibility. So, the true question that I should be asking myself is not whether I can imagine my zombie twin, but rather whether I can conceive her. Is my zombie twin conceivable?

Chalmers suggests that, in order to determine this, I need to carefully describe the situation and then check that “there is no hidden contradiction lurking in the description.” The procedure is no different than when we consider a mile-high unicycle. Undoubtedly, no such thing as a mile-high unicycle exists in a real world, but its description is coherent. Similarly, the analysis of a zombie scenario does not reveal any contradiction, thus showing that the described situation is coherent and the idea of a zombie conceivable (Chalmers 1996: 96).

I will admit that the zombie scenario seems conceivable, but it is not entirely clear to me how I reached this conclusion. The description is so short—a creature physically, functionally, and psychologically identical to me, embedded in an identical environment, minus conscious experiences—that there is not much chance of discovering any contradiction. Does this quick procedure suffice for me to confidently claim that my zombie twin is conceivable? And surely even if my zombie twin is indeed conceivable, it does not automatically mean that it is metaphysically possible? Does it?

3. The Characterization of Conceivability

Chalmers tackles this issue by further explaining the notion of conceivability. It is an epistemic notion, defined in epistemological and maybe also psychological terms. There are three dimensions of difference between the notions of conceivability: prima facie vs. ideal conceivability, negative vs. positive conceivability, and primary vs. secondary conceivability.\(^5\) A sentence \(S\) is prima facie conceivable for a subject when \(S\) is conceivable for that subject on first appearances: after some consideration the subject finds that \(S\) passes the tests which are criterial for conceivability. This is the subject's initial reaction to the scenario, but this kind of conceivability is too weak, since it is tied to contingent cognitive limitations. A worry is that the person did not conceive the scenario in sufficient detail to discover a well-hidden inconsistency—as in the case when I worried that I came too quickly to the conclusion that my zombie twin is conceivable—and that maybe further reasoning would disprove this. What is needed here is an ideal conceivability for which one needs justification that cannot be rationally defeated, i.e.\(^5\)

---

\(^3\) For more on the distinction between imagining and conceiving, as well as on Descartes' position, see "Introduction" in Gendler & Hawthorne (2002).

\(^4\) In Chalmers 2010, he speaks of both conceivability and imaginability, but it is clear from the exposition that the outcome is achieved by rational reasoning.

\(^5\) The three distinctions and consequent varieties of conceivability are stated in Chalmers (2010: 143–148), but for a detailed discussion, see Chalmers (2002).
ideal reasoning is required that cannot be defeated by any further reasoning.

However, the distinction between prima facie and ideal conceivability does not help me allay my doubts. My worry was that my reasoning was not good enough and that it could turn out that I did not conceive a real possibility after all. How can I know that my reasoning cannot be defeated by better reasoning? Have I done enough, or did I settle for prima facie conceivability of my zombie twin? According to Chalmers, although we are nonideal beings we still know many things that are ideally inconceivable (e.g., that $0 = 1$), and that are ideally conceivable (e.g., that someone exists). We also know that certain things about the world are knowable a priori, but others cannot be known a priori even by an ideal reasoner. Similarly, in discussing consciousness we have good reason to know that certain things about the world are conceivable (e.g., that I have conscious experiences and my zombie twin does not). Second, I need to fill in as many arbitrary details as possible and try to see if the situation remains coherent. That is, I need to provide a more detailed description of my zombie twin's physical state to see if it is really identical with my physical state. This surely cannot just be assumed to be the case, since according to physicalism the description of my physical state should also describe my conscious experience. Accordingly, it is possible that my zombie twin must be missing some physical properties in order to be without consciousness, or maybe the whole idea is simply incoherent. Moreover, given that today science is not yet able to explain consciousness, maybe there will be further scientific discoveries and we will acquire new physical concepts, or the old ones will be improved and it will turn out that zombies are not conceivable.

Chalmers responds to some objections to the conceivability of zombies in his *The Character of Consciousness* (2010: 154–157). It seems that in assessing the zombie hypothesis he does not rely so much on the comprehensive description of microphysical facts and on proving that there is no contradiction lurking around, but on his intuition that the physical and the phenomenal concepts are in principle so different that there is no way that the phenomenal concepts could apply to the same entities as the physical concepts.

Chalmers' rejection of physicalism and his zombie argument are informed by two ideas. One is the idea that modal knowledge is acquired through a priori means (modal rationalism), specifically by analyzing concepts (conceptual analysis). The second is that the metaphysical supervenience relation enables the reductive explanation which, to a great extent, proceeds by a priori means.

Now, if physicalism is understood as a metaphysical supervenience thesis, then to be true it must fulfill the following requirement: “for any logically possible world $W$ that is physically indiscernible from our world, all the positive facts

---

6 For example, Patricia Churchland claims that we are at this point unable to explain consciousness because neuroscience is still underdeveloped, and not because it is in principle unexplainable (Churchland 1997).
true of our world are true of \( W' \) (Chalmers 1996: 42). This also means that everything is reductively explainable in terms of microscopic physics, and that there is an entailment from the basic physical facts (fundamental entities and properties of completed physics, as well as the fundamental laws of physics) to all other positive facts, including phenomenal facts. This means that, in principle, if we possess phenomenal concepts, we should be able to read off phenomenal facts from the specification of microphysical facts (including the microphysical laws). In order to see how the application conditions of a phenomenal concept can be satisfied by fixing physical facts, we only need to be able to analyze what it takes for an entity to satisfy the intension of a phenomenal concept. But this we should be able to achieve, according to Chalmers, mostly a priori if we know the intensions of phenomenal as well as physical concepts, since intension is “a function specifying how the concept applies to different situations” (Chalmers 1996: 54). If we know both, we can determine their connection even without empirical information about our world. However, Chalmers claims that between the phenomenal and the physical there is no such conceptual link, and that there is an explanatory gap between the physical level and conscious experience; namely, physical properties are best explained as structural or dynamical properties, but phenomenal properties are not suited for such explanation, and therefore physics in principle cannot explain consciousness. Physicalism must be false.\(^7\)

This conviction of Chalmers is also behind his denial that it may turn out after some new scientific discoveries that zombies are not conceivable after all. He points out that it is highly unlikely that the structural-dynamic character of physics and physical concepts would be changed by new scientific developments, so the phenomenal concepts should already today have a character that would link them appropriately with physical concepts, but this is not the case. The only vaguely possible candidate would be functional concepts. This brings me to my second objection that one cannot just assume that my zombie twin's physical state is identical to mine, but needs to be argued for. The view that explains consciousness in wholly functional terms is called analytic functionalism.\(^8\) According to this view, to say that someone is conscious only means that they have a state that functions in an appropriate way in their cognitive system and in their behavior (Chalmers 2010: 155). This means that by definition, any functional duplicate of mine is also conscious. On this view zombies are inconceivable, and more generally, there is no epistemic gap between physical and phenomenal truths. It also means that explaining consciousness simply means explaining the various functions. But, as Chalmers argues, there is a strong intuition that there is something further to explain, an intuition which is shared by the majority, so functional analysts owe us very strong arguments to oppose it. They do not provide them, so their view should be rejected (ibid: 111–114).

Analytic functionalism is, in my opinion, suffering from the same disease as Chalmers' project: it relies too heavily on conceptual analysis. How do we get from knowledge of our concepts to knowledge of mind-independent reality, from conceptual to metaphysical? Our concepts are often inadequate to the true nature of things; just take the example of water. Moreover, we sometimes have concepts of things that do not exist after all, for example phlogiston. How can conceptual analysis show that physicalism, the substantive thesis about the nature of our world, is false? How can conceptual analysis on its own tell us what is metaphysically possible and what is metaphysically necessary? Surely this depends on the very nature of things in the world and not only on our conceptual scheme.\(^9\) To what degree do our concepts mirror the objective reality? Moreover, there is a persistent feeling that our cognitive limitations crucially determine what we are going to count as possible. By distinguishing between prima facie and ideal conceivability, Chalmers tries to reduce the dependence of the method on the ability of a thinker, but the distinction seems to me to be too abstract to produce correct results. Concerning the issue of the application of concepts, Chalmers denies that they are provided by a definition (at least not in a majority of cases), but rather by intension, which gets determined from the detailed consideration of scenarios. We presuppose that a certain scenario obtains and then ask ourselves what we would say in that case. By considering a certain number of scenarios, we notice certain regularities and are thus able to determine the extensions of concepts. But what kind of scenarios—a hypothetical situation that

---

\(^7\) For the exposition on why physics in principle cannot explain phenomenal consciousness, see especially Chalmers 1995.

\(^8\) See, for example, Lewis (1966) and Armstrong (1968).

\(^9\) For a similar position, see Lowe (2013).
might turn out to be true (a conceptual or epistemic possibility), or that might have been true (metaphysical possibility)?

4. The Link from Conceivability to Possibility

The most common objection in the literature is the objection to the conceivability-possibility entailment, and it draws especially from the existence of a posteriori necessities that were persuasively argued for by Kripke (1980). The identity statement “Water is H₂O” is the most often cited example of a posteriori necessity, namely metaphysically necessary truth that is knowable only a posteriori. A posteriori necessities present a threat for modal rationalism, for the a priori route to modal knowledge. Take the example of water. If we know the identity of water with H₂O only a posteriori, i.e., only after the scientists discovered its chemical structure, then first of all, we cannot come to know this necessity only by a priori means, and moreover, we can coherently conceive many situations concerning water which do not represent metaphysical possibilities. For example, we can conceive a situation in which water has some different chemical structure, say XYZ, which makes it conceivable that water is not H₂O, yet it is not metaphysically possible. From this it follows that conceivability is not a very good guide to possibility. So why would we think that from the conceivability of zombies follows their metaphysical possibility?

Chalmers tackles the problem of a posteriori necessities by distinguishing two kinds of conceivability—the primary one and the secondary one, which are linked to two different kinds of possibility. The distinction is based on the fact that within two-dimensional semantics there are two ways in which we can evaluate a sentence across possible worlds. In nontechnical terms, the distinction in the case of positive conceivability could be expressed as follows: one can imagine a situation as actual—as a way the world might actually be—and on this basis consider if it follows that S is the case. If it does, then S is primarily conceivable. The extensions of the involved concepts are fixed through their primary intensions. On the other hand, one can imagine a situation as counterfactual—as a way the world might have been. If the imagined situation is coherent and S is the case, then S is secondarily conceivable; concepts are associated by secondary intensions (Chalmers 2002: 156–158).

Primary conceivability is a reliable guide to primary possibility, and secondary conceivability is a reliable guide to secondary possibility. Primary possibility is a familiar conceptual possibility, but Chalmers also calls it epistemic possibility since we consider a particular situation solely on the basis of our a priori knowledge, suspending our empirical knowledge as it were. Accordingly, primary conceivability is always an a priori matter (what might conceivably be the case) and is the relevant type for modal rationalism. Secondary conceivability is linked to secondary possibility; Chalmers also calls it subjunctive possibility (what conceivably might have been the case). This is a familiar metaphysical possibility. When we evaluate a particular situation from the perspective of secondary possibility, we acknowledge that the character of the actual world is fixed, and ask: if the situation had obtained, would S have been the case? Since the application of words in situations considered counterfactually often depends on their reference in the actual world, secondary possibility is often a posteriori.

If we apply this distinction to the case of the sentence “Water is not H₂O,” we get the following result: a priori determined primary intension of a concept fixes reference in the actual world. The primary intension associated with water is something like “watery stuff” or “the clear, drinkable liquid in our environment.” The secondary intension is determined a posteriori and picks out reference in counterfactual worlds. The secondary intension associated with water is “H₂O.” Chalmers does not say specifically, but H₂O seems to have the same primary and secondary intension: “H₂O.” Now, if we evaluate the sentence according to the primary intensions involved, the primary intensions of water and H₂O differ, so it is primarily conceivable and primarily possible that water is not H₂O. However, when evaluated according to the secondary intensions, the sentence is not secondarily conceivable or possible since the secondary intensions of water and H₂O are the same.

For modal rationalism secondary possibility is clearly not acceptable, since in many cases the evaluation is possible only after a priori knowledge is supplemented with empirical information about the actual world. The central role is therefore given to primary conceivability, which is entirely determined a priori, and as a result constitutes a reliable guide to the primary possibility. The primary intensions of involved concepts are used in this way.
in the zombie argument also, and the demonstrated possibility of a zombie is therefore the primary possibility. But aren't secondary intensions required here, since they are the ones that ultimately lead to metaphysical modality? As we have seen in the case of water, if evaluated according to the primary intensions the result is not metaphysical, but rather an epistemic possibility.

Chalmers argues that although primary conceivability does not entail metaphysical possibility, there is still a link between the two, since when evaluating sentences—whether associated with primary or secondary intensions—we are evaluating them across the same space of possible worlds (Chalmers 2010: 146). As he says, the distinction between primary/epistemic and secondary/metaphysical possibility “is not a distinction at the level of worlds, but at most the distinction at the level of statements” (Chalmers 1996: 68). Thus, the world that is primarily conceivable and therefore epistemically possible (a way that a world might actually be), is also metaphysically possible (a way the actual world is not but could have been). For example, when we imagine a Twin Earth situation in which watery stuff in the oceans and lakes is XYZ, we have access to a metaphysically possible world. It is the same case when we conceive a zombie world using the primary intensions—it is a metaphysically possible world. Furthermore, “the primary intension determines a perfectly good property of objects in possible worlds.” For example, there is nothing wrong with the property of being watery stuff, as well as with the property that is picked out by the primary intension of a phenomenal concept. Consequently, argues Chalmers, “[i]f we can show that there are possible worlds that are physically identical to ours but in which the property introduced by the primary intension is lacking,” then physicalism is still proven to be false (Chalmers 1996: 132).

Chalmers explains his solution in more detail, especially in The Character of Consciousness (2010: “The Two-Dimensional Argument against Materialism”), but this basic outline should suffice. Despite careful exposition of the argument within the framework of two-dimensional semantics, my main question is still unanswered: how can conceiving of something be sufficient to establish the possibility of that something? As we have seen, Descartes believed that when conceiving, the intellect is revealing the essences of things which are external to our minds. If we can clearly and distinctly conceive something, that thing is metaphysically possible. While he too rashly judged the intellect to be infallible, I believe he was on the right track when he identified the essences of things as the source of metaphysical necessity and possibility. Conceivability can be a guide to metaphysical possibility only if it manages to become something more than just knowledge of our concepts. Chalmers’ suggestion that we determine the extension of concepts by considering their use in various hypothetical situations could be a way to latch on to something more objective, and in this way to shake off the curse of seemingly superficial and hopelessly mind-dependent concepts. However, his epistemic two-dimensional semantics—in which epistemic possibilities correspond to metaphysically possible worlds—is actually an attempt to subsume metaphysical modality under the conceptual.

The idea that epistemic and metaphysical modalities are being evaluated across the same space of logically possible worlds is counterintuitive. These are two distinct kinds of modality that derive from different sources. Epistemic modality—which is usually referred to as conceptual, but is also logical modality in a broader sense—is concerned with what is possible or necessary given our concepts and the connections among them. Chalmers focuses more on the apriority of such truths, the fact that we can know them without consulting empirical information, and accordingly calls them epistemic. An epistemic possibility is, as Chalmers puts it, “a specific way the actual world might turn out to be, for all one can know a priori” (Chalmers 2010: 216). Metaphysical modality, on the other hand, is not modality concerning our concepts, but the mind-independent reality. It is about what could or could not have been the case, given how our world in fact is. Modality concerns the world itself, which we take in its totality. We do not limit ourselves to the content of our concepts and wonder what kind of worlds could go with them, if we varied empirical data. We start with all possible knowledge we have.

---

10 In contemporary metaphysics, the view that modality is grounded in essence and that metaphysical modality should be characterized in terms of essences or identity of things and not the other way around is persuasively defended by Kit Fine (1994). In modal epistemology, Hale (2013) and Lowe (2012) develop accounts of modal knowledge as being derived from essentialist knowledge.

11 Vaidya similarly claims that the identification of metaphysical modality with logical modality leads to the counterintuitive view that the sources of modality are the same (Vaidya 2008: 198).
about our actual world and wonder in what ways it could have been different and in what ways it could not have been. If in the first case we ask ourselves, for example, what is our concept of water, then in the second case the question is what is the essence or nature of water, that which makes it the thing it is. Clearly, to conceive a world on the basis of our conceptual knowledge cannot mean that we determined one way our world could have been. Intuitively, we see a clear distinction between the two possibilities. Chalmers says that what connects both possibilities is the fact that they are both “ultimately grounded in what it is rationally coherent to suppose” (Chalmers 2010: 187). Sure, we certainly employ our reasoning when we are determining both kinds, but as I said, they are grounded in two distinct domains, and the metaphysical kind cannot be limited to the rational domain, as Chalmers suggests, since it concerns the nature of the independent reality.

Chalmers goes on to argue that all modal phenomena are best analyzed within one modal space, and that there is no reason to postulate two separate spaces of worlds. He considers also the possibility that one could try to introduce metaphysically possible worlds as a subset of logically possible worlds, and proposes a few counterexamples to the strong necessities which would go with such a construction of a modal space (Chalmers 2010: 166–184). There is another option he does not consider: a unified modal space of metaphysically possible worlds and conceptual modality as a restricted kind of metaphysical modality. Such a view is proposed by Kit Fine, who identifies the source of metaphysical necessity in the identity of all objects and sees conceptual necessity as a restricted kind of metaphysical necessity, since the source of its truth is the identity of concepts, a more restricted class of objects (Fine 1994: 9).

As we have seen, the thought-experiments against a certain theory are successful if they persuade us that the theory wrongly explains a certain example by appealing to our intuitions. The main problem with the zombie argument is that the idea of a zombie is counterintuitive. A zombie world is a complete physical duplicate of our world, but instead of conscious beings it is inhabited by zombies. A zombie is a complete physical duplicate of a conscious being, but lacks conscious experience; it nevertheless functions and behaves exactly like the latter. Now, if a zombie and a zombie world are indeed metaphysically possible, then this is a problem for physicalism as well as dualism. Physicalism is false and one needs to adopt some kind of dualism. However, while Chalmers may be proven right and a nonphysical explanation of consciousness is needed, there is a new awkward thing to be explained: why do we have consciousness at all? If a zombie functions and behaves exactly like us, and if it does not reveal any deficiency, then what is the point of having consciousness at all? It seems to be something we can easily do without, which goes against what we firmly believe.

References