SEÇÃO:EPISTEMOLOGIA E FILOSOFIA DA LINGUAGEM

Elucidating the Conceivability Argument

Elucidando o argumento da conceptibilidade

Aclarando el argumento de la concebilidad

Julia Telles de Menezes[1]
orcid.org/0000-0003-1302-6077
juliatelles@gmail.com

Abstract: It shall be examined how anti-physicalist arguments give rise to the tension between those aspects of our everyday life (with focus on phenomenality) and the thesis of physicalism. The debate over the subjective character of consciousness, or as it is sometimes called: "the hard problem of consciousness" (CHALMERS, 1996), is considered to be the greatest challenge to physicalism. Many philosophers posit this as a matter that cannot be solved, regardless of scientific progress, for it is beyond the scope of what science can find out about the world. If they are correct, the consequence is that the idea of physicalism itself fails. The paper is divided in two parts. For the first part we will deal with Chalmers' version of the conceivability argument as well as the semantic apparatus of the two-dimensional framework required to make the appropriate link between conceivability and possibility. At the end of this we shall take a look at Kripke's version of the conceivability argument against physicalism.

Keywords: Conceivability. Phenomenal consciousness. Two-dimensional semantics.

Resumo: O presente trabalho examina como argumentos antifisicalistas representam a tensão entre certos aspectos da nossa vida cotidiana (fenomenalidade) e a tese metafísica do Fisicalismo. O debate acerca do caráter subjetivo da consciência, ou do também chamado "Problema Difícil da Consciência" (CHALMERS, 1996), é considerado um dos maiores desafios à tese fisicalista. Muitos filósofos pensam que o problema é "difícil" pois a questão não pode ser resolvida empiricamente, ela é inteiramente conceitual. O presente artigo é dividido em duas partes. Na primeira parte analiso a versão de Chalmers do argumento daquilo que pode ser concebível, bem como o aparato semântico bidimensional necessário para conectar os domínios conceituais e modais. Na segunda parte tratarei da versão kripkiana dos argumentos.

Palavras-chave: Conceptibilidade. Consciência fenomenal. Semântica bidimensional.

Resumen: El presente trabajo examina cómo los argumentos antifisicalistas representan la tensión entre ciertos aspectos de nuestra vida cotidiana (fenomenalidad) y la tesis metafísica del fisicalismo. El debate sobre el carácter subjetivo de la consciencia, o el llamado "Problema Difícil de la Consciencia" (CHALMERS, 1996), es considerado uno de los mayores desafíos a la tesis fisicalista. Muchos filósofos piensan que el problema es "difícil" porque está más allá del alcance de lo que la ciencia puede descubrir sobre el mundo. El trabajo está dividido en dos partes. En la primera parte analizo la versión de Chalmers del argumento de la concebilidad, así como el aparato semántico bidimensional necesario para establecer el vínculo apropiado entre la concebilidad y la posibilidad. En la segunda parte, el trabajo se ocupa de la versión de Kripke del argumento de la concebilidad contra el fisicalismo.

Palabras clave: Concebilidad. Consciencia. Semántica bidimensional.

The state of the debate concerning the relationship between the physical and the mental has made at best a modest progress in the last centuries. Of course, since the cartesian formulation of the so-called "mind-body

---

1   Universidade Federal Fluminense (UFF), Niterói, RJ, Brasil.

problem" to the most recent contributions in the Philosophy of Mind, it is undeniable that the debate has significantly benefitted from the developing of formal and conceptual tools within the analytic tradition. Modal logic deepens and clarifies the concepts of metaphysical and epistemic possibilities, which are central to the original cartesian argument, allowing for the *conceivability argument* to become more refined than the original version. Semantic and logical tools make the relationship between epistemic and ontological domains explicit, hence, one no longer needs to resort to the notion of God to bridge the gap between imagination and possibility; Moreover, reflection on cognitive psychology also refines the concepts deployed in attending to our own experience, which is the case of phenomenal concepts[2]. The debate's state of the art can still be seen as a dilemma though, the same dilemma present in Descartes' time. If our starting point is Physicalism then the so-called *Causal Argument* supports its truth. If we, nevertheless, accept a core premise of conceivability arguments against Physicalism, then we are compelled to turn to the next available alternative, which is Dualism. Now, assessing Dualism would confront us with the causal argument, and we would come back to our starting point.

The point of this paper is to unpack one of the horns of the dilemma, and not to solve it. My aim is to show the workings of different versions of the conceivability argument and contemporary conceptual tools we use to make sense of the argument. In other words, the goal of the paper is to elucidate a couple of versions of the conceivability arguments, particularly the two-dimensional argument against materialism, and comparing it with the modal argument advanced by Saul Kripke. Section one will briefly show the cartesian version of the argument, section two will be dedicated to examining the most technically complicated version of the conceivability argument (as formulated by Chalmers (1996, 2010). Finally, section three presents the kripkean version of the argument.

1 Locus Classicus

The *locus classicus* for the conceivability argument is Descartes' Meditation VI. Descartes aims at deriving the distinctness of mind and body from an epistemic assumption; one can form a clear and distinct idea of one's mind existing without one's body. The fact that mind and body may be conceivable as distinct entities is a tool to infer that they could indeed be distinct, and the possibility of them being distinct is then used to conclude that they are indeed distinct. Therefore, physicalism is false.

> (1) I can conceive clearly and distinctively that I, a thinking thing, can exist without my extended (i.e., physical) body existing.
>
> (2) Anything that I can conceive of clearly and distinctly is logically possible.
>
> (3) If it is logically possible that X (mind) exists without Y (body), then X (mind) is not identical to Y (body).
>
> (4) Therefore, I, a thinking thing, am not identical with my extended body.

The problem with this seminal formulation is the vagueness with regards to the notion of "conceiving a clear and distinct idea". If one conceives clearly and distinctly X without Y, then it is possible that X exists without Y since God must be able to produce in reality any conceivable distinction in one's mind. The link between conceivability and possibility depends on the proof of God and God's ability to produce anything that is conceivable. Descartes, however, does not present a detailed constraint to the type of conceivability that entails metaphysical possibility, that is, to the kind of mental act (conceivability/imagination) that would be immediately committed to the existence of what it is conceived, nor he presents an argument for the entailment in question. Thus, it is not clear which kind of conceivability leads to possibility or even if any kind of conceivability would lead to metaphysical possibility at all. Faced with that difficulty, Chalmers formulates an updated version

---

2 Here it is important to notice that matters concerning my treatment of anti-physicalist arguments such as the conceivability argument are *entirely* conceptual and therefore independent of any empirical contributions to the general area of Philosophy of Mind.

of the argument with the aid of his interpretation of the two-dimensional semantics in order to build a licit link between conceivability and possibility.

2 Two-dimensional argument against materialism

The term "qualia" refers to phenomenal aspects of our mental lives, that is, what it is like to undergo some experience, such as what it is like to see red, to smell freshly ground coffee, new mown grass etc. However, which (sort of) mental states have qualia? It is still controversial to ascribe this particularity of experience to mental states such as beliefs and desires[3]. Cognitive states are paradigmatic examples of mental states with no *phenomenal feel*. My belief that 2+2=4 or that the Earth is oval has nothing that *it is like*—at least not in the same sense as feelings of pain and other perceptual experiences do. So, we need to distinguish between mental states that have phenomenal feel from mental states that do not have phenomenal feel. Typically, we say that perceptual states have phenomenal feel, whereas cognitive states do not.

Physicalists need to address powerful arguments involving these phenomenal aspects (qualia) of our mental lives. These arguments are about our epistemic, subjective situation confronted with an objective view of the world. This confrontation opens an epistemic gap between the way we introspect aspects of our experience and the physical explanation of reality. We will focus on one central argument against physicalism, namely, *the conceivability argument* (Chalmers 1996) in two versions, the zombie version and the kripkean version. The argument consists in the assertion of an epistemic gap between the physical and the phenomenal domain, and then proceed to claiming a metaphysical gap. The conclusion is the falsehood of physicalism. It turns into a contradiction the apparent difficulty of accommodating qualia within a physicalist framework.

Let us consider a creature—physically and functionally—identical to me but lacking subjectivity. The creature behaves just like me; it is molecule for molecule identical to me, but it lacks the subjective character of consciousness; it is experientially empty; it has no feeling such as 'Oh! So, this is what it is like to see the beach for the first time'. While I have a certain feeling when I taste Swiss dark chocolate, my zombie-twin would react in the same way as I would: she would present the same behavioral responses but would lack the special feeling I get when I taste Swiss dark chocolate. The creature processes the same information as I do: she processes the same perceptual data as me, thus she produces the same behavioral outputs as I do. Nevertheless, she lacks the phenomenal experience which I possess. As Chalmers (1996: 95) puts it: 'there is nothing like to be a zombie.'

It is plausible to deny the *nomological* possibility of zombies since there are not and there could not be such creatures in the actual world. Nonetheless, we are dealing here with the *metaphysical* possibility of a creature with absent qualia. The idea that physicalism is incompatible with the metaphysical possibility of zombies is pretty straightforward: If I had a zombie-twin (a physical duplicate of me) then there could be, according to physicalism, no difference *simpliciter* between us. Contrapositively: If there were a difference at the phenomenal level, then, according to physicalism, we could not be perfect twins. The conclusion is that these phenomenal states are not entirely physically determined. The metaphysical possibility of an absent qualia creature immediately violates the formulation of minimal physicalism[4]:

> (1) Any world which is a *minimal* physical duplicate of *our* world is a duplicate *simpliciter* of our world.

To consider the possibility of zombies is to consider that (1) is contingent. Considering the metaphysical possibility of the existence of zombies is to consider a duplicate of our world

---

[3]   There is a growing debate in Philosophy of Mind as to whether phenomenality can be ascribed to intentional states such as beliefs and desires. The debate is known as both 'Cognitive Phenomenology' and 'Phenomenal Intentionality'. I will not engage in this matter here. On the contrary, I will consider the traditional view that ascribes phenomenality to a specific kind of state.

[4]   "Minimal Physicalism" stands for the minimal requirements for a physicalist theory of mind.

lacking some special feature, and that alone violates (1). Following Chalmers (1996) we can now see how the zombie argument is formulated:

(P1) $P \wedge \neg Q$ is conceivable.

(P2) If $P \wedge \neg Q$ is conceivable, $P \wedge \neg Q$ is metaphysically possible.

(P3) If $P \wedge \neg Q$ is metaphysically possible, physicalism is false.

C.: Physicalism is false.

Let P stand for all physical truths in the world and Q stand for all the phenomenal truths. In (P1) the physical properties are kept constant, whereas the phenomenal properties vary. As explained previously, to conceive of a physical duplicate lacking phenomenal states is to conceive minimal physicalism as false. Further, as conceivability implies metaphysical possibility, the metaphysical possibility of zombies is inconsistent with physicalism. The conceivability argument is clearly valid. Physicalists need to show that at least one of the premises is false. Objections to the argument will typically question the first two premises: (i) are zombies conceivable? If they are conceivable, (ii) does it follow that they are possible? The proponent of the conceivability argument must answer positively to both questions. Consequently, physicalists will say no to either the first or the second question. We shall now examine the possible physicalist reactions.

2.1 Zombies are not conceivable

One option is to reject (P1) which asserts the conceivability of zombies. Analytic functionalists choose this path by claiming that the meaning of phenomenal terms can be analyzed in functional terms. Terms that designate phenomenal states, such as 'pain', have their meaning fixed by whatever plays the functional role of pain: whatever causes behavior typically related to pain. If what plays the functional role of pain is the physical stuff underneath it, then the physical stuff, for example, 'stimulation of c-fibers', plays that functional role in any minimal physical duplicate of our world. If this is the case, there would be no way of conceiving zombies; the relevant functional role of 'Q' is

played by 'P'; so, whatever instantiates P will also instantiate Q. Therefore, the conceivability of zombies is rendered impossible. Nevertheless, analytic functionalists must explain at least why zombies *seem* conceivable.

Those physicalists might suppose that our conceiving of zombies is somewhat deficient, not meeting ideal standards. The notion of conceivability, which the proponent of the argument employs here, is the one that abstracts from our limited cognitive capacities. Since conceivability at stake here is not ideal conceivability, we are subject to error regarding the conceivability of many propositions. One paradigmatic example is the conceivability of complex mathematical truths. It is said of the Goldbach conjecture that its truth *and* falsity are conceivable, but that cannot be the case since one of the two options is a priori truth and the other a priori false. These only seem conceivable because our limited reflection skills make that error. The problem with this kind of response is that it presupposes the complicated notion of idealized conceivability.

To reject (P1) one must claim Q is a priori deducible from P by an ideal reasoner. I will agree with Chalmers regarding the strong intuition behind the truth of (P1), and against the a priori deducibility of Q from P. It is quite intuitive that given the complete microphysical conception of the actual world, the falsity of 'Q' would not be ruled out, despite the level of knowledge and cognitive capacity the reasoner has.

2.2 Link between conceivability and possibility

The most popular physicalist response to the conceivability argument consists in rejecting the link between conceivability and possibility. There are different kinds of possibilities, there is metaphysical possibility, logical possibility, nomological possibility etc.. The relevant kind of possibility here is metaphysical possibility. Consider, first, the contrast between nomological and metaphysical possibility. It is conceivable that tele-transportation exists, but the existence of tele-transportation is surely not *nomologically* possible. The laws of physics that rule our world

do not allow for such possibility. However, if the world were ruled by different physical laws, tele-transportation would be a tenable possibility. This is a metaphysical possibility: If the world were to be different in such-and-such ways, x would be the case, so x is metaphysically possible. If there were no way the world could have turned out that x would be the case, we say that x is not metaphysically possible. Examples of metaphysical impossibilities are contradictory thoughts: there is no way a world could turn out in which square circles would be metaphysically possible.

However, many philosophers reject this link mainly by focusing on counterexamples to the alleged bridge between conceivability and possibility. Two famous examples are: (i) the Goldbach conjecture or its negation are conceivable but one of them is impossible and, (ii) a posteriori identities are conceivable though metaphysically impossible. The issue with a posteriori identity statements is that, after taking Kripke's considerations into account regarding the identity between rigid designators, we are inclined to agree that every identity is necessary, therefore, this makes them metaphysically impossible to separate. Nonetheless, it is quite reasonable to imagine them coming apart, i.e., to imagine that Hesperus is not Venus or that Gold is not Au79. These counterexamples suggest that we are dealing with a very specific notion of conceivability that needs to abstract away from any rational limitations that constraint our reasoner's abilities to conceive, among other things. In order to argue in defense of the conceivability-possibility link, one needs to specify which notion of conceivability yields metaphysical possibility. Chalmers (2002) makes an inventory of kinds of conceivability and concludes that only one particular kind of conceivability is safe guide to metaphysical possibility. This is where Chalmers' two-dimensional framework comes in to explain what is required to safely pass from conceivability to possibility.

2.3 Two-dimensional semantic framework

"Two-dimensional semantics" designates a set of semantic theories within intensional semantics to which meanings are intensions. Two-dimensional semantics postulates in addition that the truth-value of certain sentences holds a double dependence *vis-à-vis* possible worlds. To represent the two ways that truth-values depend on possible states of affairs, two-dimensional semantics systematically assigns a pair of intensions to each linguistic expression: a *primary intension* and a *secondary intension*. First, one-dimensional semantics ascribes only one intension to each linguistic expression. This intension picks out the way truth-values depend on facts, whereas the two-dimensional framework aims to pick out another kind of dependence, viz. the truth value of a sentence *vis-à-vis* what it means or conveys. The two-dimensional semantics generalizes the double-indexing strategy developed to deal with certain expressions, such as indexicals, demonstratives and tense terms ('here', 'now', 'that'…). This is an uncontroversial application of the apparatus. A more ambitious interpretation of two-dimensional semantics generalizes the uncontroversial applications to apply to *all* sort of expressions. Such ambitious interpretation seeks to isolate a priori aspects of meaning. This generalization departs from two core ideas: there are two ways a linguistic expression depends on possible worlds. First, the primary extension of an expression depends on the nature of the actual world in which the expression is uttered. Second, the secondary extension of an expression depends on the nature of the world in which the expression is counterfactually evaluated. Corresponding to these two kinds of dependence, there are two kinds of intension. An intension is a function that takes an expression relative to a world. So, the intension of the expression e is a function from a possible world w to the object that satisfies e. The extension of a sentence is a truth-value in a particular world, whereas the intension is the proposition expressed. The two-dimensional framework stipulated two kinds of dependences of expressions on possible worlds, we have two ways of considering possibilities: (i) the possibilities represent the way the actual world might have turned out to be, which is equivalent

to "considering a possibility as actual" or to consider that the world we are evaluating is *our* world. (ii) The other way to consider possibilities is to "consider a world as counterfactual". In the latter, the actual world is already fixed and the extension of the linguistic expressions will have the same truth value at counterfactual worlds as they do at the actual world. Or to put it in Chalmers' vocabulary, the possibility that represents the world as actual is a *primary possibility* and the possibility that represents the world as counterfactual is a *secondary possibility*.

Considering a possible world *w* as actual is to consider the possibility that *w* is *our* world, that is, to consider the possibility that the actual world could have turned out to be different. Thus, it is possible to consider that our world is such that the watery stuff in it is XYZ and not H2O. This is how Putnam's doppelgänger at Twin-earth would perceive his world when reflecting on the reference of Twin-water across worlds. It is a reflective exercise that takes our actual world and raises the hypothesis that other worlds could also be actual. Then we should evaluate meanings across *w* as if *w* were actual. By contrast, considering a world as counterfactual is to think of a different possibility under the condition that the meaning of the expression is fixed in the actual world. Therefore, according to the two-dimensional semantics, if an expression is evaluated relatively to a world *w,* it has two intensions as a result, depending on how *w* is conceived (actually or counterfactually). As Chalmers puts it, it is primarily possible that water is XYZ but it is not secondarily possible that water is XYZ, but it is secondarily possible that water is H$_2$O.

The two kinds of intensions characterized above represent two dimensions of meaning. At this point, it is convenient to mobilize semantic matrices to visualize the double-dependence that two-dimensional framework stipulates and each of the corresponding intensions. Consider the following sentence containing an indexical term:

(2) I am sick.

The relevant possibilities are: at the world *i*

Mary is the utterer and at the world *j* Peter is the utterer of (2). Mary is sick in both worlds, *i* and *j*: Peter is not sick in neither *i* nor *j*. We can represent the double-dependence of the truth value of (2) by mobilizing the semantic matrix:

$$\begin{pmatrix} & i & j \\ i & 1 & 1 \\ j & 0 & 0 \end{pmatrix}$$

The worlds represented in the vertical axes are worlds considered as actual, or primary possibilities. The worlds represented in the horizontal lines are considered as counterfactual. The secondary intensions of (2) are represented in the horizontal lines of the matrix: If we consider *i* as actual, then (2) is true are *j* considered as counterfactual. At the world *i*, 'I' identifies Mary, so the term 'I' rigidly designates Mary in *j*. Mary is sick in both *i* and *j*, since the actual world is where the reference of the terms is fixed, once the reference is fixed, we are in a position to evaluate the sentence in a counterfactual world. Nevertheless, if we consider *j* as actual (the second line of the matrix), then (2) is false at *i* considered as counterfactual. At *j*, the term "I" picks out Peter, therefore, the term "I" rigidly designates Peter at *i*. The primary intension of (2) is represented by the diagonal of the matrix. The primary intension is true at *i*, since, by considering *i* as actual and by evaluating (2) in the actual world, the output will be the "true". In the actual world, "I" picks out Mary and she is sick at *i*. When considering *j* as actual, "I" picks out Peter in *j* and Peter is not sick in *j*, hence the result is "false".

Possible worlds *w* play the role of contexts of utterances that determine the extensions in *w*. The extension of an indicative sentence is a truth value: true or false. The primary extension of a sentence depends on the world in which the sentence is uttered. In this case, the context of utterance determines the truth-value of the sentence. The secondary extension of an expression in *w* depends on the worlds considered as counterfactual—it is no longer the nature of the world in which the sentence is uttered.

Now consider the sentence (3):

(3) There is water here.

The matrix below represents the two dimensions of meanings of the sentence (3). The possibilities w, v, u represented in the vertical axes are *worlds considered as actual*, and the possibilities represented by horizontal lines are *worlds considered as counterfactuals*. The relevant possibilities are: at w 'water' designates H2O, at v 'water' designates XYZ and at u 'water' designates KLM.

$$\begin{pmatrix} & w & v & u \\ w & 1 & 0 & 0 \\ v & 0 & 1 & 0 \\ u & 0 & 0 & 1 \end{pmatrix}$$

The matrix represents the two ways to evaluate a sentence: the secondary intension of (3) is a function from worlds considered as counterfactual to extensions. Keeping the actual meaning of "water" (H_2O) and evaluating (3) in counterfactual worlds will yield the following results: "water" designates H_2O in every possible world so (3) will be true in w but false in v and u considered as counterfactual since only in w "water" designates H_2O. The secondary intension of (2) is represented in the line of the matrix.

The primary intension of (3) is represented by the diagonal in the matrix. Considering w, v and u as actual, (3) will be always true, since regarding primary intension, "water" will pick out whatever plays the watery role in the circumstances of evaluation, and at w is H2O, at v is XYZ and at u is KLM. We can observe that the primary intension is represented by the intersection points between worlds considered as actual and worlds considered as counterfactual.

2.4 Dimensions of conceivability

With the semantic apparatus in hand, we can now comprehend the conceivability argument as formulated by David Chalmers. The argument operates by producing ontological conclusions from epistemic premises. The real work in this argument is done by bridging the epistemic and modal domains *via* the two-dimensional framework. Chalmers makes an inventory of different notions of conceivability which will serve as candidates to entail metaphysical possibility. Those new dimensions of conceivability should also accommodate widely recognized counterexamples to the link. He distinguishes between three dimensions of conceivability: prima facie vs. ideal, negative vs. positive, primary vs. secondary. This is a requirement to assess both (P1) and (P2) of the argument. We shall focus on the two latter kinds of conceivability:

Negative vs. positive conceivability

Negative conceivability: S is conceivable if and only if S cannot be *ruled out* through a priori reasoning. (p.143)

Positive conceivability: S is positively conceivable when one can coherently imagine a situation in which A is the case. (p.144)

Negative conceivability explicitly resorts to the notion of a priority and yields close connections with conceptual analysis. It is defined so as to exclude any contradictory sentences such as "round squares", "married bachelors" etc.. Hence, per definition, any sentence which does not contain any a priori contradiction is conceivable. Examples of negative conceivability are "water is not H_2O", "the moon is made out of cheese", "pigs can fly". Conceivability is negative as it is defined in terms of *ruling out* what cannot be excluded by purely a priori reasoning. On the other hand, positive conceiving of a sentence S consists in the ability to *imagine* a coherent situation which verifies S; it is a definition centered in the faculty of imagination. Defining conceivability in terms of 'what can be ruled out a priori' is negative conceivability.

Prima facie vs. ideal conceivability

Prima facie conceivability: S is prima facie conceivable (for a subject) when that subject is unable to rule out the hypothesis expressed by S by a priori reasoning on *initial considerations*. (p.143)

Ideal Conceivability: S is ideally conceivable when the hypothesis expressed by S cannot be ruled out a priori even on *ideal reflection*.

Prima facie conceivability is tied to the subject's contingent cognitive limitations. The absence of ideal cognitive capacities may lead to mistakes in a priori reasoning, such as judging S to be prima facie conceivable on initial consideration and then later, upon deeper reflection, seeing that S is not really conceivable. Alternatively to prima facie conceivability, we have ideal conceivability that abstracts away from the subject's cognitive limitations and requires that the utterer of S has ideal cognitive capacities. Some sentences certainly fail to be even prima facie inconceivable, such as simple mathematical falsehoods, like "2+2=5", other sentences, such as complex mathematical truths are prima facie conceivable as false but, ideally inconceivable as false.

We can combine prima facie and ideal conceivability with negative and positive conceivability. We have already considered examples of prima facie and ideal *negative* conceivability. A sentence S is *prima facie positively conceivable* if a subject can imagine a situation that she takes to be coherent (on first reflection) verifying S. A sentence S is *ideally positive conceivable* if its coherent imaginability cannot be ruled out a priori on ideal reflection. One paradigmatic case of positive conceivability in philosophy is exemplified in Descartes' conceivability notion. He claims to have "clear and distinct ideas", which is equivalent to imagine a scenario that is coherent on sustained reflection and which verifies some sentence S.

It is clear that ideal conceivability is a better candidate to entail metaphysical possibility than prima facie conceivability. Prima facie conceivability is susceptible to failures in reasoning resulting in falsely considering a sentence as conceivable or in failing in see its inconceivability. Examples that illustrate the asymmetrical advantage of ideal over prima facie conceivability involves the conceivability of complex mathematical truths. The inferiority of prima facie conceivability is made clear by obvious counterexamples against the entailment between conceivability and possibility. Here is one:

The Goldbach conjecture is claimed to be both conceivable as true *and* false, but that cannot be the case since one of the two options is a priori true and the other a priori false. Both options seem conceivable due to our limited reflection skills. The distinction between prima facie conceivability and ideal conceivability accommodates this counterexample: the Goldbach Conjecture is prima facie conceivable as false *and* as true, but it is ideally conceivable either as false *or* as true.

Primary vs. secondary conceivability

Another counterexample to the link between conceivability and possibility, which is not accommodated by distinguishing between prima facie and ideal conceivability concerns the so-called kripkean modal hybrids. This counterexample requires an additional dimension of conceivability:

Kripke's notorious analysis of the necessary a posteriori yields the following results: It is claimed that necessary a posteriori sentences are conceivable as false despite their being metaphysically impossible: "Water is not H2O" is conceivable as true, but, because the terms involved in the identity statement are rigid designators, the statement is necessarily false, hence not metaphysically possible.

According to Chalmers, there is a sense of conceivability in which the sentence "water is not H_2O" is conceivable, and a different sense in which it is not. This is where the two-dimensional framework comes in, to aid in the distinction between an additional dimension of conceivability: primary and secondary conceivability.

Primary conceivability: S is primarily conceivable when it is conceivable that S is actually the case.

Secondary conceivability: S is secondarily conceivable when S conceivably might have been the case.

There are two senses of conceivability in play here, one is Kripke's sense in which the sentence

(4) Water is not H_2O.

is conceivable but impossible, and another sense in which the sentence is not even conceivable. This requires the distinction above. The sentence (4) is primarily conceivable but not secondarily conceivable. This distinction corresponds to two ways of considering possibilities which we have previously contemplated: primary possibility is tantamount to considering a world as actual, while secondary possibility is tantamount to consider a world as counterfactual. Primary conceivability is tied to a priori knowledge in the sense that, for all we know a priori, it is conceivable that water is not H_2O. Secondary conceivability takes into consideration empirical facts about the actual world and empirical facts about our world that exclude the possibility of water not being H_2O. This marks a different approach in understanding Kripke's analysis of the necessary a posteriori: "water is not H_2O" cannot be ruled out a priori, so it is primarily conceivable. However, the sentence can be empirically ruled out, hence it is not secondarily conceivable since "water" designates H_2O in every possible world.

We have now two additional senses of conceivability: on the one hand, we are evaluating what *plays the watery role*; on the other hand, we ask about the actual reference of "water". Combining primary and positive conceivability, "Water is not H_2O" is primarily positive conceivable in the sense that we can coherently imagine a situation in which watery stuff—the liquid that fills rivers and lakes, that we drink when we are thirsty, that falls from the sky etc.—picks out something other than H_2O. But the sentence is not secondarily conceivable if we are to inquire about the reference of 'water' in whatever counterfactual scenario, since it is always H_2O.

This is the distinction that Chalmers suggests as the way to verify which notion of conceivability is the best guide to metaphysical possibility and to account for the alleged counterexamples to the link between conceivability and possibility. Instances of the necessary a posteriori such as 'water is not H2O' have two different intensions associated with it: they have a *contingent primary intension* and a *necessary*

secondary intension. Two ways of considering the modal status of the statement correspond to two ways of considering conceivability and possibility, as previously pointed out.

(5) Water is H_2O

is primary conceivable if one can conceive of a possible world where the primary intension is true (strongly tied to a priority). (5) is secondarily conceivable if one can conceive of a world where the secondary intension is true (this is not tied to a priority, on the contrary, this is the way to represent empirical facts of the world). "Water is not H_2O" is primary conceivable as we can conceive a XYZ-world lacking H_2O, nevertheless containing watery stuff, and it is not secondarily conceivable since the secondary intension is contingent: Taking the semantic facts of the actual world (where water is H_2O) as fixed, the statement is false at the XYZ-world.

Now, we can say that a statement is primarily possible at a world that verifies the statement (where the primary intension is true) and secondarily possible at a world where the secondary intension is true. What kind of conceivability entails possibility? Primary conceivability entails primary possibility and secondary conceivability entails secondary possibility. The link between conceivability and possibility is not controversial if we accept Chalmers's dimensions of conceivability and possibility. It is clear that (cf. Chalmers 2010):

(CP+) Ideal primary positive conceivability entails primary possibility.

(CP--) Ideal primary negative conceivability entails primary possibility.

This distinction leads to a refinement of the conceivability argument (Chalmers 2010: 148):

(P1) $P \wedge \neg Q$ is 1-conceivable.

(P2) If $P \wedge \neg Q$ is 1-conceivable, $P \wedge \neg Q$ is 1-possible.

(P3) If $P \wedge \neg Q$ is 1-possible, physicalism is false.

C.: Physicalism is false.

However, primary possibility presents no threat to physicalism, so (P3) is false. The falsity of physicalism requires the secondary possibility of $P \wedge \neg Q$. Primary conceivability is a safe guide to primary possibility but it is not a safe guide to secondary possibility in the examples considered so far. There is still a gap between primary possibility and secondary possibility. Chalmers proposes a unique way to close the gap between primary possibility and secondary possibility. For, there are certain cases in which primary conceivability does entail secondary possibility.

In standard cases of theoretical identities, taking (4) "water is not H_2O" to be conceivable but not possible is actually to take (4) primary conceivable, but not secondarily possible. Primary conceivability is not in general a good guide to secondary possibility. Rather, secondary conceivability is a good guide to secondary possibility and since (4) is not secondarily conceivable, it cannot be secondarily possible by that inference. This explains that conceivability is usually not a guide to possibility: it is the wrong king of conceivability we are considering. However, there are some special cases in which primary conceivability entails secondary possibility. And this is central to the conceivability argument.

If some linguistic expression Q has coinciding primary and secondary intensions, then the same possibilities will verify Q and satisfy Q since intensions are defined as functions from possibilities to truth-value. If Q has the same truth-value regardless of the possibility in which Q is evaluated, then there is no gap between primary and secondary possibility. In order for the sentence to be secondary possible, both P and Q must have coinciding primary and secondary intensions. then the same possibilities will verify Q and P.

In this sense, primary conceivability of $P \wedge \neg Q$ will also be secondary conceivability, so primary conceivability will, after all, imply secondary possibility. Chalmers thinks that there is only a gap between primary possibility and secondary possibility if the primary and secondary intension of the expressions in the sentence differ. If they, instead, return the same truth-value, then primary possibility

and secondary possibility will also coincide.

But what kind of special cases are these in which primary conceivability entails secondary possibility? They are those involving linguistic expressions that designate phenomenal properties, hence, cases involving phenomenal terms and microphysical terms like H_2O. It is somewhat uncontroversial that phenomenal terms have coinciding primary and secondary intensions. The reason for this *stability* of phenomenal terms is quite straightforward: we have evidence that primary and secondary intensions differ when the intension of the linguistic expression and its extension seem to come apart. However, when the linguistic expression designates a phenomenal property, there can be no conceiving of phenomenal properties as different from their appearance, since it is the appearance of phenomenal properties that makes them what they are. There is a clear difference between the appearance of contingency of phenomenal terms and the appearance of contingency of physical terms. For there is a potential dissociation between appearance and reality in the case of "water" and "heat", on the one hand, which does not occur in the case of conscious phenomena such as "pain", on the other hand. It is clear that something can look like water, or satisfy the watery role, but not be water, but it is clearly false that something might feel like pain and fail to be pain, since pain is essentially what feels like. The two-dimensional representation of phenomenal concepts must distinguish them from other rigid designators like "water'" that have a constant secondary intension, but a variable primary intension: the secondary intension of phenomenal concepts coincides with their primary intension. Of course, the coincidence of primary and secondary intensions of any phenomenal concept is no independent evidence for the claim that it affords us insight into the very essence of experiences.

To close the gap between primary and secondary possibility, microphysical truths P must also be semantically stable—have coinciding primary and secondary intensions. There is an intuitive sense in which terms which designate

microphysical truths like "mass", "H_2O" are not twin-earthable for the same reason as phenomenal terms are not twin-earthable, these terms already designate their referents essentially. There is no way the world could turn out to be that "H_2O" would not be H_2O. Of course, this is the opposite from rigid designators, which have a variable primary intension. However, it could be argued that P does not have coinciding intensions. The primary intension of a physical term P could be whatever plays the P role, whereas the secondary intension of P is tied to the property that *actually* plays the P-role. However, let us consider that P has coinciding primary and secondary intensions for the sake of the argument.

Now we can correctly display the formulation of the conceivability argument:

(P1) $P \wedge \neg Q$ is 1-conceivable.

(P2) If $P \wedge \neg Q$ is 1-conceivable, then $P \wedge \neg Q$ is 1-possible.

(P3) P and Q have coinciding 1 and 2-intensions. (P and Q are semantically stable)

(P4) If $P \wedge \neg Q$ is 1-possible, then it is 2-possible. (from (P1), (P2) and (P3))

(P5) If $P \wedge \neg Q$ is 2-possible, then physicalism is false.

C.: Physicalism is false.

The fact that P and Q have coinciding primary and secondary intensions assures the entailment form primary possibility to secondary possibility. Hence physicalism is really threatened by the conceivability argument. Of course, at this point one would argue against the coincidence of primary and secondary intensions of phenomenal terms or of microphysical terms. But we will grant that for the sake of the argument.

3 Kripke's modal argument

Kripke's conceivability argument against materialism follows Descartes' original pattern, but it is grounded on his own views about identity and modality. Consider the identity:

(8) P = Q

where Q stands for "pain" and P stands for a brain state like "stimulation of c-fibers":

(P1) If $P \wedge Q$ are a rigid designators, then it is necessary that P is Q.

(P2) $P \wedge Q$ are rigid designators.

(P3) $P \wedge \neg Q$ is conceivable.

(P4) If $P \wedge \neg Q$ is conceivable, then $P \wedge \neg Q$ is possible.

(P5) $P \wedge \neg Q$ is possible. (from P3 and P4)

C.: P is not Q. (from P1, P2 and C5)

(P1) and (P2) are assured by Kripke's semantics. Q is a rigid designator by stipulation. P is a rigid designator because "pain" cannot pick out something other than the feeling of pain; there is no way that pain can come apart from the appearance of pain, so it designates pain across possible worlds. According to Kripke's semantics, identity between rigid designators is always necessary. (P3) is to a certain extent uncontroversial. We can indeed conceive of pain and stimulation of c-fibers coming apart as we can conceive of the falsity of any necessary proposition provided it is a posteriori. (P4) is the key premise of conceivability arguments in general— the link between conceivability and possibility. Kripke argues against the link regarding other kinds of identity with regards, in particular, to theoretical identity statements. He argues by explaining away the appearance of contingency of those identities. However, he thinks that the appearance of contingency of psychophysical identity cannot be explained away. Because there is no asymmetry between appearance of pain and pain[5] as there is in theoretical statements.

According to Kripke, the trick to explain away the appearance of contingency is to point out that the identity only seems contingent because of how the referents of the terms are fixed. The reference of "water" and the reference of "H_2O" coincide only contingently: It is a contingent

---

[5] This is the same linguistic feature presented in Chalmers' zombie argument which I called "semantic stability".

fact that what has the appearance of water is H_2O. To illustrate this, Kripke asks us to consider a qualitatively identical situation in which "heat" does not designate "molecular motion" and yet someone, who is in this qualitative epistemic situation would not be able to distinguish both situations. Heat would feel like heat in both situations, but since in the twin-heat scenario, there is no molecular motion, we say that there is no heat. Because we cannot qualitatively distinguish both situations, we can conceive of the identity being false.

In the psychophysical identification, in contrast, there is no gap between the qualitative epistemic situation and the actual situation. They both coincide, so the appearance of contingency remains without explanation. This is so because, contrary to the theoretical identity statements, it is not a contingent fact that pain feels like pain. Consider now an epistemic qualitatively identical scenario where there is pain but no stimulation of c-fibers. In this twin-pain scenario, there must be something else that produces the sensation of pain. Nevertheless, this is still a scenario in which there is pain. This contrasts with the twin-heat scenario, where there is the sensation of heat without molecular motion. The disanalogy in the twin-pain scenario is revealed because what plays the pain-role must be pain regardless of what pain is (stimulation of c-fibers or something else). The appearance of the phenomenal state and the phenomenal state do not come apart.

To put this in Kripke's technical terms, the reason why the strategy applied to standard theoretical identification does not work in psychophysical identification is that, in the former case, the reference of "water" is fixed via the referent's contingent properties: the reference of "water" is picked up by superficial and contingent features such as being watery stuff. In contrast, the reference of a phenomenal term such as 'pain'' is not fixed via its contingent properties, rather, it is fixed directly by its essential immediate phenomenal property. In theoretical cases, the terms are twin-earthable: it is possible that their appearance and their nature come apart, in phenomenal case they

are not. Hence, we cannot explain the appearance of contingency. In a nutshell, if they *seem* to come apart and we cannot explain away their separation, then they *are* different.

Kripke's argument was originally formulated to refute the Type Identity Theory. Nevertheless, the argument can also be adapted to attack our formulation of physicalism. We just have to substitute the first two premises of the argument and thus obtain the zombie argument. Instead of the identification thesis, we can take the psychophysical conditional $P \to Q$. The psychophysical conditional must be necessary for physicalism to be true. If the conditional $P \to Q$ is necessary, then any essential property of P must entail an essential property of Q. Granting the possibility of the mind existing without the body requires either abandoning the necessary connection between them, or showing that the possibility of distinction is merely an appearance. According to Kripke, such explanation is not available, hence the psychophysical conditional is false.

Conclusion

Chalmers' and Kripke's conceivability arguments both depart from the logical/conceptual distinction between the physical and the phenomenal to arrive at the possibility of their distinction. Both arguments make the same point, in fact, the only difference between them is the way they justify the inference from conceivability to possibility. Kripke justifies this in terms of his direct reference theory and how words have their meaning fixed, whilst Chalmers does it by means of his interpretation of two-dimensionalism.

If the two-dimensional analysis of a posteriori necessities is correct, it should work either for psychophysical identities or conditionals. If "pain" is identical to "stimulation of c-fibers", then it is not secondarily possible that there be a physically identical world with no pain (zombie-world). But then what are we conceiving when we conceive of zombie-worlds? We are primarily conceiving that there are zombies and we are conceiving of a world where the primary intension of "pain is stimulation of c-fibers" is false. To find this

proposition, we must locate the primary intension of "pain", and it seems that here the primary intension is something like "the unpleasant feeling that I get when I pierce my ears" and—assuming the a posteriori identity holds—the secondary intension is the basic physical description of "stimulation of c-fibers". The primary intension corresponds to a priori facts about pain, whilst the secondary intension corresponds to empirical facts about pain. So, according to two-dimensional semantics, we are allowed to infer that there is a scenario in which "that unpleasant feeling" does not pick out anything despite the fact that there is stimulation of c-fibers. This is the zombie world, since it involves the supposed physical part of pain without the phenomenal part. Thus, since two-dimensionalism vindicates the inference from primary conceivability to secondary possibility, the conceivability of zombie worlds assures their primary possibility (at least in the phenomenal case), so physicalism as stated in the psychophysical conditional, is false.

The two physicalist reactions to the conceivability argument previously considered are: the rejection of the conceivability of zombies and the rejection of the link between conceivability and possibility. The first reaction is implausible: we have accepted that $P \wedge \neg Q$ is conceivable, since it is not a priori deducible that $P \wedge \neg Q$. Conceivability is a priori consistency. If P and Q are conceptually independent, it is reasonable to accept that $P \wedge \neg Q$ is conceivable. The second physicalist reaction is to reject the link between conceivability and possibility. This route is not so trivial, for early attempts to reject this link were based on the existence of counterexamples examined previously such as the Goldbach conjecture and instances of necessary a posteriori. However, two-dimensional semantics explains why the so-called counterexamples are not counterexamples at all, since the link between conceivability and possibility is far more subtle than initially presupposed. As explained, the conceivability argument depends on the distinction between different dimensions of

modality and conceivability. If this dependence is correct, then we are left with the work of rejecting the generalized version of two-dimensionalism to break the connection between conceivability and possibility. This should be one way to go. Another option consists of breaking the connection by mobilizing phenomenal concepts. That would require new definition of physicalism. Instead of a priori physicalism, one should explore the prospects of developing a version of a posteriori physicalism, in which the connection between phenomenal truths and physical truths is still necessary but a posteriori. The strategy that mobilizes phenomenal concepts to defend physicalism from the epistemic arguments such as the conceivability argument is called the *phenomenal concept strategy*[6]. Any account of phenomenal concepts should explain: why are phenomenal concepts conceptually different from physical concepts --- this would contemplate the epistemic gap, or the conceivability of zombies— and how can physicalism be a posteriori if the terms in question are semantically stable. The second constraint would account for the change in the definition of physicalism.

The analyses of both arguments should make explicit how physicalists should proceed. The conceivability argument threatens a priori physicalism, that is, the metaphysical thesis that the phenomenal supervenes with metaphysical necessity on the physical and that mental truths are a priori entailed by physical truths. As I hope to have shown by elucidating the two-dimensional argument in section two, the fact that there is a conceptual disconnection between phenomenal truths and physical truths, as stated by the first premise of the argument—about conceivability of zombies--- falsifies the second part of a priori physicalism, that is, there is no a priori entailment between phenomenal and physical truths. The metaphysical possibility of zombies (P and not Q) is a direct violation of the first part of a priori physicalism---supervenience with metaphysical necessity. Now what is left for the physicalist is to explore the possibility of a posteriori physicalism.

---

6   The label is coined by Daniel Stoljar (2005). See Stoljar for the detailed solution.

Therefore, physicalist responses which try to compromise, that is, to accept the epistemic gap of the conceivability argument, but to block the ontological conclusion, came to be so popular. But this should be a topic of another paper. My goal here was merely to unpack what came to be one of the most difficult formulations of the most pressing argument against materialism, particularly the link between conceivability and possibility presupposed by each version of the argument. The structure of conceivability arguments, from its original (cartesian) version to the highly formalized version (Chalmers' version) through the kripkean version. All versions are meant to show that the argument still poses a serious threat to any physicalist theory.

References

CHALMERS, D. *The Conscious Mind*. Oxford: Oxford University Press, 1996.

CHALMERS, D. Does conceivability entail possibility? *In*: GENDLER, T.; HAWTHORNE, J. (ed.). *Conceivability and Possibility*. Oxford: Oxford University Press, 2002. p. 145-200.

CHALMERS, D. The foundations of two-dimensional semantics: Foundations and applications. *In*: GARCIA--CARPINTERO, M.; MACIA, J. (ed.). *Two-Dimensional Semantics*. Oxford: Oxford University Press, 2006.

CHALMERS, D. *The Character of Consciousness*. Oxford: Oxford University Press, 2010.

KRIPKE, S. *Naming and Necessity*. Cambridge, Mass.: Harvard University Press, 1980.

LOAR, B. Phenomenal states. *Philosophical Perspectives,* [*S. l.*], v. 4, p. 81-108, 1990.

SCHROETER, L. Two-dimensional semantics. *In*: ZALTA, E. N. (ed.). *The Stanford Encyclopedia of Philosophy.* Stanford: Stanford University, 2021. (Spring 2021 Edition). Disponível em: https://plato.stanford.edu/archives/spr2021/entries/two-dimensional-semantics. Acesso em: 10 jul. 2021.

STALNAKER, R. Assertion. *In*: COLE, P. (ed.). *Syntax and Semantics*: Pragmatics. New York: Academic Press, 1978.

STALNAKER, R. *Content and Context*. Oxford University Press, 1999.

STOLJAR, D. Two conceptions of the physical. *Philosophy and Phenomenological, Research,* [*S. l.*], n. 62, p. 253-281, 2001.

STOLJAR, D. Physicalism and phenomenal concepts. *Mind and Language,* [*S. l.*], v. 20, n. 5, p. 469-494, 2005.

Julia Telles de Menezes

Doutora em Filosofia pela Goethe Universität, em Frankfurt am Main; professora adjunta da Universidade Federal Fluminense (UFF), em Niterói, RJ, Brasil.

Endereço para correspondência

Julia Telles de Menezes

Universidade Federal Fluminense

Departamento de Filosofia

Campus do Gragoatá/ Blocos N

Rua Professor Marcos Waldemar de Freitas Reis, s/n

São Domingos, 24210-201

Niterói, RJ, Brasil