

Originalni naučni rad
UDK 111:141

DANIEL KOSTIĆ

**THE TWO-DIMENSIONAL ARGUMENT AGAINST
PHYSICALISM AND THE CONCEPTUAL ANALYSIS¹**

This paper is divided into three sections. In the first section I briefly outline the background of the problem, i.e. Kripke's modal argument (Kripke 1980). In the second section I present Chalmers' account of two-dimensional semantics and two-dimensional argument against physicalism. In the third section I criticize Chalmers' approach based on two crucial points, one is about necessity of identities and the other is about microphysical descriptions and a priori derivation.

1.

The conceivability argument starts with the thesis to which a materialist must be committed: if the mental states are identical to the brain states ($P=Q$), then a situation in which this identity falls apart should not be possible. This is so because, according to Kripke (Kripke 1980), all identity statements involving rigid designators, if true at all, must be necessary true. The conceivability arguments depart from the claim that a situation in which the brain states take place but phenomenal states are lacking altogether is conceivable. It is argued, then, if such a situation is conceivable, it is possible. Finally, if such a situation is possible, then it is metaphysically possible,

¹ I would like to thank Zivan Lazovic for very helpful and insightful comments on this paper. I am also indebted to Stephen Tiley for proofreading this paper and for providing very useful comments on it.

because as we recall identity has to hold necessarily in order to be true. If this is the case then materialism is false.

Now, how do we come from a purely epistemic claim about what is conceivable to the modal claim about what is possible?

According to Kripke, the only thing we need to make this step is a lack of contradiction in conceiving such a situation. Consider an example with the natural kinds terms. Namely, there is no contradiction in conceiving a situation in which water is not H₂O. However, because “water” rigidly designates H₂O, the argument actually shows that, although conceivable, it is not possible that water is not H₂O. How does this step of the inference work? Since it is an empirical discovery that water is H₂O, the reference of the identity statement is not fixed *a priori*. In other words, the state of affairs in our world might have been different, so water might have turned out to be something else in our world (depending on the actual empirical discovery). Therefore, there is nothing in our knowledge that would contradict the conceivability of the claim “water is something other than H₂O”. Just to give a contrastive example, given the definition of a triangle, a geometrical figure that has three angles, it is neither *a priori* conceivable nor possible for a triangle to have four angles.

Why does this step from conceivability to metaphysical possibility break down in the case of “water = H₂O”? According to Kripke, it is because the reference of the concept of water is fixed to a microphysical description, which in our world is H₂O. On the other hand, if it turned out that the microphysical description of water in our actual world was not H₂O but something else, say XYZ, then the reference of “water” would be fixed to XYZ in the actual world. According to Kripke, rigid designation of natural kind terms is always tied to microphysical description². This, on the other hand, means that whatever we discover at the microphysical level to play the role of water, fixes the reference of “water”. In a sense, it is not the empirical discovery that fixes the reference, because the reference is fixed beforehand, but whatever we believe to play the role of water. This is how rigid designation of the natural kind terms actually works according to Kripke. Therefore, it is only conceivable that water might not have been H₂O, but it is not possible. On the other hand, if some possible world at which water is not H₂O, was actual, then “water is H₂O” would be false considered from that world. And this is where the two-dimensional semantics steps in.

2 Kripke (Kripke 1980) argues that rigid designators get their reference fixed by an act of “baptism”. In the case of natural kinds the scientific community, for example, decides that the reference of certain concepts is tied to certain “essential” properties, such as atomic numbers, or any other microphysical description.

2.

According to Chalmers' interpretation of two-dimensionalism, concepts have two intensions, which are functions from truth conditions to possible worlds and vice versa. The intension that goes from the truth conditions in the actual world to the counterfactual worlds is *the primary intension*. The primary intension tells us what fixes the reference in the actual world. The secondary intension goes from possible worlds to truth conditions in the actual world and tells us how the reference in the counterfactual world is fixed. In other words, a term S is verified in some possible world W considered as actual, if and only if its primary intension is true at W. A term S is satisfied in some possible world if and only if its secondary intension is true there. The two-dimensional semantics is here employed to show why the step from conceivability to metaphysical possibility in the case of (P = Q) works and why it breaks down in the case of (water = H₂O). Let's have a closer look at what is going on here.

In order to utilize his interpretation of two-dimensional semantics, Chalmers distinguishes several kinds of conceivability. Conceivability is here to be understood as an epistemic notion. The first one is what he calls "*prima facie* conceivability", which requires that the subject should not be able to rule out a hypothesis solely by *a priori* reasoning. The second one is ideal conceivability which goes a bit further in trying to abstract from the cognitive capacities of a subject, and claims that a hypothesis is ideally conceivable if it cannot be ruled out *a priori* even on ideal rational reflection (Chalmers 2009).

The opposite case from these two is the already mentioned example with a triangle - given the definition of triangles as geometrical bodies with three angles, conceivability of the triangles with four angles is ruled out *a priori*.

These two kinds of conceivability are examples of negative conceivability. However, there is also a positive conceivability. This sort is best known from Descartes's thought experiments (Descartes 1996). Positive *prima facie* conceivability is the case when a subject can imagine a situation that she takes to be coherent and as the one in which the hypothesis is true. As with negative conceivability, here we also have ideal positive conceivability. We say that a situation is ideally positively conceivable if its *prima facie* positive conceivability cannot be defeated.

However, the most important kinds of conceivability according to Chalmers are primary and secondary conceivability. These two kinds of conceivability can be based on either positive or negative conceivability. What distinguishes them is not just an epistemic perspective or cognitive capacities, but also a sense in which we say something is conceivable. So,

we say, for example, that “water is not H₂O” is primary conceivable, exactly as it is stated, that is, that water is actually NOT H₂O (which in Kripkean terms means that the whole world is otherwise, not just water and H₂O), whereas in the case of secondary possibility we take a sense of conceivable in which “water is not H₂O” merely seems conceivable but is not conceivable. An example for the sense in which we think of the secondary conceivability is to say that when we say “water is not H₂O” we only mean that watery stuff (that seems like water) is not H₂O, but water as we know it is still H₂O.

When we look at the distinction between primary and secondary conceivability in the light of the distinction between the *a priori* and the *a posteriori* knowledge, it is clear that primary conceivability can only be made plausible in the domain of the *a priori*. That is, a hypothesis is primary conceivable if it cannot be ruled out purely on *a priori* reasoning, without any reference to empirical or extra-linguistic knowledge.

So what is the link between primary and secondary conceivability on the one side and metaphysical possibility on the other?

The primary intension of the sentence “water is not H₂O” ($\circ\supset S$ hereafter) is true at the counterfactual world W considered as actual, and its secondary intension at W₁, say our world considered as counterfactual, is false, because at our world water is H₂O. To put it differently, a world verifies S iff its primary intension is true at that world, and a world satisfies S iff its secondary intension is true there considered from the actual world.

So, in order to prove that these steps of inference in the conceivability argument follow *a priori*, Chalmers’s reasoning is based on the distinction between primary conceivability and possibility and secondary conceivability and possibility. How does this work? We said that, for example, water is not H₂O is primary conceivable and therefore primary possible because “water is H₂O” is not an *a priori* truth. Knowledge of “water is H₂O” is not based on the meaning of the terms, linguistic rules, etc, but it is an empirical discovery, which we then employ to understand meaning of the terms. So, for all we know water might have been something other than H₂O in our world. Unlike the case of a triangle, wherein a four-angled triangle is simply incoherent even to conceive of, because if we grasped the meaning of the concept “triangle”, we know, without any reference to extra-linguistic or to empirical knowledge, that a four angle triangle is an incoherent concept, and therefore not possible. So when we say “water is not H₂O” is possible, it only means that there is nothing in our knowledge that would contradict to this statement *prima facie*, in a sense, this is not a statement about just water and H₂O, but the statement about the whole world and its nomological structure. Now, Chalmers wants to show that if we apply the conceivability argument to the case of water/H₂O, what is actually claimed

simply leads to inconsistency, by showing that although in the case of water/H₂O the statement “water is not H₂O” is *prima facie* conceivable, thus primary possible, it is not conceivable and therefore not possible in the sense in which water is still H₂O and there is some watery stuff that has the same macro properties of water, but which is not H₂O, that is “water is not H₂O” is not secondary possible. In the case of primary possibility, if water was not H₂O but otherwise, then as Kripke puts it, the whole world would be otherwise and the statement “water is not H₂O” would be true. This is why two-dimensional analysis comes very handy in these complicated cases. We need the secondary possibility to see what is going on with the secondary intension considered from some centred world.

Let us see the structure of the argument with natural kinds like water and H₂O. Let *S* be the proposition that water is not H₂O.

1) *S* is conceivable;

• This premise says that *prima facie* (it could not be ruled out *a priori*) it is conceivable that water is not H₂O.

2) If *S* is conceivable, *S* is 1-possible;

• This premise says that if the statement is *prima facie* conceivable, then it is primary possible.

3) If *S* is 1-possible, then *S* is 2-possible;

• This premise says that if the proposition “water is not H₂O” were primary possible, then it would be secondary possible, which is not true in this case.

Whereas in the case of the phenomenal the argument goes right through, from primary possibility to secondary possibility, and given the physicalist commitment thesis, the conclusion is not compatible with it - therefore physicalism is false. Let us have that argument formalized as well:

4) $(P \& \circ \supseteq Q)$ is conceivable;

5) If $(P \& \circ \supseteq Q)$ is conceivable, $(P \& \circ \supseteq Q)$ is 1-possible;

6) If $(P \& \circ \supseteq Q)$ is 1-possible, then $(P \& \circ \supseteq Q)$ is 2-possible ;

7) If $(P \& \circ \supseteq Q)$ is 2-possible, materialism is false;

8) Therefore, materialism is false.

So what is going on here compared to the water/H₂O case? The argument here says that if a hypothesis is conceivable then it is primary possible, in another words it is not possible to rule out that hypothesis solely on the ground of *a priori* reasoning. This is the point where epistemic and modal principles meet, because if a conceived hypothesis is not possible to rule out *a priori*, then it is possible. This is the meaning of primary possibility, or the link between conceivability and possibility in Kripkean terms, which has been criticized in particular for this step. In an attempt to block these principle objection, Chalmers goes further in claiming that although the statement “water is not H₂O” is primary possible, the identity of water

with H₂O is established only *a posteriori*; however, once established as an identity it holds necessary. We might say that there are no *a priori* reasons that would contradict to conceiving of a situation in which water is not H₂O in a strict sense (something other than H₂O plays the water role) - this is a sense of the primary possibility – but it is not possible that water is H₂O in the actual world and yet something else in a counterfactual world - this is the sense of the secondary possibility. This distinction clearly shows that primary conceivability does not lead to secondary possibility in the case of “water is not H₂O”. However, “water is not H₂O” is compatible with the inferences from primary conceivability to primary possibility and from secondary conceivability to secondary possibility.

The argument against the identity of water and H₂O breaks down when it comes to the inference from primary conceivability to secondary possibility, because the statement has the primary and secondary intensions that do not coincide. In other words it is not coherent to think of the situation in which a world W1 verifies the statement’s primary intension and a world W2 satisfies its secondary intension, given that they are separated (do not coincide), that is it is not coherent to hold that water at W1 is H₂O and water at W2 is XYZ, since “water is H₂O” is verified at W1 it cannot be XYZ. On the other hand, in the case of phenomenal the primary and secondary intensions of the statement do coincide, which in effect allows for the inference from primary possibility to secondary conceivability, and further to secondary possibility. This is so because, as Kripke noted, in the case of phenomenal there is no dissociation between appearance and reality, to paraphrase him: pain is identified with its immediate quality, i.e. for him how it appears to be in pain subjectively is to be in pain. To put it into more technical terms, dissociations between water and H₂O, or between heat and molecular motion, seem conceivable, but they are only primary conceivable, thus primary possible, because their meanings have two modal dimensions, one that goes from the possible world to truth values or primary intension, which tells us what plays the water role in the centred world; and the other intension that goes from truth values to possible worlds or the secondary intension. Now, since in the case of natural kind terms, primary and secondary intensions do not coincide, it is because of the way in which we learn about their primary intensions (objective, third person perspective), in the case of phenomenal consciousness primary and secondary intensions collapse into one, because of the immediacy of phenomenal knowledge which dissolves the distinction between appearance and reality, in a sense that whatever properties of our experience appear to us, they are not merely appearing, they are real. In other words, for Kripke, painfulness of pain is its essential property, and when when we conceive of a situation in which C-fiber firing is not pain, we can’t have an alternative explanation for this

situation **as have in** the case of heat and molecular motion. Therefore if it seems that the identity between pain and C-fiber firing can fall apart it does fall apart.

What is going on in the case of $(P = Q)$ is that it seems that both W1 and W2 verify the statement. In a sense primary and secondary intensions of the statement “the phenomenal is not neurobiological” seem to coincide. That is the reason why the two-dimensional argument in this case goes through from primary possibility to secondary conceivability and further to secondary possibility. This argument structure purports to be saying that both statements - “pain=C fiber firing” and “C fiber firing =pain”- seem to be true, which is a contradiction.

To put it in other words, in the case of natural kind terms, like “water” or “heat”, there seem to be a clear dissociation between the real nature of things and how they appear in our experience, i.e. we can imagine that something might look like water or heat whereas in fact it is not. Since the identity statement “water is H_2O ” is not based on *a priori* knowledge, there is no *a priori* contradiction in conceiving of a situation in which water is not H_2O , and thus if the hypothesis cannot be ruled out *a priori*, then it is possible. What makes it break down is the fact that given the *a posteriori* necessity of the identity statement “water is H_2O ”, it is not possible that water is something else at W2 and still H_2O at W1. However, what is possible is that if water was something else in the actual (centred) world, then its primary intension would be fixed to whatever that “something else” is, to whatever plays the water role in the counterfactual scenario. In this case the statement “water is not H_2O ” would be true in the actual world. But this only means that the whole world would be otherwise. This makes a strong case, then, that primary and secondary intensions of phenomenal terms do in fact coincide. This in effect makes the two-dimensional argument involving the phenomenal/neurobiological to go through to the secondary possibility, which ultimately means that physicalism leads to contradiction, assuming that the primary and the secondary intensions of phenomenal terms coincide.

3.

My criticism of this account is focused on two points. One is that Chalmers does not provide any argument in support of the claim that coinciding of the primary intensions occurs in all identity statements that involve concepts which are based on distinct modes of presentation. In other words, he believes that in the case of natural kinds the counterfactual situations are conceivable but not metaphysically possible because the primary intensions of the concepts involved diverge. Whereas in the case of

psychophysical identity primary and secondary intensions coincide and therefore what is conceivable is *ipso facto* metaphysically possible. But this does not follow necessary from his account of the two-dimensional semantics. Conceivability of an identity falling apart does not necessarily stem from the diverging primary intensions of the concepts involved. Some concepts, such as phenomenal concepts, may refer directly and thus won't have a priori connections with their referents. If this is the case then it would be conceivable that identity involving such concepts can fall apart but it would not mean it is metaphysically possible for them to fall apart.

If this is right, then it casts a serious doubt about whether this account of two-dimensional semantics is universal applicable to the natural kinds. Apparently Chalmers (Chalmers 1996, 2009) and Chalmers and Jackson (Chalmers and Jackson 2000) maintain that a posteriori necessity of identities involving natural kinds stems from two sources: a priori analysis of meaning, i.e. conceptual analysis, or a priori derivation of any macro-physical fact from the body of complete microphysical descriptions. Block and Stalnaker (Block and Stalnaker 1999, pp. 14-16) use an example of the "explanatory gap" in explaining life to argue against this point. If Chalmers and Jackson were right, then we would be analyzing the concept of "life" in terms of functional roles such as digestion, reproduction, locomotion, respiration and so forth. However, it might have happened that some living organisms never digested, respired, reproduced or moved but still being alive. Perhaps they were based on some other processes to sustain life, e.g. some creatures made out of pure energy so they never had a need to digest, respire or reproduce to be considered alive. If this is right then the analysis of life in terms of digestion, reproduction, locomotion, respiration, etc., would not be a priori. It would merely describe how living things in our known universe function, but that would by no means imply that all living things function in the same way. As Block and Stalnaker put it: "Closing the explanatory gap in the case of life has nothing to do with any analytic definition of 'life', but rather is a matter of showing how living things around here work" (Block and Stalnaker 1999, p. 15). Indeed, the analysis of meaning in terms of the descriptions of the causal or functional roles can't be a priori and it also can't serve universally as a basis for the necessity of identities in the case of natural kinds.

Another problematic issue with Chalmers' take on the two-dimensional semantics is the notion of a priori derivation of any random macro-physical fact from the complete microphysical descriptions. In our case it is a lack of such derivation for phenomenal concepts. Namely, Chalmers and Jackson (Chalmers and Jackson 2000) claim that an explanation of identity that involves natural kinds has to include two steps, one which is a priori, based on some a priori analysis of meaning, as it was already explained, and

the other has to be a microphysical description. However, Block and Stalnaker (Block and Stalnaker 1999) show that in some cases either a macrophysical concept is not inferred from the microphysical description or if it is based on the microphysical description then the inference is not a priori. They use the following example.

9) 60 percent of the earth is covered with H₂O.

10) H₂O is the stuff that plays the right kind of causal role in explaining our use of the word 'water'.

11) Therefore, 60 percent of the earth is covered with water.

The assumption must be that the inference from 9) and 10) to 11) is now mediated by the following conceptual analysis:

12) Water is the stuff that plays the right kind of causal role in explaining our use of the word 'water'." (Block and Stalnaker 1999, pp. 25-26).

The crucial step in the argument that is supposed to grant that 11) follows a priori from the premises 9) and 10) is of course hidden premise 12) However, 12) is not a conceptual truth about the meaning of the word "water". At best it is an abbreviation of an a posteriori and empirical analysis of the right kind of causal role of water. For all we know some other liquid might have played that role.

If this kind of conceptual analysis is available for natural kinds, then it is also available for cases involving phenomenal concepts. Furthermore, it certainly does not constitute a difference between the conceptual analysis of natural kinds and the analysis of phenomena concepts.

In conclusion, Chalmers' approach to fortifying the entailment from conceivability to metaphysical possibility by using two-dimensional modal semantics does not succeed for two reasons. One is that there is no justification for the assumption that diverging primary intensions are the only ground for the necessity of the identity. Phenomenal concepts don't have a priori connections to their referents and still can establish necessary identities. This in effect undermines the very idea of relying on the analysis of primary intensions in examining necessity of certain statements. The second reason is that conceptual analysis does not help us to derive concept's a priori component from the microphysical descriptions. It turns out that this kind of analysis is not universally applicable to explanations of identities of all natural kinds and even if it was there is nothing that prevents us from using it for the psychophysical identity.

Literature

- Balog, K., 1999. Conceivability, Possibility, and the Mind-Body Problem. *The Philosophical Review*, 108 (4), pp. 497-528.
- Benardete, J. A., 1958. The analytic a posteriori and the foundations of metaphysics. *Journal of Philosophy*, 55 (12), pp. 503-514.
- Block, N. and Stalnaker, R., 1999. Conceptual Analysis, Dualism, and the Explanatory Gap. *Philosophical Review*, 108, (1), pp. 1-46.
- Blumenfeld, J. B., 1975. Kripke's refutation of materialism. *Australasian Journal of Philosophy*, 53 (2), pp. 151-156.
- Braun, D., 2006. Names and Natural Kind Terms. In: Lepore E. and Smith B. eds., 2006. *The Oxford Handbook of Philosophy of Language*. Oxford: Oxford University Press.
- Byrne, A., 1999. Cosmic Hermeneutics. *Nous*, 33, pp. 347-383.
- Carnap, R. 1947. *Meaning and necessity: A study in semantics and modal logic*. Chicago Illinois: Chicago University Press.
- Chalmers, D., 1996. *The Conscious Mind*. New York: Oxford University Press.
- Chalmers, D. and Jackson, J., 2001. Conceptual Analysis and Reductive Explanation. *Philosophical Review*, 110, pp. 315-360.
- Chalmers, D., 2002a. Does Conceivability entail Possibility. In: Gendler T. S. and Hawthorne, J., eds. *Conceivability and Possibility*. Oxford: Oxford University Press.
- Chalmers, D., 2009. The Two-Dimensional Argument Against Materialism. In: McLaughlin B. P. and Walter S., eds., 2009. *Oxford Handbook to the Philosophy of Mind*. Oxford: Oxford University Press.
- Davies, M. and Humberstone, I. L., 1980. Two Notions of Necessity. *Philosophical Studies*, 38, pp. 1-30.
- Descartes, R., 1996. Meditations on first philosophy with selections from the objections and replies. In Cottingham, J., ed., 1996. *Cambridge texts in the history of philosophy*. Cambridge: Cambridge University Press.
- García-Carpintero, M. and Macià, J., eds., 2006. *Two-Dimensional Semantics*. Oxford: Oxford University Press.
- Gertler, B., 2002. Explanatory reduction, conceptual analysis, and conceivability arguments about the mind. *Nous*, 36, pp. 22-49.
- Hempel, C., 1969. Reduction: Ontological and Linguistic Facets. In: Morgenbesser, S., Suppes, P. and White, M., eds., 1969. *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*. New York: St. Martin's Press.
- Hill, C. S., 1997. Imaginability, Conceivability, Possibility, and the Mind-Body Problem. *Philosophical Studies*, 87 (1), pp. 61-85.

- Hill, C. and McLaughlin, B., 1999. There Are Fewer Things in Reality that Are Dreamt of in Chalmers' Philosophy. *Philosophy and Phenomenological Research*, 59, pp. 445-454.
- Jackson, F.,1998a. From Metaphysics to Ethics: A Defence of Conceptual Analysis. New York: Oxford University Press.
- Kripke, S.,1980. Naming and Necessity. Cambridge Massachusetts: Harvard University Press.
- Levine, J.,1983. Materialism and Qualia: The Explanatory Gap. *Pacific Philosophical Quarterly*, 64, pp. 354-361.
- Levine, J.,1993.On Leaving Out What It's Like. In: Davies M. and Humphreys G., eds., 1993. *Consciousness: Psychological and Philosophical Essays*, Oxford: Blackwell.
- Levine, J.,1998. Conceivability and the metaphysics of mind. *Nous*, 32(4), pp. 449-480.
- Levine, J.,1999. Conceivability, Identity, and the Explanatory Gap. In: Hameroff, S. R. and Chalmers, D.J. eds., 1999. *Toward a Science of Consciousness III: The Third Tucson Discussions and Debates*. Cambridge, MA: MIT Press.
- Levine, J.,2001. *Purple Haze: The Puzzle of Conscious Experience*. Cambridge, MA: MIT Press.
- Papineau, D.,2002. *Thinking about Consciousness*. Oxford: Oxford University Press.
- Putnam, H.,1975. The Meaning of 'Meaning'. In: Gunderson, K.,ed., 1975. *Language,Mind and Knowledge*. Minneapolis: University of Minnesota Press.
- Stalnaker, R.,1978. Assertion. In: Cole P.,ed. *Syntax and Semantics: Pragmatics*. Vol. 9, New York: Academic Press.
- Tye, M., 2003. A theory of phenomenal concepts. In: O'Hear A.,ed., 2003. *Minds and Persons*. Cambridge: Cambridge University Press.

DANIEL KOSTIĆ

DVODIMENZIONALNI ARGUMENT
PROTIV FIZIKALIZMA I POJMOVNA ANALIZA

(Apstrakt)

Ovaj tekst se bavi dvodimenzionalnim argumentom protiv fizikalizma. Tekst ispituje ulogu primarnih i sekundarnih intenzija u logičkoj derivaciji metafizičke mogućnosti iz zamislivosti u modalnim argumentima protiv fizikalizma. U tekstu se dokazuje da ovaj pristup ima dva fatalna nedostatka. Jedan je da argument ne nudi opravdanje za tvrdnju da su primarne intenzije pojma koje se razilaze jedina osnova za nužnost identiteta. Drugi je da pojmovna analiza ne uspeva da dokaže da su makro pojmovi a priori izvodivi iz potpunih mikrofizičkih opisa.