oner's Dilemma, for example, where cooperation and defection seem to have equal force, the pattern turns out to be unstable similar to the matching pennies game. The pattern reflects the fact that on the assumption that the opponent cooperates, it is rational to take advantage and defect, but on the assumption that both defect, it is rational to jointly cooperate.

A plausible definition should produces patterns that respect all assessments of the undisputed cases, *i.e.*, it should yield the correct categorical verdicts. But it should equally answer to our intuitive assessments even in the pathological cases, such as the one in Figure 2 where despite the pathologicality we can make some categorical assessments.

That the Nash equilibrium does not produce satisfactory answers in cases like the one just mentioned was noted by Luce and Raiffa already.[6] They concluded that a unified theory of the non-cooperative games was not possible without complicating the problem with more contextual information:

> It is unfortunate (or fortunate, depending upon your viewpoint) that a unified theory for all non-cooperative games does not seem possible. The only alternative seems to be to complicate the problem by introducing more initial information in the form of boundary and initial conditions—information referring to personality traits, psychologies of the players, etc. [LucRai57, p. 104]

I like to think that by viewing the concept of rational decision as being governed by a circular definition, we both unify and simplify the theory of non-cooperative games. It is unified by relying exclusively on payoff maximization, no extraneous information needs to be introduced. It is simplified by providing the most natural and direct method to model the circularity inherent in these situations. Both of these advantages are afforded by the revision theoretic semantics which allows us to use circular definitions without inviting the inconsistencies they are traditionally associated with.

[6] The same goes, I claim, for the solutions based on restricting idealizations addressed here. I do not claim that only the revision approach avoids this problem. There may be other approaches that provide satisfactory answers. However, the restrictive approaches do currently seem to be dominating the field to a point that makes my claim reasonably strong even without going into the other approaches.

FotFS
II

Benedikt **Löwe**, Wolfgang **Malzkorn**, Thoralf **Räsch** (*eds.*)
Foundations of the Formal Sciences II
Applications of Mathematical Logic in Philosophy and Linguistics
Papers of the conference held in Bonn, November 10-13, 2000

Two-Dimensionalism and the Metaphysical Possibility of Zombies

Daniel Cohnitz*

Heinrich-Heine-Universität Düsseldorf
Philosophisches Institut
Universitätsstraße 1
Gebäude 23.21/04.73
D-40225 Düsseldorf

E-mail: `cohnitz@phil-fak.uni-duesseldorf.de`

Abstract. Two-dimensionalism is a formal framework used in formal semantics, epistemology and the philosophy of mind. The technical background dates back to the early seventies, in particular to Krister Segerberg's paper "Two-Dimensional Modal Logic". The mathematical tools developed in that tradition can be used to model the relations between two semantical properties of concepts or expressions, which, according to two-dimensionalism, can be conceived to be two kinds of intensions. I shall present the general ideas of two-dimensionalism, and give a brief reconstruction and discussion of one application in the philosophy of mind.

1 Introduction

In this paper, I shall give a brief presentation of the tool "two-dimensional modal logic". I shall leave out the more technical aspects, but I shall

Received: April 12th, 2001;
In revised version: May 10th, 2001; June 1st, 2001;
Accepted by the editors: June 15th, 2001
2000 *Mathematics Subject Classification*. **03A05 03B45**.

* I would like to thank Oliver Beermann, Dieter Birnbacher, Axel Bühler, Anneli Mottweiler, John Perry, Barry Smith, and an anonymous referee for helpful comments on earlier drafts of this paper.

© *2003 Kluwer Academic Publishers. Printed in The Netherlands, pp. 55–62.*

explain the main idea behind it and look at one famous application in the philosophy of mind.

I think there are at least three good reasons to become more familiar with two-dimensional modal logic. One is that it seems to be a promising tool for dealing with various perennial semantical problems. The semantics of belief ascription, the problem of the essential indexical, and Frege's puzzle about non-trivial identities allegedly are all solvable within the two-dimensional framework.[1] A second reason is that David Chalmers tries to use two-dimensional modal logic as a backup for his modal epistemology, in particular in order to distinguish two different notions of conceivability. A third reason for having another look at two-dimensional modal logic is that David Chalmers used this framework to explicate his arguments against physicalism. I shall reconstruct the use made of two-dimensional modal logic in the latter debate, drawing on a paper by Alex Byrne.[2] Finally I shall briefly sketch how to evade the dualist conclusion by a move which does not exclude a two-dimensional semantics. This move is motivated by some ideas of John Perry, but, as far as I can see, it is used for the first time as a direct attack on the two-dimensional version of the argument.

2 A word on the technical background

The formalization of two dimensional modal logic usually referred to as the first is Krister Segerberg's in the paper "Two-dimensional Modal Logic" of 1973. Segerberg himself refers to Hans Kamp and Arthur N. Prior, noting that an anonymous referee of his own paper suggested that Frank Vlach was the inventor of two-dimensional modal logic.

However, in the paper "Two notions of necessity" (1980), Martin Davies and Lloyd Humberstone seem to have constructed a language of modal logic which comes closest to the one I shall address here. In "Two notions of necessity" they basically enriched the conventional language of modal logic with the operators 'actually' and 'fixedly' and extended S5 by adding appropriate axioms. The resulting system S5$\mathcal{AF}$ is proven to be complete in in appendix 10 of Martin Davies' book "Meaning, Quantification, Necessity."[3] Anybody interested in the technical background of two-dimensionalism might wish to have a look there.

3 Chalmers' two-dimensional argument against physicalism

The doctrine of physicalism is usually put either as an identity or a supervenience[4] thesis according to which everything (mental) supervenes on the physical (all mental states are physical states). Since identity claims are necessarily true (if they are true), it follows that if someone is in pain at t, then there is a true sentence ϕ, drawn from some appropriate physical vocabulary –of so called "completed physics"– such that $\ulcorner \phi \supset$ someone is in pain at $t\urcorner$ is necessarily true.[5]

It seems rather obvious that the identity of physical and mental states is not a truth we discovered by a priori reasoning. In order to prove the aposterioricity of this identity it is usually claimed that zombies are logically possible, i. e. creatures which are physically indiscernible from us, but lack all conscious states.[6] Thus, supposing that someone is in pain at t, many physicalists hold that there is a true physical sentence ϕ such that $\ulcorner \phi \supset$ someone is in pain at $t\urcorner$ is not just necessary, but also a posteriori. Now, David Chalmers argues that this kind of "a posteriori" physicalism is false. His argument takes the semantic framework of "two-dimensionalism" as its starting point.

3.1 Two-dimensionalism

In this section I draw on the presentation in [Byr00, p. 1–8] and [Cha96, p. 131–136]. The matrix-idea is taken from Byrne's paper.

As it is usually understood, the intension of a singular term τ is a function that takes each possible world w to the referent of the extension of τ in w, if it has one. Thus, the intension of 'the chancellor of Germany

[1] *Cf.* [Cha02b] .

[2] To get a complete reconstruction of the use of two-dimensionalism in Chalmers' argument, *cf.* [Byr00] or [BloSta$_1$99].

[3] *Cf.* [Dav$_2$81, p. 263–267].

[4] I will concentrate on identity only. Chalmers takes it that physicalism is best understood as a supervenience thesis, and hence argues against this supervenience. I think that physicalism should be formulated as an identity claim. Arguments in favor of my position can be found in [Per$_1$01]. For the purpose of the present paper the difference doesn't matter.

[5] *Cf.* [Kri$_1$72].

[6] For problems concerning the conceivability of zombies in this sense *cf.* [Per$_1$01,Pol$_1$00,Coh$_1\infty$].

in 2000' is a function that takes the actual world to Gerhard Schröder, and a possible world in which Helmut Kohl is still chancellor to Helmut Kohl. The intension of 'Gerhard Schröder', as Kripke argued, takes every world in which Schröder exists to Schröder. The intension of a sentence σ is a function that takes each possible world w to the truth value that (the proposition expressed by) σ has at w.

When we evaluate some singular term –say 'Gerhard Schröder'– at a world w, we are asking what the referent of 'Gerhard Schröder', used with the actual meaning, is at w, we are *considering w as counterfactual.*[7]

Now consider the familiar example of 'water' and suppose for simplicity that in addition to the actual world $@$, there are exactly three possible worlds: w_1, a twin earth where Putnam's potable fluid XYZ is found in the lakes and oceans, etc.; w_2, a world where the fluid in the lakes and oceans is a mixture of 95% H_2O and 5% XYZ; and w_3, a world where H_2O is found in the lakes and XYZ in the oceans. Then the intension of 'water' (again, if Kripke and Putnam are right) can be displayed in the following *one-dimensional* matrix:

Intension of 'water'

$@$	w_1	w_2	w_3
H_2O	H_2O	H_2O	H_2O

According to Chalmers, a word like 'water' in fact has two intensions associated with it—the *primary intension* and the *secondary intension.* The secondary intension is what we have been simply calling 'the intension'.

The so called primary intension of a singular term is again a function from worlds to extensions, but this time it picks out what the referent of the singular term would be if that world turned out to be actual. Hence we fix the reference of a singular term in the possible world under consideration. In the case of 'water' we can say (in a rough approximation) that the primary intension of 'water' picks out the dominant clear, drinkable liquid in the oceans and lakes, it picks out the *watery stuff* in a world.

To display both intensions, the intension-matrix for 'water' should be extended into two dimensions as follows:

[7] This terminology appears nonsensical to some. Within this framework we are considering $@$ sometimes as counterfactual and sometimes some world $w \neq @$ as actual. I kindly ask the reader to swallow this for the purpose of the present paper.

Intension of 'water'

Counterfactual — ↓ Actual	$@$	w_1	w_2	w_3
$@$	H_2O	H_2O	H_2O	H_2O
w_1	XYZ	XYZ	XYZ	XYZ
w_2	H_2O	H_2O	H_2O	H_2O
w_3	H_2O-or-XYZ	H_2O-or-XYZ	H_2O-or-XYZ	H_2O-or-XYZ

Here the rows represent what the secondary intension of 'water' would be if the row-world turned out to be actual, or were considered as actual. If w_1 turned out to be actual, then 'water' would refer to XYZ in every world. If w_2 turned out to be actual –the world where the clear potable fluid is a 95/5 H_2O/XYZ mix– "we would probably have said that [H_2O] but not [XYZ] was water"[8]—at least that's the way Chalmers thinks about it. In w_3, we probably would have said that both were water, if that world would be the one in which we fixed the reference. (Note that I dropped intra-world variation for simplicity. In a full-fledged two-dimensional framework such variations could be represented by replacing the world considered as actual with so called *centered-worlds*: an ordered pair of a world and at least a marked individual and time.)

The primary intension of 'water' is represented in this matrix by the diagonal from top left to bottom right. It is that function that assigns H_2O to $@$ and w_2, XYZ to w_1, and H_2O-or-XYZ to w_3. So, our original intension is now extended into a function F from $W \times W$ to referents, where W is the space of possible worlds. The secondary intension of 'water' is the Function $F(@, x)$, and the primary intension is the function $F(x, x)$. (If we were to consider intra-world variation in this framework, we would have a space of centered possible worlds W^* and the intension would be a function F from $W \times W^*$ to referents.)

Since the primary intension regards questions about what our terms would refer to if the actual world turned out in different ways it is determined a priori. Simply understanding 'water' already allows one to recover the two-dimensional matrix. Since the secondary intension is not determined a priori, as it depends on how things turn out to be in the actual world, we can recover the matrix a priori, but cannot know a priori what row we are on. This is supposed to hold for all rigid designators.

[8] *Cf.* [Cha96]; compare [Byr00].

(For non-rigid singular terms the secondary intension is a simple copy of the primary intension and so is determined a priori.)

But let's now turn to the intension of sentences, of 'Water is H_2O' in particular. What are the possible-worlds truth-conditions of the proposition expressed by this sentence? Following Kripke and Putnam again, it is the function that assigns The True to every world. More generally, what are the truth-conditions of the proposition that would be expressed by this sentence if world w turned out to be actual? In Chalmers' two-dimensionalism the secondary intension and primary intension are obtained by, respectively interpreting the sentence in accordance with the secondary and primary intensions of its terms. Thus 'Water is H_2O' is supposed to have the same secondary intension as 'H_2O is H_2O' and the same primary intension as 'The watery stuff is H_2O'. Again in a metaphorical way, the matrix can be recovered a priori, whereas we won't know a priori which row we are on.

As we all know, 'Water is H_2O' is one of Kripke's examples of a necessary a posteriori sentence. Chalmers thinks the two-dimensional apparatus can explain why this is so:

> The primary intensions of 'water' and 'H_2O' differ, so that we cannot know a priori that water is H_2O; the associated primary intension of the sentence is not necessary (it holds in those worlds in which the watery stuff has a certain molecular structure). Nevertheless, the secondary intensions coincide, so that 'Water is H_2O' is true in all possible worlds when evaluated according to the secondary intensions – that is, the associated secondary intension of the sentence is necessary. Kripkean a posteriori necessity arises just when the secondary intensions in a statement back a necessary intension, but the primary intensions do not.[9]

Modeling a posterioricity and necessity in this way, Chalmers can extract his crucial premise:

(2D) For any sentence S, S is a priori iff S has a necessary primary intension.

3.2 The argument

Now the argument against physicalism runs as follows:

> Some notation: for all sentences α, $\ulcorner \alpha^p \urcorner$ is a sentence whose secondary intension is the same as α's primary intension. Let Φ be a physical sentence that expresses a proposition true at exactly those worlds that are physical duplicates of the actual world. Let Ψ express some mental fact that is not an a priori consequence of the physical facts. So $\ulcorner \Phi \supset \Psi \urcorner$ is a posteriori and, if physicalism is true, also necessary. According to (2D), then, if physicalism is true the primary intension of $\ulcorner \Phi \supset \Psi \urcorner$ is contingent (the secondary intension is of course necessary).
>
> Now, either the primary intension of Φ is the same as its secondary intension, or not. Suppose first, that they are the same. Then the primary intension of $\ulcorner \Phi \supset \Psi \urcorner$, that is, the secondary intension of $\ulcorner \Phi^p \supset \Psi^p \urcorner$, is the secondary intension of $\ulcorner \Phi \supset \Psi^p \urcorner$. So the secondary intension of $\ulcorner \Phi \supset \Psi^p \urcorner$ is contingent, so the proposition it expresses is contingent, and $\ulcorner \Psi^p \urcorner$ is of course true (the secondary and primary intensions of a sentence coincide at the actual world). But, according to physicalism, for any true sentence a, $\ulcorner \Phi \supset \alpha \urcorner$ is necessary. Hence physicalism is false.
>
> Suppose, on the other hand, that the primary intension of Φ differs from its secondary intension.[...][10]

The last alternative, that the primary intension of Φ differs from its secondary intension, is excluded by Chalmers as a loophole for the physicalist. It could be one, since the contingency of $\ulcorner \Phi^p \supset \Psi^p \urcorner$ does not imply that $\ulcorner \Phi \supset \Psi^p \urcorner$ is not necessary. The cost of taking this horn would be not to know anymore "what the physical really is".[11] This doesn't seem to be a very convenient position, because it would make physicalism an empty claim.

4 No zombies in a non-empty physicalism

Nevertheless, I would recommend every physicalist to bite this bullet. Although there is not enough space available to expand the argument and discuss this move in detail, I shall briefly sketch in what way we indeed do not know what we want to mean with "physical". Consider again the zombie world: we admit to finding a world conceivable which is physically indiscernible from ours, but lacks all mental states. What about the physical changes in our world that are according to physicalism caused by mental states? How are they caused in Chalmers' zombie world which is supposed to be physically indiscernible from ours?

The answer is that conceiving a world "physically indiscernible" from ours *and* at the same time sticking to the *physicalist's* understanding of 'physical' doesn't go with the conceivability of the absence of *qualia*. Arguments that can be used for this strategy can be found in John Perry's *Knowledge, Possibility and Consciousness*. The physicalist might know "what the physical really is", but in order to explain the *aposterioricity* of this knowledge he suppresses parts of this knowledge to the effect that it allows for the possibility of zombies. Note that this suppressed part of

[9] *Cf.* [Cha96].

[10] This neat reconstruction is taken from [Byr00, 9].

[11] *Cf.* [Byr00, 10].

the physicalist's knowledge is not knowledge about inaccessible essential properties.[12] This knowledge is about physical processes which are accessible (like our knowledge of the causal efficacy of mental states), but accessible a posteriori. What this move requires though, is a different partition of primary and secondary intensions than the one Chalmers offers. But I think it would be compatible with (2D) and everything else we said so far. Making this physical knowledge "available information"[13] again will exclude the possibility of a world physically indiscernible from ours but without qualia, and all zombies disappear.

5 Conclusions

I have shown how the tool of two-dimensional modal logic is applied in contemporary metaphysics. The tool seems to have its virtues as an instrument to clarify our arguments in this rather complicated field.[14] At the same time the tool is sufficiently neutral, I tried to show that the dualist's argument built on two-dimensionalism appears to have a loophole for the physicalist. Although *prima facie* it seems too much to swallow that the physicalist doesn't know what he is in fact talking about, this strategy easily looses its problematic character if we remind ourselves how we were dragged into detecting a contingency in the first place. We were already suppressing our knowledge about the physical when we found zombies which are *physically* indiscernible from us conceivable.

[12] This move was considered in [Cha96. p 134–136].

[13] I'm using "available information" in the technical sense introduced by [Bar+97].

[14] For a more critical evaluation of this framework see [Byr00].

FotFS
II

Benedikt **Löwe**, Wolfgang **Malzkorn**, Thoralf **Räsch** (*eds.*)
Foundations of the Formal Sciences II
Applications of Mathematical Logic in Philosophy and Linguistics
Papers of the conference held in Bonn, November 10-13, 2000

Topics in Reverse Mathematics

Mariagnese Giusto
via Loreto Vecchia 9/10/A
17100 Savona
Italy

E-mail: `mariagnese@savonaonline.it, mariagnese.giusto@tiscalinet.it`

Abstract. This paper suggests motivations and goals of the program known as Reverse Mathematics, providing some illustrative examples of the many techniques and problems involved in working within subsystems of second order arithmetic, namely, in particular, RCA_0, WKL_0, ACA_0, ATR_0. Some examples from Combinatorics, – the Free Set Theorem and Ramsey's Theorem – show how some theorems of ordinary mathematics may not fit in one of the subsystems mentioned aboves. As application of Reverse Mathematics to the History of Mathematics, we comment on König's duality theorem and Cantor's proof that every countable closed set is a set of uniqueness. Also, more technically, we present some results related to Lebesgue spaces (every open covering has a Lebesgue number) and Atsuji spaces (every continuous function defined on them is uniformly continuous): we show that the known proof of "every Atsuji space is Lebesgue" needs ACA_0, and we conjecture that the statement is actually equivalent to ACA_0. Finally, we discuss some limitations for Reverse Mathematics that may lead to projects of research in this field of mathematics.

1 Introduction

In this paper we discuss some topics in Reverse Mathematics, the program started by Harvey Friedman and Stephen Simpson in the '70's and developed in many publications: the basic reference is Simpson's recent

Received: March 16th, 2001;
Accepted by the editors: April 30th, 2001.
2000 *Mathematics Subject Classification*. **03B30 03F35** 05-99 46A99.

© *2003 Kluwer Academic Publishers. Printed in The Netherlands, pp. 63–87.*