ANALYTICITY, MODALITY, AND GENERAL TERMS

Kathrin Glüer
Department of Philosophy
Stockholm University
kathrin.gluer@philosophy.su.se

Peter Pagin
Department of Philosophy
Stockholm University
peter.pagin@philosophy.su.se

ABSTRACT: In his recent paper 'Analyticity: An Unfinished Business in Possible-World Semantics' (Rabinowicz 2006), Wlodek Rabinowicz takes on the task of providing a satisfactory definition of analyticity in the framework of possible-worlds semantics. As usual, what Wlodek proposes is technically well-motivated and very elegant. Moreover, his proposal does deliver an interesting analytic/synthetic distinction when applied to sentences with natural kind terms. However, the longer we thought and talked about it, the more questions we had, questions of both philosophical and technical nature. Hence the idea of this little paper – for how better to honor a philosopher than by trying very hard to criticize him? After quickly running over some background in possible worlds semantics and setting out Wlodek's proposal against that background, we shall bring up and discuss our questions in sections 3 – 5. In the final section, we shall also make a stab at a different solution to the problem, making use of our own earlier idea of relational modality.

Printed from: *Hommage à Wlodek. Philosophical Papers Dedicated to Wlodek Rabinowicz.* Ed. T. Rønnow-Rasmussen, B. Petersson, J. Josefsson & D. Egonsson, 2007. www.fil.lu.se/hommageawlodek

1. Background

Both in what we might call model theory proper, and in (model theoretic) semantics for natural language, we approximate the notion of meaning with that of an *interpretation*. An interpretation is a function that assigns semantic values to expressions, where semantic values are the things that serve as meanings in the model or theory in question. In formal semantics for natural language, the interpretation function assigns values to natural language expressions, or regimented counterparts of them, that stand in for real meanings, whatever those are. Model theory proper, on the other hand, usually deals with formal languages, and studies the semantics for those languages in terms of models. Here, one is usually interested not in particular models, but in generalizations over models of particular kinds.

In possible-worlds semantics, a model S for a language L, be it formal or natural, is typically a structure like

$$\langle W, \boldsymbol{a}, R, I, D, i \rangle,$$

where W is a set of possible worlds, $\boldsymbol{a}$ a designated member of W (the actual world of S), R a binary relation on W (an accessibility relation), I the domain of individuals of S, D a function from members of W to subsets of I (the individuals existing in that world), and i an interpretation function.

This is complemented with a truth-definition, giving the conditions for an arbitrary formula to be true at a world w, under an assignment f of values (in $D(w)$) to variables, for instance as follows:

(TS) i) T($At_1, \ldots, t_n, w, i, f$) iff $\langle i_f(t_1, w), \ldots, i_f(t_n, w) \rangle \in i(A, w)$

ii) T$(A\&B, w, i, f)$ iff T(A, w, i, f) and T(B, w, i, f)

iii) T$(\neg A, w, i, f)$ iff not T(A, w, i, f)

iv) T$(\forall xA, w, f)$ iff for every $f' =_{/x} f$, T(A, w, f')

v) T$(\Box A, w, i, f)$ iff for every $w'Rw$ T(A, w', i, f)

vi) T$(\mathcal{A}(A), w, i, f)$ iff T$(A, \boldsymbol{a}, i, f)$.

Here $i_f(t_j, w)$ equals $i(t_j, w)$ (the referent of t_j in w) in case t_j is an individual constant, and equals $f(t_j, w)$ in case t is a variable. 'every $f' =_{/x} f$)' means every function f' that differs from f at most in what it assigns to x. 'every $w'Rw$' means every world w' that stands in the relation R to w. That A is true at a world w is equivalent to saying that $w \in i(A)$, where $i(A)$ is the set of worlds that make up the intension of A in the model. Truth simpliciter in the model, T$(A, \mathcal{M})$ (where $\mathcal{M}$ is the model) equals truth at the actual world of the model.

Since it does not play any role for what follows, we have ignored the complication that arises if D is not a constant function, i.e. if there are different individuals in different worlds. To allow for that one would have to adjust the truth definition accordingly, for instance by having separate truth- and falsity-conditions in the negation clause (TS ii). Like Wlodek in his paper, we shall abstract away from the accessibility relation (i.e. assume that every world is accessible from every world). Wlodek also ignores the distinction between the interpretation function and the variable assignment, and as that does not play a role for our discussion either, we shall follow him here, too.

In Wlodek's simplified semantic framework, then, a model is a structure

$$\langle W, \boldsymbol{a}, I, D, i \rangle.$$

The structure comprising the first four elements,

$$\langle W, \boldsymbol{a}, I, D \rangle,$$

is called a *frame*, and we can then characterize a model as a pair $\langle S, i \rangle$ of a frame S and an interpretation function i that is defined on S. As usual, this is completed with a truth definition.[1]

2. Wlodek's proposal

The notion of analyticity has been defined in a number of different ways since Kant. The version that Wlodek takes as his point of departure is canonically formulated by Quine:

(AY) A statement is analytic when it is true by virtue of meanings and independently of matters of fact

(Quine 1951, 20). The intuitive idea is that synthetic sentences depend for their truth or falsity both on what they mean and on what the world is like (in other respects), while the truth values of analytic sentences only depend on the meaning. The world might be any possible way, and the sentence is nevertheless true. However, although this is an intuitively necessary condition

[1] The revised quantifier clause of that truth definition is equivalent to

(TS′) iv) T($\forall x A, w, i$) iff for every $i' =_{/x} i$, T(A, w, i').

of being true in virtue of meaning, it cannot be sufficient. Otherwise any sentence that is necessarily true, i.e. true at all worlds, would count as true in virtue of meaning. Intuitively, we want to allow for sentences to be necessarily true for reasons other than meaning, metaphysical reasons for instance.

To interpret (AY) within possible-world semantics, one should therefore appeal to the concept of an interpretation instead. Wlodek starts by considering the principle

(DefAn) A sentence is *analytic* on a given interpretation iff it is true in all models that keep the interpretation constant.

The quantifier 'all models' has as its domain the collection of models of some given semantic theory, and (DefAn) is therefore relative to semantic theory. But, as Wlodek immediately remarks, this does not make much sense in the given framwork:

> To keep the [interpretation] constant, we need in the first place keep constant the range of the [interpretation], which means we need to keep constant the frame itself, from which we draw the entities assigned to linguistic expressions. But then there is nothing left in the model that can be varied! (Rabinowicz 2006, 351.)

On (DefAn), analyticity on an interpretation is, in a sense, reduced to truth on that interpretation, obviously contrary to intention.

Wlodek's elegant solution to this problem is the introduction of higher-level interpretation functions $\boldsymbol{i}$. Let's call them 'w-interpretations', or, following Wlodek, simply 'interpretations', if the context disambiguates. A w-interpretation assigns to an expression, not an intension, as ordinary interpretations, but *ultraintensions*. An ultraintension is a function that takes a *frame* as argument and gives an ordinary intension as value.

Thus, Wlodek characterizes a w-interpretation $\boldsymbol{i}$, defined for a language L and a collection Σ of frames so that for every frame $S = \langle W, \boldsymbol{a}, I, D\rangle$ in Σ,

(WI) i) if t is a variable or an individual constant of L, $\boldsymbol{i}(t)(S)$ is an object in I

ii) if F^n is an n-ary predicate of L, $\boldsymbol{i}(F^n)(S)$ is a function that to each world w assigns an n-ary relation on I, and

iii) if A is a sentence of L, $\boldsymbol{i}(A)(S)$ is a subset of W.

By taking $\boldsymbol{i}(A)(S)$ and abstracting on the formula argument we get a function $\lambda A(\boldsymbol{i}(A)(S))$ from formulas to intensions, defined on the frame S, i.e. an ordinary interpretation function i on S. Therefore, $\langle S, \lambda A(\boldsymbol{i}(A)(S))\rangle$ is a model. And hence, each w-interpretation $\boldsymbol{i}$ determines a set of models

$$\mathcal{M}_{\boldsymbol{i}} = \{\langle S, i\rangle : \forall A(i(A) = \boldsymbol{i}(A)(S))\}.$$

With the help if w-interpretations Wlodek then defines his notion of analyticity:

(WA) A sentence A is *analytic* on an interpretation $\boldsymbol{i}$ iff for all $S \in \Sigma$, A is true in $\langle S, \boldsymbol{i}\rangle$, i.e. iff for all $S = \langle W, \boldsymbol{a}, I, D\rangle$ in Σ, $\boldsymbol{a} \in \boldsymbol{i}(A)(S)$.

This notion we shall call ‘w-analyticity’. Strictly speaking, w-analyticity is relative to w-interpretation function. So we really have a collection of concepts, w($\boldsymbol{i}_1$)-analyticity, w($\boldsymbol{i}_2$)-analyticity, and so on. By ‘w-analyticity’ we shall intend the non-relative generic property of having one or other of the specific analyticity-properties.

Wlodek goes on to compare the notion of w-analyticity with other notions, such as logical truth, necessity and *a priority*. Logical truth comes out as w-analytic, as it should, since logical truth is truth in the actual world of all models, and analyticity on $\boldsymbol{i}$ is truth in the actual world in all models in $\mathcal{M}_{\boldsymbol{i}}$.[2]

Of special interest is Wlodek's definition of *synonymy* as identity of ultraintension. As Wlodek points out (Rabinowicz 2006, 354), it follows that predicates F and G are synonymous (on $\boldsymbol{i}$) iff

(2) $\forall x(Fx \leftrightarrow Gx)$

is analytic (on $\boldsymbol{i}$). We shall return to this below.

3. Quine's problem

Quine himself, as we all know, no sooner formulates the analyticity principle (AY) than he rejects it. His immediate reason is that the formulation refers to meanings, which Quine regards as suspect. In a later paper (Quine 1960), however, Quine develops his criticism of the idea of truth in virtue of meaning

[2] This depends on defining logical truth as truth in the actual world of all models, a definition required for making

(1) $\mathcal{A}(A) \rightarrow A$

a logical truth. (1) is true in the actual world of every model, but false in a world w if A is false in w and true in $\boldsymbol{a}$.

alone. The problem is that we have no basis for applying the concept in the crucial cases, for we don't know how to tell whether a sentence is true in virtue of meaning alone or in virtue of very general facts. Quine's example is

(3) $\forall x(x = x)$.

In this case we can say either that (3) is analytic, because true in virtue of meaning alone, or alternatively that it is synthetic but still true because of the general fact that every object is self-identical (Quine 1960, 113). This fact (if it is a fact) obtains in every possible world. And this generalizes; for every truth that is even a candidate for analyticity, there would seem to be a corresponding general fact.

Now, Quine's problem is not a problem that can be solved by means of semantics. We don't get any help with it by looking at a semantic theory. Take (3) again. We could add a clause for identity to our truth definition get the following as result:

(4) T($\forall x(x = x), \boldsymbol{i}, S, w$) iff for every $\boldsymbol{i}' =_{/x} \boldsymbol{i}$, $\boldsymbol{i}'(x, S, w) = \boldsymbol{i}'(x, S, w)$.

Does (4) tell us that (3) is true independently of facts? Or that it is true in virtue of the fact that

(5) for every $\boldsymbol{i}' =_{/x} \boldsymbol{i}$, $\boldsymbol{i}'(x, S, w) = \boldsymbol{i}'(x, S, w)$?

Formally, it tells us the latter, if anything. And we will obviously get the same result whatever formula evaluation we look at.

Of course, we might say that (5) *itself* is true in virtue of meaning, and that therefore it does not really express a state of affairs that obtains as a fact. But if this, in turn, is to come out in a semantic theory, we would need a new meta-meta-language to state it in, and in that meta-meta-language the result would be analogous. It would show us that, formally, the truth of (5) depends on some general fact specified in that meta-meta-language.

Hence, we don't really get any help with solving Quine's problem from Wlodek's proposal. But neither should we expect to get such help from it, or any other proposal like it.[3] In a sense, this is not even a criticism, since we think that Quine's problem does not have a solution. Thus, we should not really ask whether Wlodek's proposal successfully gives a version of (AY) in possible-world semantics. Rather, we should ask whether it captures (an important or interesting aspect of) our intuitive analyticity judgments, quite independently of (AY).

Intuitively, there is a distinction, and clearly an important and interesting distinction, between necessary truths such as logical truths (and analyticities) and other necessities when it comes to their dependence on general metaphysical fact. The distinction we have in mind is not an on/off distinction, but can be characterized in terms of *degrees of dependence* on metaphysical facts: Logical truths only minimally depend on general metaphysical facts, but other necessities do so to a much higher degree. One intuitive desideratum on a notion of analyticity would thus be that it accommodates this distinction. That is, amongst the necessities, it should distinguish the minimally metaphysical ones, among them the logical truths, from the (more) metaphysical necessities.[4]

---

[3] Including, of course, our own suggestion presented in the final section.

[4] This 'metaphysical desideratum' intuitively holds regardless of whether all analyticities are necessary or not. Contrary to Kripke himself, Wlodek reckons with contingent

Before we can turn to the question whether Wlodek's notion of w-analyticity captures this particular intuitive distinction, we need to get clear about a more general matter, however. For capturing, or failing to capture, any intuitive informal notion whatsoever means doing semantics for natural language. And it seems to us that it is not immediately clear that Wlodek is doing natural language semantics. So, the first thing we would like to know is what he is really after: a project in model theory proper or in natural language semantics? As we shall see, this choice has important consequences. Actually, we shall argue that for Wlodek, there is something of a dilemma here. A purely model theoretic notion of analyticity not only does not provide any help in answering the metaphysical desideratum – it threatens to be empty. If the project is natural language semantics, however, it seems to us that Wlodek's proposal turns out to be a version of two-dimensional semantics – an outcome that Wlodek explicitly wants to avoid.

4. The problem of odd interpretations

In possible-worlds semantics (for natural language), an (ordinary) intension is the counterpart to the pre-theoretic notion of meaning. This comes out in ordinary counterfactual considerations. We can take a sentence like

---

w-analyticities. His example is Kripke's sentence ‘‘The standard metre in Paris is one meter long’’ (Rabinowicz 2006, 353). For Wlodek, that is, w-analyticity and minimally metaphysical necessity are not coextensive. But even if we follow him here, it intuitively holds that, amongst the necessities, it is the minimally metaphysical ones that are analytic.

(6) Alfred likes salmon

and consider whether it is true under various different scenarios.[5] In relation to a scenario described, we check whether (6) comes out true by trying to determine whether Alfred in that scenario stands in the liking-relation to what is salmon in that scenario.

Now, strictly speaking, we should say that, with respect to a scenario w, we check whether what ‘Alfred’ denotes in w and what ‘salmon’ denotes in w stands in the relation denoted by ‘likes’ in w. However, there is a reason why we don't bother. For, as we ordinarily understand it, ‘salmon’ stands for a *projectable property* and ‘like’ for a *projectable relation*. When we want to determine the extension of ‘salmon’ in some world w, as given by a description, we check which individuals have the *property* of being a salmon, because that property, as we ordinarily conceive it, can be projected into counterfactual scenarios. Likewise, we take it that we can project the relation of *liking* into counterfactual scenarios. For the proper name ‘Alfred’, what counts is the just that we are still speaking of the same individual Alfred in the counterfactual scenario. In short, we tacitly take the possible-world intension of an ordinarily meaningful expression to *track* the projection of the associated property into counterfactual possibilities. That is why a particular, well-chosen intension can plausibly be taken to model our intuitive understanding of an expression, and why a particular, well-chosen possible-world interpretation function can model our intuitive understanding of a language.

---

[5] Let's say that describing a scenario is giving a partial description of a possible world. Then a scenario can be regarded a set of possible worlds with common characteristics.

But things change when what we are doing is not semantics for natural language, but 'pure' model theory. For, of course, not all interpretation functions of a model theory are like that. Provided all the relevant objects are in the domain of individuals, an interpretation function i_1 can assign to the predicate F in a world w_1 an extension comprised of Wlodek, Julia Roberts, the state of Texas, and The Moonlight Sonata, while in world w_2, it assigns an extension consisting solely of The Queen Elizabeth II (the ship), and in world w_3 Kanzi the bonobo and that worlds largest contiguous quantity of sulfuric acid. From the point of view of natural language semantics, however, we cannot really regard the intension assigned to F by i_1 as a *meaning* in any ordinary sense of the word, because in this collection of extensions we cannot see the outcome of any projectable property.

Now consider the status of the atomic sentence Ft under i_1, given that at w_1, i_1 assigns to t Julia Roberts, at w_2 The Queen Elizabeth II and at w_3 Kanzi. Ft comes out true at all three worlds. Suppose it goes on like that. Ft will then be necessarily true, because true at all worlds. Intuitively, this is just gerrymandering. With i_1, we have not managed to assign to Ft any *proposition* – in an intuitive sense of the word – that holds by necessity. But formally, of course, the intension of Ft under i_1 is the set of all worlds, well known as the one and only necessary proposition.

From a logical point of view, there is nothing objectionable about interpretation functions like i_1. On the contrary. In model theory we are usually interested in properties that are invariant across models or interpretations: what is true in every model, or what is true in every model given that something else is true in that model. We are characterizing 'purely' logical properties, and in so doing we quantify over the domain of models. If we were to restrict the collection of interpretation functions to those that are ''sensible'' or ''natural'' in some sense, the domain of models would be smaller, and any results about logical properties weaker and less interesting. So we do want the odd interpretations in the domain for the sake of the logical theorems.

Things change, however, if we want to characterize, in model theoretic terms, properties that *do* depend on the choice of interpretation. And this is precisely what Wlodek does with his notion of w-analyticity. W-analyticity, remember, is relative to a w-interpretation function. Although w-interpretations belong to a higher level than ordinary interpretations, any single w-interpretation $\boldsymbol{i}$ can be as odd as i_1 (but at a higher level). In particular, we can have a w-interpretation $\boldsymbol{i}_2$ such that, for some particular frame S, the ordinary interpretation function $\lambda A(\boldsymbol{i}_2(A)(S))$ is exactly i_1, and so on.[6]

Now, given the odd w-interpretation $\boldsymbol{i}_2$, consider the associated notion of w($\boldsymbol{i}_2$)-analyticity. If for any frame S, $\lambda A(\boldsymbol{i}_2(A)(S))$ is odd in a way equivalent to i_1, we can have the result that Ft comes out as w($\boldsymbol{i}_2$)-analytic.[7] Although again this is logically unobjectionable, w($\boldsymbol{i}_2$)-analyticity clearly does not cor-

---

[6] Of course, i_1 was explicitly designed with t as a non-rigid term, thus giving it as its intension a non-constant function. In Wlodek's framework, individual terms are just assigned objects, which is equivalent to just assigning them constant intensions. So that particular example does not fit Wlodek's framework. For an example that does fit the framework, consider two gerrymandered predicates F and G which at each world are assigned disjoint and otherwise unrelated extensions. Then $\forall x(Fx \rightarrow \neg Gx)$ comes out as necessarily true.

[7] With ordinary clauses, this does not work at worlds with empty domains. So either we need an example without individual constant, or else a special treatment of non-referring terms.

respond to any intuitive notion of analyticity. Intuitively, we want an analytic sentence to be true at all worlds because the terms in it stand for properties such that combined according to the syntax of the sentence they conspire to yield a proposition that holds everywhere. Gerrymandered interpretations do not deliver that.

The conclusion, then, is that to the extent that Wlodek with his definition of analyticity attempts to give a formal counterpart or approximation of some intuitive informal notion, whether he succeeds or not depends on the choice of w-interpretation. A concept like w($\boldsymbol{i}_2$)-analyticity does not fulfill the requirement, and because of the existence of odd concepts like w($\boldsymbol{i}_2$)-analyticity, neither does the generic concept of w-analyticity itself (i.e. the concept of being w($\boldsymbol{i}$)-analytic for some $\boldsymbol{i}$ or other). In fact, if Ft is w($\boldsymbol{i}$)-analytic for some $\boldsymbol{i}$, then it is safe to say that any formula is w($\boldsymbol{i}$)-analytic for some $\boldsymbol{i}$ or other, and hence the generic concept of w-analyticity is empty: every formula is w-analytic, if there are enough w-interpretations in the framework.[8]

In the next section, we shall therefore consider Wlodek's concept of analyticity under a *good* choice of w-interpretation function.

[8] It does not help to consider only abstract model frameworks, with domains of individuals just assumed rather than picked from the real world. In such a framework, no distinction can be made at all between natural and gerrymandered interpretation functions, for if a formal counterpart of an intuitive notion of analyticity requires a natural interpretation, there must be something in virtue of which the proposed interpretation function is natural rather than gerrymandered, and no such thing is offered by abstract frameworks.

5. The right interpretation

Instead of an arbitrary interpretation, we shall now consider what might be called *the right interpretation*, $\boldsymbol{i}_r$. $\boldsymbol{i}_r$ is intended to track the properties that are intuitively associated with the lexical items of English. In this section, unless otherwise noted, we mean w($\boldsymbol{i}_r$)-analyticity by 'w-analyticity'. A formula that is w-analytic, then, is a formula that is true in the actual world of every frame S under $\boldsymbol{i}_r$. We shall also designate one particular frame S_a as the *actual frame*. The actual/designated world of S_a is the ''real'' actual world. By fixing the w-interpretation in this way, the overall project is changed from doing model-theory proper to doing formal natural language semantics by means of model-theoretic techniques. As we will see, this is not merely a change of labels.

Two questions now naturally arise:

1. What happens in other frames?
2. How is the distinction between minimally metaphysical and other necessary truths effected?

As regards the first question, note that there is, or at least may well be, a w-interpretation $\boldsymbol{i}$ in the framework that gives the natural ordinary interpretation on S_a but provides nonsensical interpretations on other frames. But we want $\boldsymbol{i}_r$ to provide natural interpretations also on other frames. So what *is* natural on other frames? It turns out that Wlodek's answer to this question pretty much coincides with the answer to question 2.

The idea is that the actual world plays a decisive role for determining the semantic value of certain linguistic expressions. Wlodek writes

> While a sentence such as ''Water is H_2O'' is necessary in a model

> in which the term ‘‘water’’ picks out H_2O in the designated world, this sentence will be false (and necessarily so) in a model in which the designated world is like Putnam's Twin Earth, with a different chemistry from ours but with the same phenomenal appearance. For that choice of the designated world, the term ‘‘water’’ picks out a different chemical substance. The reason for this difference is that the analytical application criteria for ‘water’ rely on phenomenal characteristics that actually are exhibited by its referent. On Twin Earth, these criteria pick out, say, XYZ rather than H_2O. [...] Expressions such as ‘‘water’’ and ‘‘H_2O’’ are necessarily co-extensive, which means that they have the same intension. However, they are not synonymous (Rabinowicz 2006, 354).

What emerges from this is the following account: The linguistic meaning of an expression e is its ultraintension. The particular possible world w in which e is *used* determines an ordinary intension for e. If the ultraintension is a constant function, the ordinary intension will be the same, regardless of the world of use. In some cases, however, the ultraintension is not constant; Wlodek considers ordinary language natural kind terms as cases in point. According to him, the ultraintension of ‘water’ is associated with a set of phenomenal characteristics P_w. Consequently, when used in a world w, ‘water’ will acquire as its ordinary intension a function $i(\text{‘}water\text{’})$ that tracks the *substance property* that *in w* exhibits P_w. $i'(\text{‘}water\text{’})$ will be the intension of ‘water’ in a frame S' that has w as its actual (designated) world. Finally, the *right* ultraintension $\boldsymbol{i}_r$ is such that $\lambda A(\boldsymbol{i}_r(A)(S'))$ is exactly i'.

This answers both questions above. On this account, the very sentence

(7) Water is H_2O

exemplifies the distinction wanted. (7) is a necessary truth in the standard sense, true at all worlds in the actual frame. However, it is not analytic, since it is not true at all actual worlds of non-actual frames under $\boldsymbol{i}_r$. We seem to have what we wanted.

There is more to the picture, however. Presented as above, Wlodek's concept of ultraintension is analogous to David Kaplan's concept of *character* (Kaplan 1989). The character of a sentence A is a function from a context of utterance c of A to the proposition expressed by A (or by the utterance) in c. The ultraintension of A is a function from a frame S to the proposition expressed by A in S. And in fact, since for Kaplan (Kaplan 1989, 507-10) the possible world of the utterance is an ingredient in the context, there might be more than an analogy.

The question is what more can vary between frames than the selection of their actual world. Wlodek's answer seems to be that, in principle, all the other elements can vary, but when applying the idea to show that necessity does not entail w-analyticity, he emphasizes the following:

> In particular, as we vary the frame, the range of possible worlds may expand. If the designated world in the new model represents such a new possibility, then A might be false in that world, which would make it false in the new model (Rabinowicz 2006, 354).

In our terms, that would amount to there being *other* possible worlds than those in the domain W of the actual frame S_a. And the question is whether this option has to be taken seriously here. Again, the answer depends on what the overall project is. In an abstract model theory, designed to let everything vary except the clauses of the truth definition, we indeed need to vary the domain of possible worlds to achieve maximum generality. But since the project *is* natural language semantics, taking this option seriously, it seems to us, is not only not necessary, but actually misguided. In natural language semantics, we work with a fixed interpretation. And if we want to take account of other possible worlds for the sake of handling intensional vocabulary, we want all the relevant possibilities to be within the domain of the semantic theory used: The domain of possible worlds is simply the domain of *all* possible worlds. Saying that there are other possible worlds in other frames is like saying that there are more possible worlds than there actually are.

So, in natural language semantics, we don't want the domain of worlds to vary between frames. Similarly, we don't want the domain of individuals I to vary, since it is equally incoherent to say that there are more possible individuals than there are in the actual domain of possible worlds. And we don't want the domain function D to vary, either. If different distributions of individuals create different possible worlds, all these different worlds can, and should, be in the same domain of worlds. In other words, since we don't want the domain of worlds to vary between frames, we don't want the domain function to vary, either.

But if we keep both W and I and D fixed, the only thing left to be varied is the designated world of the frame. But then, if an ultraintension $\boldsymbol{i}$ of an expression e takes a frame as its first argument, this is effectively equivalent to taking a *possible world* as its first argument, viz. the designated possible world of the frame in question. And since that world is the world in which the expression is used, i.e. the world of utterance, it seems that ultraintension not only is analogous to character, but rather a restricted version of it.

Moreover, if the project is natural language semantics, Wlodek's system turns out to be a version of *two-dimensional* semantics. This is not what Wlodek wants; officially at least, he wants to avoid going two-dimensional. Before he develops his own system, he says the following about the two-dimensionalist alternative:

> But even if this suggestion is workable, we shall here abstain from going two-dimensional. As we shall see, a concept of a priori that is essentially equivalent to the one sketched above can be defined without leaving the one-dimensional framework (Rabinowicz 2006, 350).

Now, Wlodek's original truth definition is indeed one-dimensional, but when he introduces ultraintensions later on (2006, 352), he indicates with the clause for atomic formulas that the ultraintension of an expression, with the frame argument, does occur in the revised truth definition. That means that Wlodek's ultraintension effectively is a semantic property that gets evaluated at a *pair* of possible worlds, the world of utterance and the proper world of evaluation. And that is precisely the distinctive feature of two-dimensionalism.[9] Accordingly, the two-dimensionalist notions of *primary* and *secondary* intension (cf. e.g. Chalmers 1996) can be reconstructed in terms of ultraintensions. The primary intension of an expression e is precisely its ultraintension determined and evaluated at the world of utterance. Formally, with $\boldsymbol{i}(e)$ as a binary function from

[9] And in fact, even if we would let other elements of the frames vary, the system would still be essentially two-dimensionalist, since the world of utterance would then be *part* of the first argument of the ultraintension of the expression.

pairs of worlds $\langle \boldsymbol{w}_i, \boldsymbol{w}_j \rangle$ to extensions, and with $\boldsymbol{i}^1(e)$ as the primary intension of expression e and $\boldsymbol{i}^2_{\boldsymbol{w}}(e)$ as the secondary intension of e at world $\boldsymbol{w}$, we have

(8) a) $\boldsymbol{i}^1(e) = \lambda \boldsymbol{w}((\boldsymbol{i}(e))(\boldsymbol{w}, \boldsymbol{w}))$

b) $\boldsymbol{i}^2_{\boldsymbol{w}_i}(e) = \lambda \boldsymbol{w}((\boldsymbol{i}(e))(\boldsymbol{w}_i, \boldsymbol{w}))$.

One could, of course, save the one-dimensional character of Wlodek's system by leaving the ultraintensions out of the truth definition. Then, ultraintensions would only be used for characterizing meta-linguistic properties like analyticity, not for giving truth conditions of object language sentences. This would amount to removing ultraintensions from natural language semantics, and employing them only for certain philosophical purposes. But this is not what Wlodek wants, since he thinks that ultraintensions are relevant for giving the semantics for ‘ultraintensional’ (or hyperintensional) contexts such as ‘S believes that …’ (2006, 354). We conclude, therefore, that Wlodek's system really is two-dimensional, either literally or essentially.

To sum up what we have argued so far: The most basic question to be raised about Wlodek on analyticity concerns the nature of his overall project. Is it a project in model theory proper or in natural language semantics? If it is model theory proper, his notion of w-analyticity threatens to remain empty. Moreover, a purely model theoretic notion of analyticity does not provide us with a formal counterpart for our informal notion of analyticity and, thus, does not account for the intuitive distinction between minimally metaphysical truths and other necessities. If the project is natural language semantics, on the other hand, Wlodek does provide us with such an account. At the same time, however, his system, contra intentionem, turns out to be a version of two-dimensionalism.

We would like to conclude by sketching an alternative: a truly one-dimensional account of the distinction between minimally metaphysical truths and other necessities based on our earlier idea of relational modality.

6. Relational modality and general terms

In 'Proper names and relational modality' (forthcoming) and 'Relational modality' (2006) (the former more philosophical, the latter more technical), we suggest a new semantic account of the modal intuitions that Kripke made use of in *Naming and Necessity* (1980) as evidence for the rigidity of proper names. To recapitulate briefly, Kripke's idea was that a name like 'Aristotle' could not have the same intension as the description 'the teacher of Alexander', since the sentence

(9) Aristotle might not have gone into pedagogy

is simply true. If 'Aristotle' were co-intensional with 'the teacher of Alexander', the two expressions would be intersubstitutable in modal contexts, and hence

(10) The teacher of Alexander might not have gone into pedagogy

would be true as well. But (10) is ambiguous between a wide-scope reading

(10′) The teacher of Alexander is such that: possibly he did not go into pedagogy

which is true, and a narrow-scope reading

(10″) It is possibly that case that: the teacher of Alexander he did not go into which is false, while (9) does not have any such ambiguity. Hence the two expressions are not intersubstitutable and therefore not co-intensional. Assuming that synonymy coincides with or at least implies co-intensionality, we can conclude that they are not synonymous either. Generalizing from this argument, Kripke concluded that proper names are not synonymous with non-rigid definite descriptions, definite descriptions, that is, that pick out referents in possible worlds according to some contingent descriptive condition. Rather, proper names are rigid designators: they denote the same individual in every possible world (where it exists).

Our suggestion is that the intuitive difference between (9) and (10) depends on the *de re* nature of our use of our ordinary modal expressions. When asking, for instance, what might have been true of Aristotle, we are interested in the person referred to, in *Aristotle*, no matter how he is designated. And we want to know what would be true of *this very person* in counterfactual circumstances. No-one has brought this out more clearly than Kripke himself, but he does not use the semantics of the modal expressions to explain it. We suggest to do so. We think that the de re nature of ordinary modal thinking comes out in the fact that proper names (and other simple singular terms) occur referentially *when in the scope of modal expressions*. That is, in the scope of modal expressions, proper names contribute to truth conditions by referring to their actual world referents.[10]

For simple cases, this idea is implemented in the semantic theory by means of two truth definition clauses. First, there is a clause in the definition of what

[10] Actually, it is a little more complex than this.

we call ‘actua-truth’ concerning the evaluation of atomic formulas:

(A) Actua-true($Pt_1, \ldots, t_n, \boldsymbol{w}$) iff $\langle i(t_1, \boldsymbol{a}), \ldots, i(t_n, \boldsymbol{a})\rangle \in i(P, \boldsymbol{w})$.

That is, an atomic formula is actua-true at a world w just in case the referents of the terms in the *actual* world satisfy the extension of the predicate in w. The second relevant part of the truth definition is the clause that effects that (A) kicks in under the necessity operator:

(M) True(⌜ It is necessary that ϕ⌝, $\boldsymbol{w}$) iff Actua-true($\phi, \boldsymbol{w}'$) at any world $\boldsymbol{w}'$ accessible from $\boldsymbol{w}$.

From (A) and (M) we can see that even if ‘Aristotle’ is co-intensional with ‘the teacher of Alexander’, (9) comes out true. For according to the proposal, the truth of (9) depends on whether there is a world at which it is true that the actual referent of ‘Aristotle’, i.e. Aristotle, does not go into pedagogy. Since there is, (9) is true.

What we would like to do here is to sketch how this idea could be extended to general terms and how this extension could be used for drawing a one-dimensional distinction between minimally metaphysical truths and other necessities. For the sake of argument, we shall here agree with Kripke (and Wlodek) that certain sentences involving natural kind terms do express metaphysical necessities of the non-analytic kind. An example, again, would be

(7) Water is H_2O.

Now, as far as natural kind terms can be treated as *terms*, our suggestion in fact is rather simple: The idea is that natural kind terms embed like proper names under modal operators.[11] Then we can apply clause (A) again, and take it to hold in case any of the terms t_i is either a simple singular term or a natural kind term.[12]

Assume, then, that both 'water' and 'H_2O' are natural kind terms, and that they are co-referential (in the actual world), i.e. that they refer to the same *kind*. But they are *not* co-intensional, since they intuitively differ in meaning. By the present proposal, the sentence

(11) Necessarily, water is H_2O

still comes out true, since by assumption the two terms are co-referential in the actual world. This is in accordance with standard intuitions. However, on this account

(7) Water is H_2O.

is *not* necessarily true: (7) will be false at worlds where the terms have different extensions. Nevertheless, its *necessitation*, (11), is true.

---

[11] Similarly, Kripke suggested that natural kind terms, like proper names, are rigid designators.

[12] It does not at present matter so much how natural kind terms are identified. We can go either the externalist way and say that a natural kind term is a term that denotes a natural kind, or go the internalist way and say that a natural kind term is a term that is treated as such by the speaker(s). Since we treat a term as a natural kind term just if we believe it denotes a natural kind, these two alternative will yield the same intuitions about the current status of terms.

The framework of relational modality thus allows us to draw a distinction between sentences true at every world and sentences whose necessitations are true. And we further suggest that this distinction in fact makes *some* distinction between minimal or analytic and non-analytic necessity: The analytic ones are the sentences true at every world, while the non-analytic ones are the sentences whose necessitations are true.[13]

We would like to conclude by also sketching how the idea of relational modality could be applied to natural kind 'terms' where, or to the extent that, these have to be regarded as *predicates*.[14] For singular terms, we operated with a distinction between simple and complex singular terms, suggesting that these behave differently in the scope of the modal operators. An analogous distinction could be made between natural kind predicates and others, most notably those explicitly expressing the descriptive stereotype for a natural kind. However, the idea obviously cannot be that, in the scope of modal operators, natural kind

---

[13] Conversely, one can suggest that a member of the stereotype for 'water' (cf. Putnam 1975a),

(12) Water is thirst-quenching

is true at every possible world, while

(13) Necessarily, water is thirst-quenching

is false, because there are worlds where H_2O is not thirst-quenching. Hence, the proposal also allows for contingent analytic sentences (or, for those who like it, contingent *a priori* sentences).

[14] Cf. Soames 2002, esp. chapters 10 and 11 for an extensive discussion of this issue.

predicates take the *extensions* they have in the actual world.[15] Rather, the difference has to hinge upon the *properties* associated with the predicates.

Our suggestion is that ordinary natural kind predicates like ‘is water’ are associated with two different properties. First, they are associated with a descriptive stereotype. This stereotype determines a second, ‘underlying’ property: the property of being of the natural kind that instantiates, or realizes, the stereotype. For a given stereotype, the underlying kind can vary from world to world. In the actual world, H_2O instantiates the water-stereotype, but in a Putnamian Twin-world, XYZ does. In the context of modal operators, however, we suggest, ordinary natural kind predicates stand for the property of belonging to *the kind that realizes the stereotype in the actual world.* In this, ordinary non-scientific natural kind predicates actually differ both from the predicates explicitly expressing the stereotype, and *scientific* natural kind predicates like H_2O. The former always stand for the stereotypical property, while the latter always stand for the underlying kind.

For implementing this idea formally, we shall assume the following: if some particular set e (such as a set of cats, or of water quantities) is the extension of two different natural kinds k and k', then k is a *sub-kind* of k', $k \subset k'$ (or *vice versa*). That is, two natural kinds cannot partially overlap in extension. Under this assumption, there is a universal *minimal natural kind* function K such that $K(e, \boldsymbol{w})$ is the minimal natural kind k such that e is part of the extension of k in world $\boldsymbol{w}$. In case e is not part of the extension of a natural kind at all, K is undefined for the argument.

[15] That would have clearly nonsensical consequences such as making it metaphysically impossible for there to be more, or less, or even numerically different bodies of, water than there actually are.

We also assume that we have an $extension\ function$ E that, for a given natural kind k and world $\boldsymbol{w}$ gives the extension of k in $\boldsymbol{w}$. With this much of machinery, we have the following. For a predicate P, $i(P, \boldsymbol{w})$ gives the extension of P in $\boldsymbol{w}$. If that is the extension of a natural kind, then $K(i(P, \boldsymbol{w}), \boldsymbol{w})$ is the minimal natural kind k of which $i(P, \boldsymbol{w})$ is the extension in $\boldsymbol{w}$. Finally,

$$E(K(i(P, \boldsymbol{w}), \boldsymbol{w}), \boldsymbol{w}')$$

is the extension of that natural kind in world $\boldsymbol{w}'$.

As a special case we have

$$E(K(i(P, \boldsymbol{a}), \boldsymbol{a}), \boldsymbol{w})$$

which is the extension in a world $\boldsymbol{w}$ of the minimal natural kind exemplified by the extension of P in the $actual$ world. We abbreviate this $i_N(P, \boldsymbol{w})$. For instance, since we believe that the extension of ‘water’ in the actual world is the extension of the natural kind H_20, putting ‘water’ for P above gives us the quantity of H_20 in $\boldsymbol{w}$.

So, in case P is an n-place natural kind predicate, we can extend the atomic clause for actua-truth accordingly:

(A+) Actua-true($Pt_1, \ldots, t_n, \boldsymbol{w}$) iff $\langle i(t_1, \boldsymbol{a}), \ldots, i(t_n, \boldsymbol{a})\rangle \in i_N(P, \boldsymbol{w})$.

Now, we can reproduce the distinction drawn above between sentences true at every possible world and sentences whose necessitations are true for sentences containing natural kind predicates. For instance, using ‘Tiger’ as a natural kind term that picks out a natural kind in virtue of the fact that the $tiger$ stereotype picks out a natural kind on Earth in the actual world, and using ‘FT’ (‘Felis Tigris’) as a predicate that picks out tigers in virtue of other properties than the stereotype (whether or not it picks out the extension only in virtue properties essential to the natural kind), we have

(14) $\forall x(Tiger(x) \leftrightarrow FT(x))$

as sentence true in some world (for instance, the actual world) and false in others (where other creatures than those belonging to Felis Tigris satisfy the stereotype). On the other hand,

(15) $\Box\forall x(Tiger(x) \leftrightarrow FT(x))$

is true, since both terms are natural kind terms, picking out the same natural kind in the actual world.

Further, if ‘Stiger’ is a predicate abbreviating the tiger stereotype, then

(16) $\forall x(Tiger(x) \leftrightarrow Stiger(x))$

is true in every possible world, while

(17) $\Box\forall x(Tiger(x) \leftrightarrow Stiger(x))$

is false (since ‘Stiger’ is not itself taken to be a natural kind predicate).

Again, we suggest, this draws *some* distinction between between analytic and non-analytic necessity. (16) is intuitively analytic, while (15), although a true necessity statement, is intuitively non-analytic. Even though the value of this treatment remains to be determined, it is at least clearly one-dimensional.

References

Chalmers, D., 1996, *The Conscious Mind*, Oxford University Press, Oxford.

Glüer, K. and Pagin, P., 2006, `Relational modality', in Lagerlund *&* al. 2006.

Glüer, K. and Pagin, P., forthcoming, `Proper names and relational modality', *Linguistics & Philosophy* .

Kaplan, D., 1989, `Demonstratives', in J. Almog, J. Perry and H. Wettstein (eds.), *Themes from Kaplan*, Oxford University Press, Oxford.

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

Lagerlund, H., Lindström, S. and Sliwinski, R. (eds.), 2006, *Modality Matters. Twentyfive Essays in Honour of Krister Segerberg*, volume 57 of *Uppsala Philosophical Studies*, Department of Philosophy, Uppsala University, Uppsala.

Putnam, H., 1975a, `The meaning of 'meaning'', in K. Gunderson (ed.), *Language, Mind and Knowledge*, Minnesota Studies in the Philosophy of Science VII, University of Minnesota Press, Minneapolis. Reprinted in Putnam 1975b. Page references to the reprint.

Putnam, H., 1975b, *Mind, Language and Reality. Philosophical Papers vol. 2*, Cambridge University Press, Cambridge.

Quine, W. V. O., 1951, `Two dogmas of empiricism', *The Philosophical Review* 60:20--43. Reprinted in Quine 1980, 20--46. Page references to the reprint.

Quine, W. V. O., 1960, `Carnap and logical truth', in *The Ways of Paradox and Other Essays*, 107--32, Harvard University Press, Cambridge, Mass.

Quine, W. V. O., 1980, *From a Logical Point of View*, Harvard University Press, Cambridge, Mass., 3 edn.

Rabinowicz, W., 2006, `Analyticity: An unfinished business in possible-world semantics', in Lagerlund *&* al. 2006, 345--57.

Soames, S., 2002, *Beyond Rigidity: the Unfinished Semantic Agenda of Naming and Necessity*, Oxford University Press, Oxford.