HAL
archives-ouvertes.fr

Are two Dimensions Too Many? - A one-dimensional rival to two-dimensional semantics

Sacha Bourgeois-Gironde, Denis Bonnay

▶ To cite this version:

Sacha Bourgeois-Gironde, Denis Bonnay. Are two Dimensions Too Many? - A one-dimensional rival to two-dimensional semantics. 2005. <ijn_00000659>

HAL Id: ijn_00000659

http://jeannicod.ccsd.cnrs.fr/ijn_00000659

Submitted on 23 Dec 2005

Manuscript Number:

Title: Are Two Dimensions Too Many? A One-dimensional Rival to Two-Dimensional Semantics

Article Type: Original Research

Section/Category:

Keywords: Two Dimensional Semantics, Semantics for Knowledge and Belief, Modal logic, Possibility, Conceivability, Impossible worlds, Theory of reference, Transworld Identity

Corresponding Author: Dr Sacha Bourgeois-Gironde, PhD

Corresponding Author's Institution: Institut Jean Nicod / ENS LSH

First Author: Sacha Bourgeois-Gironde, PhD

Order of Authors: Sacha Bourgeois-Gironde, PhD; Denis Bonnay, MD

Manuscript Region of Origin:

Abstract: We discuss two interpretations of two-dimensional semantics (2DMS) of modality due to D. Chalmers and R. Stalnaker. The main problem with both interpretations of the formal framework is the relinquishing of rigidity for terms. They are in this sense unfaithful to an agent's beliefs. We present alternative principles to capture what we take to be agents' beliefs. These are the principles of hyper-rigidity and backward reference to actuality. We propose, then, to go back to a one-dimensional semantics which affords a satisfactory model of beliefs' reports. Our proposal, like 2DMS, addresses typical problems of representation for beliefs and epistemological difficulties related to modal knowledge.

abstract

We discuss two interpretations of two-dimensional semantics (2DMS) of modality due to D. Chalmers and R. Stalnaker. The main problem with both interpretations of the formal framework is the relinquishing of rigidity for terms. They are in this sense unfaithful to an agent's beliefs. We present alternative principles to capture what we take to be agents' beliefs. These are the principles of hyper-rigidity and backward reference to actuality. We propose, then, to go back to a one-dimensional semantics which affords a satisfactory model of beliefs' reports. Our proposal, like 2DMS, addresses typical problems of representation for beliefs and epistemological difficulties related to modal knowledge.

Are Two Dimensions Too Many? A One-dimensional Rival to Two-dimensional Semantics

Abstract. We discuss two interpretations of two-dimensional semantics (2DMS) of modality due to D. Chalmers and R. Stalnaker. The main problem with both interpretations of the formal framework is the relinquishing of rigidity for terms. They are in this sense unfaithful to an agent's beliefs. We present alternative principles to capture what we take to be agents beliefs. These are the principles of hyper-rigidity and backward reference to actuality. We propose, then, to go back to a one-dimensional semantics which affords a satisfactory model of beliefs reports. Our proposal, like 2DMS, addresses typical problems of representation for beliefs and epistemological difficulties related to modal knowledge.

Introduction

There are two basic stories concerning how the reference of proper names is fixed through possible worlds. The Hintikkean account associates names with individual concepts: each speaker associate with names some modes of identification that determine reference in belief worlds. Thus, "Cicero", as used by a speaker with a poor knowledge of ancient Rome, does not rigidly refer to Cicero, but, for example, to whomever she might take to be the most famous orator of the last years of the Republic. The Kripkean account has it that individuals, rigidly designated by the names we use for them, are the basic building blocks of possible worlds: there is in fact no problem as to how the reference of "Cicero" is fixed, because it is precisely used to refer to Cicero, whatever his occupation in Rome around 50 B.C., including or not being a famous orator.

At first sight, these two accounts seem incompatible. But one can make them compatible if one takes them to deal with two different kinds of possibility, namely, for the first one, epistemic possibility, and, for the second one, metaphysical possibility. The idea is thus that two coreferential terms can fail to corefer, rigidity notwithstanding, in a belief world, just in case the agent is mistaken as to the reference of these terms and does not know they do corefer. Starting with this idea, proponents of two dimensional modal semantics (2DMS for short) suggest to merge the two accounts into one and the same theory. In 2DMS, a sentence can be associated with two kinds of intensions, the first one along Kripkean possible worlds, which will account for the fact

that the sentence expresses a necessary or contingent propostition, and the second one along Hintikkean possible worlds, which will account for failure of substitutivity in intensional contexts and for epistemic properties of sentences.[1]

Two-dimensional modal semantics is therefore an attractive framework: it makes it possible to give a possible world analysis of content compatible with Kripke's analysis of rigidity. But this synthesis rests on an equivocation. The mere gluing of the two approaches does not account for the speaker's intuition as described by Kripke. If rigid designation, and individuals picked through it, is all what possible worlds are about, and if the content of the sentence expressed by a speaker is to be a set of possible worlds, these possible worlds should be built on the very basis supplied by rigid designation, since it is the only basis on which we have an intuitive access to possible worlds. If I can consistently believe that Hesperus and Phosphorus are not the same thing, it should be just because I can consistently of Hesperus and of Phosporus that they are not the same thing. Unfortunately, this is not allowed in the 2DMS framework, since it represents the possibility that they are not the same by resorting to a world in which Hesperus or Phosphorus or both do not refer to what they actually refer to. But as far as it is possible for me to have *de re* beliefs about Hespherus and Phosporus, any account relying on this kind of trick misrepresents what happens when I entertain the possibility that they are different, since it represents it by two *other* objects which are in fact different, while Hesperus and Phosphorus on which the agent's thought bears are actually identical. Our aim is thus to represent on the contrary this possibility that I might entertain as the possibility that Hesperus and Phosphorus – the two objects in the sky that I can rigidly refer to – are different, even if this possibility is not a metaphysical possibility (if one grants that identity is necessary).

2DMS was precisely conceived as a theory whose ideal was to reconcile first-person intuitions about meaning and content and "official" semantic content. Sometimes the two coincide, but in most interesting cases there is a conflict between intuitions and actual content. It is solved by 2-dimensionalist in terms of theoretically prescribed Kripkean content on one side, and Fregean content, on the other side, the latter meant to reflect the agent's intuitions about meaning. However, the agent may resist this explanation and be convinced of her mentally having a firm grip on a Kripkean content. One basic critical point is that 2DMS is in some important respect cognitively inadequate because it too soon departs from possible Kripkean intuitions the agent may want to preserve in contexts in which 2DMS intends to make sense of the agent's beliefs by reinstating Fregean semantics.

One reason why we wish intuitions to support our semantic theory is that in one major area of applications of 2DMS we have nothing but intuitions, namely modal ones, to account for. The agent has some intuition about what could be the case. She may have some further specific intuition about the way she uses the terms that shape up her modal intuition. But in fact our only access to the agent's modal intuition is through the way she thinks she uses her terms and her report that she used it rigidly or descriptively. In contexts of that sort, in order to make correct predictions about what modal judgements are expressed, it is crucial that our theory fits these semantic intuitions. Unlike 2DMS we claim that in most cases modal intuitions are supported by Kripkean intuitions concerning the semantics of the relevant terms, which means that their use is intended to be rigid. Fregean contents were reinstated by 2DMS theoreticians in order to make sense of seemingly contradictory or irrational modal beliefs like the ones which imply a violation of the identity of referred to individuals.

We defend an alternative way to connect modality and rationality by keeping all along one single kind of semantic content and subsequently by defusing the systematic transition between intuition and possibility Fregean contents were supposed to allow back in the hands of 2DMSists. This needs not be taken as an empirical point about the nature of modal intuitions but as an important conceptual possibility that may make empirical and intuitive sense. To put it bluntly: If I believe that water could have a chemical structure different from H2O, what I may simply mean is that I believe of H2O that it could be something different from H2O. I may well believe an impossibility about H2O and water rather than a possibility about something I did not obviously intended to refer to. We prefer a theory which predicts this kind of semantic lucidity and modal (really modal) illusions to one which attributes to the subjects an unintended change in the use of their terms. We try to present such a theory and its comparative advantages and costs with 2DMS.

We first present the 2DMS setting, emphasizing how the addition of a new dimension can help to solve the problem of putting together the Kripkean account of proper names and the possible world account of content. This point is a bit tricky because of the various interpretation of the formal setting of 2DMS: Stalnaker's metasemantic interpretation seems immune to some of the criticisms aimed at Chalmers's semantic interpretation. But we see in the second section that the objections we direct at 2DMS bear on both interpretations. Our main point is that one should remain faithful in hostile contexts to Kripkean intuitions of rigidity which we think continue to adequately characterize the semantic behavior of singular terms and natural substances terms in such contexts. When we deal with the same problems that 2DMS was

introduced to solve, we do not say that the subject is not mistaken in believing what she believes, but we at least suppose that she knows what she believes. This epistemic candor carries over to the semantic level by keeping afloat intuitions of rigidity in those places where Fregean contents were reinstated by 2DMS. We are consequently led to extend the principle of transworld identity to worlds figuring in the content of beliefs deemed contradictory; hence an enrichment of modal space with a species of impossible worlds, which we prefer to its duplication by 2DMS. The formalism is introduced in the third section. We spell out a few semantic requirements and our vision of modal space in paragraphs 3.2 and 3.3. It is remarkable that our one-dimensional modal semantics (1DMS) both accommodates traditional objections (paragraphs 2.1) that were formulated against 2DMS as well as our own. As we said, our particular objections – in paragraphs 2.2 and 2.3 – hint back to the idea that semantic predictions must be subjectively adequate to the subject. We think that ordinary modal intuitions are better dealt with by incorporating in our semantics so-called requirements of hyperrigidity and backward reference to actuality. Those principles state respectively that the subject keeps the reference of her terms, once rigidly fixed, in every context and that she remains lucid about the kind of departure from actuality her intuitions may impart. We can easily translate those requirements in our model, which in one case (hyperrigidity) 2DMS refuses to do by principle, so to speak, and in the other (backward reference to actuality) it simply cannot do, as we show below. This mono-dimensionalist approach is not free lunch. The main hypothesis supporting its implementation is that proper names are directly used to refer to individuals in possible situations. Each name comes with a witness for it, and though it will generally be the case that such a witness is identical in a certain world with certain objects, no such thing is necessary from the point of view of the subject. This implies that we allow for impossible variation of the identity relations in certain worlds, i.e. we allow for some impossible worlds in order to complete the picture of the modal space. But it will be argued in paragraph 3.4. that our reliance on impossible worlds is relatively immune to usual criticisms against impossible worlds.

1. Aims and scope of two-dimensional modal semantics

1.1. THE FORMAL APPARATUS OF 2DMS

Two-dimensional modal semantics is a modal semantics according to which expressions are evaluated with respect to pairs of worlds instead of worlds simpliciter. The idea is that ϕ is true at (w_1, w_2) iff it would be true in w_2, considered as counterfactual, had w_1 been the actual world. Let's consider the following statements:

(ϕ) Hesperus appears in the evening sky.

(ψ) Hesperus is Phosphorus.

Now let w_0 be the actual world, and w_1 a possible world in which the morning star is still Venus but in which the object appearing in the sky on the evening is Sirius instead of Venus. In (w_0, w_0), reference is fixed in the actual world, and sentences are evaluated in the actual world as well, so that ϕ and ψ are both true in (w_0, w_0). In (w_0, w_1), reference is still fixed in the actual world, so that "Hesperus" and "Phosphorus" both refer to Venus, but the circumstances of evaluations are given by the counterfactual world w_1. Because reference fixing has not changed, ψ is still true, and because facts are different, ϕ is false. In (w_1, w_1), reference is fixed in w_1, therefore "Hesperus" refers to Sirius, whereas "Phosphorus" refers to Venus as before. As a consequence, ψ will be false, but as the circumstances of evaluation are given by w_1, ϕ is true, since w_1 is possible world in which Sirius appears in the evening sky.

Statements are thus interpreted by a function with two arguments, this is Stalnaker's notion of propositional concept (Stalnaker 1978). A propositional concept $f(x_1, x_2)$ can express two distinguished propositions:

i) first the horizontal one, $f(w_0, x)$, corresponding to the proposition usually associated to the statement, all reference fixing takes place considering the actual world as actual. To fix terminology, this is Chalmers's secondary intension (Chalmers 1996) and Jackson's C-intension (Jackson 1998).

ii) The other proposition of special interest is $f(x, x)$, the diagonal proposition, which gives the truth value of the statement in a world, that world being considered as actual. This is Chalmers's primary intension and Jackson's A-intension.

1.2. EXPECTED ADVANTAGES

This formal device hopes to solve two kinds of intricate problems.

representation problem Rigid designation generates various semantic puzzles. More precisely, the problem is to make Kripkean rigid-

ity and facts about content compatible with interpreting proposi-
tions as sets of possible worlds. One special aspect of this problem
concerns identity statements: if Kripke is right, all true identity
statements 'a=b' are necessarily true. They have therefore the
same truth value in every possible world, and hence the same
propositional content, and this content is trivial. The notion of
propositional content does not seem to give then an adequate
representation of the informative content of these statements.

Another side of the problem concerns beliefs: as long as 'a' and
'b' have the same referent, the propositional content of '$\phi(a)$' and
'$\phi(b)$' will be the same, nevertheless, it seems perfectly rational for
an agent to believe both $\phi(a)$ and $\neg\phi(b)$ in case he does not believe
that $a = b$. How then can we represent this kind of belief?

epistemological problem This is is the looser problem of providing
a better understanding of the post-Kripkean interplay between the
distinctions a priori / a posteriori (epistemic) and necessary /
contingent (metaphysical). On the traditional view of epistemic
logic, epistemic alternatives for an agent are represented as sets of
possible worlds, and possible worlds are the metaphysically pos-
sible worlds (what else?). How then could it be that something
necessary is not a priori known or at least knowable?

In a nutshell, 2DMS solves the representation problem by resorting
to primary intensions. "Hesperus is Phosphorus" does not lead to a
necessary diagonal proposition, just because in (w_1,w_1), the object
which is the evening star and the object which is the morning star
according to w_1 are different. The same idea works with belief: "Hes-
perus is a planet" and "Phosphorus is a planet" are interpreted by
different diagonal propositions. In the world w_1 discussed above, the
morning star happens to be Venus; 'Hesperus is a planet' is false in
(w_1,w_1) whereas 'Phosphorus is a planet' turns out to be true.

Given that a single statement expresses two different propositions,
horizontal and diagonal ones, we have two notions of being true ev-
erywhere. Necessity is interpreted as usual as necessity of the horizon-
tal proposition. Epistemically, one must first grant that the diagonal
proposition is determined by the descriptive content of terms - rigid
or not. The idea is that one determines the referent of 'Hesperus' in w
considered as actual by asking which object is the evening star in w. The
presupposition is that 'Hesperus' has as descriptive content 'the first
heavenly body appearing in the evening sky' and that's why 'Hesperus
appears in the evening sky' is true in every (w, w). 'Hesperus appears
in the evening sky' is a priori. More generally, always true diagonal

propositions are true in virtue of the descriptive content, or meanings of the terms. Therefore, these propositions are a priori.[2]

1.3. SEMANTIC VS METASEMANTIC INTERPRETATIONS

Even though 2DMS is a valuable tool in providing the formal setting to solve these problems, it must be recognized that the validity of its solution is highly dependent upon non-trivial theses about meaning and belief. The formalism in and of itself is by no means an argument for these theses. In particular, there is a persisting disagreement over the claims of 2DMS to model and explain a priori knowledge: Stalnaker (Stalnaker 2001) has proposed to label *semantic* the interpretation of 2DMS according to which the diagonal proposition is a component of meaning and *metasemantic* the interpretation according to which the diagonal proposition only represents the possible variations of meanings depending on the facts.

The semantic view, defended by Chalmers and Jackson, rests upon a commitment to descriptivism: to determine the value of 'Hesperus' along the diagonal line, one appeals to the descriptive content of the name. "Descriptive content" here is tantamount to some kind of Fregean sense. The existence of such a sense is a necessary condition for truth everywhere along the diagonal line to be interpretable as a priori truth. This new form of descriptivism has recently come under heavy attack: the core argument is that the other kind of intensions the semantic interpretation presupposes simply do not exist, because the kind of knowledge they imply cannot be tied to knowledge of meanings. Arguments here are the same old ones as Kripke's against the descriptive theory of proper names. The competent user of a name does not in general possess substantial identifying knowledge. The speaker might fail to have such knowledge, by ignorance (many people do not know enough about Cicero to be able to identify him uniquely) and by error (while our beliefs about G'odel's actions might be false, they would not prevent us from referring to him). Therefore, mastering the use of a name does not presuppose knowledge of some kind of identifying meaning.

The metasemantic view isn't committed to such descriptivism. No link between a priori truth and truth everywhere along the diagonal line is suggested. Moreover, there are no a priori truths, since if semantic facts had been different, meanings could have been radically different so that "The meter stick is one meter long" could have meant "Michael Jackson is five feet ten inches tall". The solution of the epistemological problem is partial: on the metasemantic view 2DMS handles well a

posteriori necessary statements, but it falls short of explaining a priori contingent ones.

2. Objections to two-dimensionalism

We first consider criticisms of 2DMS found in the literature. They bear mainly upon the semantic interpretation, but we will argue that the way out suggested by the metasemantic interpretation is not as safe as it seems. We propose some new criticism of what's wrong in general with the 2DMS strategy: we introduce two principles governing our modal intuitions, hyperrigidity and backward reference to actuality. We show then how 2DMS violates them.

2.1. Descriptivism strikes back

At first sight, objections against neo-descriptivism are aimed only at the semantic interpretation. The specificity of the metasemantic interpretation is that, though it uses the diagonal "line" to solve the representation problem, it does not presuppose that line to be part of meaning. But, if the diagonal proposition is not part of meaning, how can it play a role in solving representation problems? This question is twofold: if the diagonal proposition is not in general the content of the sentence, how can it be sometimes the proposition expressed? How can we attribute to speakers the necessary knowledge to determine this horizontal line, if the knowledge in question isn't semantical?

Stalnaker's answer to the first question (Stalnaker 1978) is Gricean. "Under certain conditions, the content of an assertion is not the proposition determined by the ordinary semantic rule, but instead the diagonal proposition of the propositional concept determined." And this happens when some principles about assertions are violated:[3]

- Non-triviality: A proposition asserted is always true in some but not all of the possible worlds in the context set.[4]

- Context-independence: The same proposition is expressed relative to each possible world in the context set.

The diagonal proposition comes into play because because some conversational maxims would be violated, had the content been the usual one, so that the content has to be reinterpreted along the metasemantic line.

As to the second point, the determination of the diagonal proposition seems to involve reflexive semantic competence on behalf of

the speakers. Speakers are supposed to be able to reflect on the fact that words could have meant something other than what they mean, given the pragmatic fact that they cannot mean what they usually do. Therefore, through the metasemantic interpretation, Stalnaker wants to have it both ways: the diagonal proposition does not depend on the identification of a descriptive content but it is nevertheless available to solve the representation problem. But this tricky strategy lays itself open to criticism.

- The access to the metasemantic dimension presupposes nearly as much of the speakers as would be needed to fill a descriptive content. If speakers have to master a Kripkean theory of reference to be able to determine the diagonal proposition expressed by an assertion, why should we refuse to construe this as constituting the Fregean sense of proper names?

- The Gricean account works well in the case of possibly trivial assertions, because it is clear that in those cases, some other interpretations have to be found. But there is no such risk of trivialization with ordinary beliefs involving proper names, when they are taken one by one, so why should we diagonalize in those cases? And if it is argued that we face contradiction when one puts together various such beliefs, so that in general their content should be the diagonal proposition, this will yield an unexplained systematic difference between the content of simple assertions and beliefs ascriptions.

- The semantic interpretation has some strong initial plausibility in the case of ready-made descriptive names like Evans' Julius, who is taken to be whoever invented the zip (Evans 1977).

We think one major lesson should be drawn from this first set of problems: a solution for the representation problem should be independent of one's position with respect to the epistemological problem. It should not depend explicitly on questionable neo-descriptivist assumptions. It should instead be able to solve the epistemological problem insofar as these assumptions are true (be it for the whole class of proper names and natural kind terms or just for a restricted and perhaps artificial subclass of such expressions) and it should not make any hidden assumption as to the metasemantic reasoning abilities of the speakers.

2.2. HYPERRIGIDITY

One of the ideas behind 2DMS, according to (Chalmers 2004), is to reconcile Kripkean and Fregean intuitions. Rigidity is ensured at the

horizontal level, and matters of informativeness and cognitive content are dealt with at the level of the diagonal proposition, either with some full-blooded Fregean sense along the lines of the semantic interpretation or with something more elusive and not semantic on the metasemantic interpretation.

But is 2DMS really faithful to the Kripkean intuitions to the end? When a lay astronomer, say O'Leary, believes that Hesperus is not Phosphorus, he believes of Hesperus and of Phosphorus that they are not the same star. It happens that Hesperus and Phosphorus are not two different stars, but one and the same planet. If, following Kripke, identity statements of this form are necessary whenever true, this means that the lay astronomer considers possible some impossible state of the world, and moreover believes of this very possibility that it is actualized.

Surprisingly enough, the formal account of O'Leary's belief 2DMS puts forward is substantially different from the informal one we have just sketched. On the semantic interpretation, "Hesperus is not Phosphorus" is not a priori false because there is some possible world w such that, on considering that world as actual, the objects satisfying descriptive meanings of "Hesperus" and "Phosphorus" would be different. That is, "Hesperus is not Phosphorus" will be true in w because at least one object other than Hesperus, alias Phosphorus, alias Venus, satisfies the definite description attached for example to "Hesperus". The way the reference of "Hesperus" has been fixed on the metasemantic account is more complicated, but the upshot is the same: the sentence will be true only because the proper names have been reinterpreted to refer to something different from their actual referent. This trick allows the proponent of 2DMS to find a (truly) possible world to back up O'Leary belief in something impossible, but this trick amounts to a renunciation. For w to be the witness we need, the requirement of rigidity has to be relaxed. Maybe here the neo-descriptivist can bite the bullet and show that abandoning rigidity is necessary to account for descriptive meaning. The proponent of the metasemantic view could also reply that Kripke is not betrayed in so far as we are on a metasemantic account, so that the name "Hesperus" does not only change its reference, but becomes a different name. But anyway, O'Leary is betrayed at least as much as Kripke: O'Leary belief was about Hesperus and Phosphorus, he believed that they are different objects, he would not regard the idea of another object coming into play to become the referent of "Hesperus" as an adequate way to construe his belief, and that is also precisely the reason why we, as well-informed astronomers, think he is believing something impossible.

Another unwelcome consequence is that our disagreement with O'Leary on the 2DMS approach to the representation problem turns into a verbal dispute. It is not really that we regard as impossible a world O'Leary regards as possible and actual, it is rather that O'Leary regards as possible that 'Hesperus' and 'Phosphorus' are names different from what they are. Surely this description of the disagreement is not faithful to the intuitions of the speakers. On the contrary, it seems that we are engaged in a substantial disagreement with O'Leary.

Nevertheless, one could argue that the 2DMS approach is in keeping with Kripke's own suggestions in *Naming and Necessity*. According to Kripke, the epistemic possibility that Hesperus and Phosphorus turn out to be two different stars is not really the possibility that Hesperus is not Phosphorus. Such possibility does not exist, since the identity between Hesperus and Phosphorus is true and necessarily true. Rather we envisage the possibility of a situation in which we would be in the same epistemic state as we are and in which 'Hesperus' and 'Phosphorus' would be used to name other objects. But here it is important to distinguish two claims: the metaphysical claim that the epistemic possibility does not give way to a genuine possibility, and the representational claim that the epistemic possibility should be construed as a genuine possibility if we modify the language. The first claim is established by usual argument in favor of rigidity, but the shift of the first claim to the second one is a non sequitur. Moreover, this shift seems to be in conflict with Kripke's view about access to possible worlds. According to Kripke, names should be regarded as rigid designators because worlds are not things waiting around there to be explored and described. On the contrary, our access to counterfactual situations depends on our ability to use proper names to fix what we are speaking about. When we entertain the idea that Aristotle could have never studied with Plato, there is no back-up description of Aristotle $\phi(x)$ such that what we are thinking about is that $\imath x\phi(x)$ could have never studied with Plato. Similarly, our access to a counterfactual situation according to which Hesperus is not Phosphorus should not depend on some qualitative description, but be the direct consequence of an act of referring. And this contradicts Kripke's claim that this epistemic possibility should be construed as a genuine possibility modulo a change in language .[5]

Supporters of 2DMS have no clear counter-arguments with respects to the hypothesis of epistemic rigidity, that is to say, the fact that a subject preserves, or may preserve, the habitual reference of his rigid terms in all contexts, including in contects in which an essential property of the objects to which the terms in question customarily refer is negated. Epistemic rigidity mlay be accompanied, to some ex-

tent, by the consciousness that the subject expresses an impossibility regarding the objects that form the habitual reference of his terms, and nevertheless think that she expresses well that impossibility with respect to those objects. It is just when the subject is not conscious of expressing an impossibility, but does not change the habitual reference of his terms, that she may be said to be victim of modal illusion. Faced with the idea of preservation of intensional identity of the reference of rigid terms, the 2-dimensional semanticist deploys in an unacceptable manner an argument of Kripkean derivation. Essential properties of an object, so it goes, would justify the change in reference of the terms in the context where the predication of those properties with respect to the object is implicitly or explicitly negated. In other words, the 2-dimensionalist is tempted to use the metaphysical thesis that there are essential properties in service to a certain theory of reference and thus do as if violations of essentialism may justify locally a change in the reference of the terms concerned. But the Kripkean order of reason is precisely the opposite : intuitions of rigidity regarding proper names and natural kind terms give a precise logical semantic basis (through the notion of identity across possible worlds) to the independent metaphysical notion that objects have some of their properties essentially. The rigidity thesis is not itself dependent on the existence of essential properties.

From our point of view, the lesson to be drawn from the second set of criticisms is that the solution to the representation problem should not consist in giving up rigidity or in appealing to change in language. That is we should satisfy the following principle:

Hyperrigidity principle: Rigid terms do not loose their rigidity when they are embedded in knowledge claims or in illusory modal claims.

The partial solution to the epistemological problem 2DMS brings about explains away our modal illusions, that is the fact that we can conceive of impossible state of affairs. We can conceive, and we can even believe that water is not H2O; but according to 2DMS, this does not mean that we can conceive some metaphysically impossible state of affairs, rather what we conceive is to be taken into account through a change as to which world is regarded as actual and reference fixing. In contradistinction with such a solution bypassing the subject's intuitions, we put forward the hyperrigidity principle as a subjective adequacy constraint on formalization of content.

2.3. BACKWARD REFERENCE TO ACTUALITY

Our second argument against 2DMS and our corresponding positive principle concerns its lack of expressive power relative to a number of simple cases : cases in which the contrast between the impossibility considered by a subject and the actual state of affairs is explicitly entertained. For example " Water could have had a different chemical structure from the one it actually has ". Modal intuitions typically present this kind of contrast. From a cognitive standpoint, that corresponds to a certain lucidity on the part of the subject with respect to the possibility or impossibility that he considers. More exactly, the fact of considering that modal intuition does not imply that the subject is ignorant about the actual state of affairs, nor that she refuses to consider that the model judgment that she formulates is not relative to actuality. In the case in which the contrast with actuality is thus sufficiently indicated in the modal judgment of the epistemic agent, 2DMS becomes an inadequate formal framework to give an account of such a judgment. The difficulty for the 2DMS framework arises from a logical contradiction that appears in the scope of the modal operator associated to the procedure that consists in taking a certain world as counter-actual (at the basis of the selection of a primary intension), where the backward reference to actuality is an explicit component of the modal judgment.

Let's specify the intuitive contrasts at stake and see when they give rise to a contradiction or not, and let's consider this as an interesting test of expressivity of 2DMS in the face of natural modal intuitions.

Modal intuitions are contrastive in the sense that they imply conscious departures from actuality. Contrasts can come in several ways. In one way I fix actuality as I know it and wonder how different than it is it could have been or might be. My intuitions then relate to actuality and I consider its possibly alternative states. In some such occasions I may further suppose that one of these states is the actual one and I mentally cancel out the contrast that made that alternative state salient. In another way I focus on the alternative, counterfactual state and I remain lucid about its implied departures from actuality. I never pretend that it is an actual state. 2DMists have respectively labeled these intuitive procedures 'considering a world as actual' and 'considering a world as counterfactual', and associated with them dual logical, linguistic and conceptual realizations. Our point, here, is that considering a world as actual is tantamount to an attempt at canceling out the contrastive relation our modal intuition bears to actuality whereas considering it as counterfactual consists in keeping this contrast under focus. This point bears on its sleeves an attempted conciliation of a formal account with

14

significant cognitive features of our modal intuitions. The problem is that not all kinds of contrasts and only rather unsubstantial ones will be properly captured by a 2DMS account. This is due to a contradiction arising in the scope of the modal operator associated with the 'considering a world as actual' procedure when a backward reference to actuality is an essential feature of the corresponding modal intuition. In some cases however this reference to actuality is not essential and apart from particular epistemic conditions which block the soundness of the projection into counteractuality mechanism even in those cases, 2DMS has a grip on them. Let's see some examples.[6]

(1) Water could have a chemical structure different from H20.

(2) Water could have had a chemical structure different from the one it actually has.

(1) is the kind of metaphysical impossibility 2DMS is fit to account for. For (1) to be true, according to 2DMS, reference-fixing must takes place in a world different from the actual one: if the drinkable stuff found in lake and oceans is XYZ in w_1, then, as long as we consider w_1 as actual, the reference of "water" will be XYZ. The important point is that for (1) to be true, a new "actual" world must be selected, that is the contrast between w_1 and the truly actual world must be canceled, we have to do as if w_1 was the actual world. Formally, let's assume, for the sake of simplicity, that "water" and "H20" are proper names. We will use a rigid function f which associates to each natural substance its chemical structure. (1) is then parsed as $\Diamond f(a) \neq b$. $f(\imath x A \phi(x)) \neq b$ will be true in the pair of worlds (w_1, x) where w_1 is a world considered as actual differing from the actual world as to which substance plays the role of water, because the reference of a will be determined through $\imath x A \phi(x)$ by jumping to w_1 whatever x is.

Now (2) being a metaphysical impossibility, such a modal illusion should be explained away in two-dimensional terms in the same way, by postulating a counteractual world in which a chemical structure which H20 does not actually possess applies to water in that world. A first reading of my modal intuition is that it seems to me that it could actually be that water, which actually has chemical structure H20, has chemical structure XYZ. However a backward reference to actuality in the content of the modal judgement entails a contradiction. The apparent solution would be to apply the same innocuous variation procedure as in (1). But it won't do for a simple reason. (2) does not mean that the chemical stuff chosen as water in a certain counteractual condition has a different chemical structure than water has. I mean that actual water has, counteractually, a different chemical structure than it actually has. Actual water and its actual chemical structure irreducibly figure in the content of my modal intuitions as the M2D

should faithfully intend to spell it out, and this generates an explicit contradiction in the 2DMS account of that intuition.

Let's make this point again in formal terms. (2) must be parsed as $\Diamond f(a) \neq [\imath y Ay = f(a)]$. Expanding a, we obtain $\Diamond f(\imath x A\phi(x)) \neq [\imath y Ay = f(\imath x A\phi(x))]$. But, on the usual interpretation of A, if f is rigid, [7] $f(\imath x A\phi(x)) \neq [\imath y Ay = f(\imath x A\phi(x))]$ is false in every pair or worlds.[8] Therefore, 2DMS cannot take into account the conceivability of (2), and this is because its explanation of (1) rests upon a shift in actuality which prevents from taking into account the backward reference to actuality made explicit in (2).

Proponents of 2DMS might wish to deny that (2) is really conceivable; but this way out would be *ad hoc*. If it is conceivable that water is not H2O, it is because it is conceivable, even though it is not metaphysically possible, that water has a chemical structure different from the one it actually has. No account allowing for (1) but not for (2) to be conceivable will do. On the contrary, a satisfactory account of conceivability must allow for such backward reference as "different from the one it actually has". 2DMS has a problem with that, because it rests upon considering change as to which world is actual. In contradistinction with that limitation, a satisfactory solution should abide by the following principle:

Backward reference principle Modal claims, even illusory ones, are grounded in the ability to conceive possibilities explicitly differing from reality.

3. One-dimensional semantics revisited

3.1. Rigid designation and impossible worlds

According to possible world semantics, the content of a sentence can be represented as a set of worlds, namely the set of worlds in which the sentence is true. We can take this representation to be a mere semantic machinery aimed at deriving correct predictions (e.g. concerning which sentences entail which ones). Nevertheless, something more is expected from a correct semantics it should reflect faithfully what it is for a speaker to understand a sentence. This is the gist of possible world semantics: understanding a sentence boils down to knowing which possible states of affairs verify it. In this sense, semantic knowledge involves not only the ability to evaluate a sentence in the actual world, but also the ability to evaluate the sentence in a possible state of affairs, that is the ability to conceive of what the world had to be like for the sentence to be true or false. The representation of content based on

possible worlds has then to be linked with speakers' abilities to conceive of possible of worlds. On the contrary, if we take the possible worlds used to represent content to be cognitively inaccessible to the speakers, semantic competence remains a mystery, because speakers do not have access to the content as characterized by the theory.

The point of the criticisms against the descriptive elements common to both semantic and meta-semantic interpretations of 2DMS is that this access remains mysterious: in general, speakers do not possess a descriptive content attached to proper names, nor do they master a theory of how worldly facts influence reference fixing for proper names. On the contrary, the Kripkean story about possible worlds provides us with a clear account of our acquaintance with possible worlds. A possible world is just an ascription of certain properties to the individuals we refer to in our language. I can refer to Cicero using the word "Cicero" and on this basis, I can imagine that some properties apply to him and that some others do not. Thus, I can conceive that Cicero was not an orator, and this ability is grounded in the use of "Cicero" as a directly referential term, which calls Cicero before our mind's eye so to speak and makes for a possibility in which we can attribute to him a wide range of properties, without paying any kind of respect to a putative descriptive content of the name. But as easily as I can conceive that Cicero was not an orator, I can play with the idea that he was not Marcus Tullius, so that I can consistently believe that Cicero is not Marcus Tullius. Here the Kripkean story falls short of giving an explanation in terms of real possibility, because if "Cicero" and "Marcus Tullius" both directly refer to some object, there is no room for a failure of coreference. But if what we are after is an account of our modal competence, that is an account of conceivability, in contradistinction with an account of metaphysical possibility, there is no reason to stop here. The object summoned up by the use of the word "Cicero" and the object summoned up by the use of the words "Marcus Tullius" do not have to be the same: as far as I use them to conceive of various situations, it can be the case that they are identical, but it is equally possible that they are not. What I can conceive is constrained by the language I use, I cannot imagine that Cicero is not Cicero, because in this case there are no two names to be used to evoke two different objects. On the contrary, metaphysical possibility does not limit what I can conceive, I can perfectly well imagine that Cicero is not Marcus Tullius, because I can confidently use these two names to evoke two objects which might be different. According to this view, when we try to model conceivability, we should not take identity to be fixed once and for all: this means that it should be considered as a varying relation

which can hold or fail to hold between the objects which are summoned up by the names.

Our talk of names "summoning up" objects and of identity being a "varying" relation might seem mysterious: where do these objects come from, if they are not the real objects dealt with by metaphysical possibility? how can identity be something else than sheer identity? Here comes the substantial thesis, which determines our modeling choices: in terms of conceivability, each name directly refers to an object which we take to be some kind of witness for the name. This object does not have to be an object which actually exist or which could metaphysically exist: it is just so to speak the shadow of the name on which our conceiving abilities rely on. Of course, if the possibilities we imagine are to be full-fledged metaphysical possibilities, these shadow objects have to identical to some real objects: a conceivable state of affairs about Cicero and Marcus Tullius is a real possibility if and only if the mental witnesses for Cicero and Marcus Tullius are both identical to the Roman man who was actually Marcus Tullius Cicero.

Now, it might seems that we endorse two contradictory claims. First, that names rigidly refer to a witness, second that coreferential names can fail to corefer in some possible worlds. But a contradiction follows only if identity is taken to be strongly necessary, that is if two objects which are identical in one world have to be identical in every world. This fact is a byproduct of the usual interpretation of equality as a logical notion, namely as real identity on every world. But if we deal with witnesses, it is very natural to construe equality as non-logical notion, namely a congruence relation on models: a witness can be identical to a real object in the actual world, and identical to no real object in another world (failure of reference), or might fail in a possible world to be identical to the object it is identical to in the actual world (when I don't know if Cicero is Tullius).

To sum, we have just advocated a framework based on the idea that names rigidly refer to witnesses which are together with actual objects the building blocks of possible worlds. Our task in the next sections will consist in giving such a semantics and showing that it does as well as 2DMS without being subject to the criticisms we addressed to 2DMS.

Before stating the principles our semantics should abide by, let us mention two consequences of our witness-driven approach to possible worlds, though, in the interest of space, we shall not develop them in this paper. First, fiction names and referring names are treated on a par: "Cicero" is associated with a witness it rigidly refers to in all conceivable worlds, and the same is true for "Pegasus", the only difference between the two is that, if I know that Pegasus is nothing but a fiction, there is no belief world in which its witness is identical to

a real object.[9]. We take this similarity to be a welcome consequence of our approach: if our ability to conceive of various possibilities involving Pegasus or Cicero is triggered by our use of these names to summon up a witness which is their fixed reference across conceivable worlds, there is no reason not to treat "Cicero" and "Pegasus" on a par, since they are both just proper names we can use to engage in imagining possibilities, quite independently of the fact that some people actually got acquainted with Cicero whereas nobody ever had the pleasure to meet Pegasus.

Another consequence of our approach is the following asymmetry between conceivability and (metaphysical) possibility. The second one is language independent: Cicero cannot be anyone but Marcus Tullius, and this fact remains unaltered by "Cicero" and "Marcus Tullius" being two different names. On the other hand, conceivability is indeed language dependent: the weird conceivable world in which Cicero is different from Marcus Tullius mirrors the linguistic fact that we have two names at hand to refer to the same real world individual. Once again, we think this is a rather welcome consequence: the conceivable state of affair in question cannot be made sense of independently from the language we use to describe it.[10]

3.2. SEMANTIC REQUIREMENTS

We want to model at once knowledge and possibility, therefore we will use a languages with three kinds of modalities: on top of an undetermined modality $\Diamond_C$ corresponding to conceivability, we add a modality $\Diamond_M$ to deal with metaphysical possibility and the necessary / contingent distinction, and another one $\Diamond_E$ to deal with epistemic or doxastic possibility and the a priori / a posteriori distinction. Moreover, each language L goes with a synonymy function syn_L from names to open formulas. To make the link with Evans' descriptive names and their account in 2DMS, the fact that $Julius = \imath x Actually\ \phi(x)$, where $\phi(x)$ is x invented the zip, will be regarded as a semantic property of the Language L, namely that $syn_L(Julius) = \phi(x)$. In contradistinction to 2DM formal languages studied in (Davies and Humberstone 1980), we have no actuality operator. We do not speak of the synonymy relations explicitly *in* the language[11] syn_L will be defined nowhere if they are no descriptive names, and if some strong version of neo-descriptivism is true, syn_L will be a total function assigning to each name a its sense as given by a definite description $\imath x\phi_a(x)$.

We want the following principles to be validated by our semantics:

1. Non-trivialization: $a = b \wedge \Diamond_E a \neq b$ is satisfiable though $\Box_C a = a$ is valid.[12]

2. Aprioricity:

   if $syn_L(a) = \phi(x)$, then $\Box_E\phi(a)$, but this does not imply $\Box_M\phi(a)$

   $\Box_M a = b \nvDash \Box_E a = b$

3. Necessity of identity: $a = b \rightarrow \Box_M((\exists x\ x = a \vee \exists x\ x = b) \rightarrow a = b)$ is valid.

4. Hyperrigidity: $\exists x\ (x = a \wedge \Box_C(\exists y\ y = a \rightarrow x = a)$ is valid.

5. Backward reference to actuality: $\Diamond_C f(a) \neq b$ is satisfiable with f rigid and b L-synonymous with $\imath y\ y = f(a)$.

These principles are intended to ensure that the representation and epistemological problems can be given a solution (requirements 1. and 2.), such that this solution is both Kripkean in a way reminiscent of 2DMS (requirements 3. and partially 4.) and immune to our previous objections (requirements 2., 4. and 5.).

More precisely, 1. is meant to ensure that the trivialization of content that paves the way to the representation problem does not arise.

2. is intended to capture the fact that semantic facts provide a priori knowledge and that a priori knowledge is not tied to metaphysical necessity, as (Kripke 1980) has shown. Kripke's examples of a priori and contingent statements ϕ will lead to situations in which $\Box_E\phi(a)$ but not $\Box_M\phi(a)$ is true. The case $\Box_M\psi$ true and $\Box_E\psi$ false corresponds to a necessary a posteriori ψ. The solution offered to the epistemological problem is neutral with respect to neo-descriptivism: a priori knowledge will be explained only in so far as some form of descriptivism is true (that is only in so far as some synonymy is given) and it does not presuppose some hidden metasemantic reasoning abilities on behalf of the speakers.

Necessity of identity statements is expressed by 3., the simpler $a = b \rightarrow \Box a = b$ will not do, unless existence is assumed to be necessary. But even though there is someone who is Michael Jackson in the actual world, it seems to be metaphysically possible that Michael Jackson does not exist. But being granted that Michael Jackson is Peter Pan, there is no possible world in which either Michael Jackson or Peter Pan but not both do not exist.

To be fully Kripkean, one needs rigidity; in our setting, rigidity in all metaphysically possible world is a particular case of hyperrigidity, if we assume that every metaphysically possible state of affairs is conceivable. 4. expresses hyperrigidity as a property of conceivable content. If I can conceive something about a, which is the case as soon as a is part of my language, then *ipso facto* I can use a in *de re* attitudes. Under the

reasonable assumption that epistemic alternatives are conceivable, 4. implies $\square_E\phi(a) \vDash \exists x \; x = a \land \square_E\phi(x)$.

As long as 5. is validated, our semantics will be immune to the objection relying on backward reference raised against 2DMS. Nevertheless, one could object that 5. is far from implementing the backward reference principle itself, precisely because we have no actuality operator. This objection is partially misguided, because an operator actually can be at least defined in certain contexts. To express that something is different from what it actually is identical to, that is to express $y \neq \imath x \; A\phi(x)$ one can use $x \neq b$ with b synonymous with $\imath x \; \phi(x)$.

3.3. One-dimensional semantics with metaphysically impossible worlds

In order to solve the representation problem, it is necessary to extend in some sense the range of possibility. Following the possible worlds account of content, for "a=b" to be informative, there must be some possible world in which it is false that "a=b". If the Kripkean thesis of the necessity of identity statements is right, then one has to look beyond the normal possible worlds to find the worlds excluded by "a=b". The solution of 2DMS is very special: no new possible are added, but possible worlds, usually construed as counterfactual alternatives to the actual world, are reconstrued as actual alternatives (this is the gist of the second dimension). This is why the 2DMS solution is both nice and problematic: it is nice, because no new worlds are needed to account the informativeness of necessary true statements, but it is problematic because necessary statements can become false, only if they are reconstrued as statements about different things (semantic interpretation) or as statements in another language (metasemantic interpretation). A more straightforward solution to the representation problem would be to take into account more worlds than just the metaphysically possible worlds and more generally to distinguish *at the worlds level* between metaphysical and epistemic possibilities.

Which new worlds could be added beyond the metaphysically possible ones? Our guiding intuition is that the agent accesses possible worlds via rigid terms like proper names. The solution consists then in adding to the stock of individuals witnesses or counterparts for names, so that what can be imagined thanks to names can be realized in a possible - possibly metaphysically impossible - world. Technically, the trick is to interpret equality not as identity but as a congruence; this is standard model-theoretic practice. [13]

We now spell out the one-dimensional modal semantics with identity (1DMS) we propose as a rival to 2DMS. Let L be a language with

equality, N_L is the set of rigid terms, that is proper names of L and terms of the form $f(a)$ if f is a function expressing some necessary property of a, syn_L a synonymy function for L. A 1DM model for L will be a 8-uplet $\langle W, w_0, R_C, R_M, R_E, D, Q, V \rangle$ such that:

- W is a set (of 'worlds').

- w_0 is a distinguished element of W ('the actual world')

- D is another set (of 'individuals'), containing witnesses for each rigid term in L, that is $D \supset \{wit_a\}_{a \in N_L}$.

- Q is a function from W to D ($Q(w)$ is the set of individuals which 'exist' in w).

- V is a function from $W \times Voc(L)$, such that $V(w, e)$ is made out of $Q(w)$ in a way matching the syntactic type of e (V 'interprets' the non logical basic expressions of L in the different worlds). Two specials constraints concerns rigid terms and the interpretation of equality. We do not require V to be totally defined for names[14], but if $V(w, a)$ is defined, $V(w, a) = wit_a$ and if $wit_a \in Q(w)$, $V(w, a)$ is defined. This will ensures that names are rigid designators.

  Similarly, if $f(a)$ is a rigid term and if $V(w, f(a))$ is defined, $V(w, f(a)) = wit_{f(a)}$, and if $wit_{f(a)} \in Q(w)$, $V(w, f(a))$ is defined.

  As to $' ='$, we require that $V(w, =)$ is a congruence on $\langle Q(w), V(w) \rangle$.

- R_C is a relation on W.

- R_M is a relation on the set of metaphysically possible worlds. A world w is metaphysically possible iff $V(w, =)$ and $V(w_0, =)$ agree on $Q(w_0) \cap Q(w)$, and if $V(w, a)$ is defined and $\langle a, b \rangle \in V(w_0, =)$ then $V(w, b)$ is defined.

- R_E is a relation defined on the set of epistemically possible worlds such that $R_E \subseteq R_C$. A world w is epistemically possible iff if $syn_L(a) = \phi(x)$, then $V(w, a)$ is defined iff $\exists!x\ \phi(x)$ is true in w and if $V(w, a)$ is defined, then $\phi(a)$ is true in w.

It is now routine to check that the semantic requirements hold.

Non-trivialization holds because we allow the equality relation to vary across possible worlds: $\langle a, b \rangle \in V(w_0, =)$ does not force $\langle a, b \rangle \in V(w, =)$ for a $w \neq w_0$. This freedom comes from the fact that we decide to construe $=$ by a congruence relation instead of construing it by the identity relation. Of course this is in a sense cheating, because if in the real world w_0, there is a real object d such that $d = wit_a = wit_b$, there

can be no metaphysically possible world where d and the two witness wit_a and wit_b split. But this is metaphysical cheating, not epistemic: from an epistemic point of view, this is perfectly possible, because the only way for a speaker to track objets named through possible worlds is to use the names, and nothing prevents him from thinking that the objects tracked by a and b *are* different, in which case for example the real object a might split into wit_a and wit_b, sharing out a's properties. Finally, since "$=$" is interpreted by a reflexive relation, there is no problem with $a = a$.

The first part of aprioricity, namely if $syn_L(a) = \phi(x)$, then $\Box_E \phi(a)$, follows directly from the definition of epistemically possible worlds. But it will in general not be the case in metaphysically possible worlds that $\phi(a)$ holds, it will be false in all metaphysically possible worlds which are not epistemically possible, which leaves room for a priori contingent statements. In the same manner, $\Box_M a = b \nvDash \Box_E a = b$ is ok because of the epistemically worlds which are not metaphysically possible. Moreover, as expected, 1DMS is neutral with respect to the truth of neo-descriptivism: the first part of 2. has content only if one accepts descriptive names, but the formalism is perfectly compatible with a language coming with no synonymy relation. That's why we say that 1DMS is more neutral with respect to the choice of a theory of meaning than 2DMS, because 2DMS has to systematically explain how the diagonal proposition is determined, and this is not an easy task on both interpretations of 2DMS.

The rigidity requirement results from the definition of metaphysically possible world.

Hyperrigidity holds because there is a built-in witness for the first existential quantifier of $\exists x \ (x = a \wedge \Box_C (\exists y \ y = a \rightarrow x = a)$ in the model, namely wit_a.

Requirement 5., which corresponds to the backward reference principle, will be satisfied as well, because even though $f(a) = b$ is true in the actual world and $f(a)$ rigidly refers to $wit_{f(a)}$, $f(a) \neq b$ is true in a world w where $(wit_{f(a)}, wit_b)$ is not in $V(w, =)$. This world w is neither metaphysically nor epistemically possible: it is not metaphysically possible, because $V(w, =)$ is substantially different from $V(w_0, =)$ and it is not epistemically possible, because $b = f(a)$ is false in w. This is the desired result, because 7. says that something is conceivable which we know *a priori* to be false, and which, if false, is necessarily false.

Satisfying at once 1., 2. and 4. is distinctive of 1DMS. Let's draw the comparison with other modal semantics.

Usual modal semantics In a classical modal setting, "$=$" is treated as a logical symbol, so that it is interpreted on all models by

identity. Then, either names are treated as rigid expressions or not. If they are, if $a = b$ is true somewhere, it is true everywhere[15], therefore 1. will not hold, that's why there is a representation problem with usual modal logic. If not, 3. and 4. will of course fail. For example, in Hintikka's approach (Hintikka 1971), singular terms are associated with "individual concepts", which are ways of picking out the individuals they refer to in worlds considerer as possible by the agent. Then in general names are not rigid. (1DMS) departs therefore radically from this kind of approach because it insists on the acceptance of 4., that is on the acceptance of rigidity in all kind of intensional contexts. It does not embrace some kind of counterpart theory similar to (Lewis 1968): witnesses and objects in general are taken to be transworld entities, so that we do not need of course to resort to a counterpart relation across worlds.

2DMS Here the problem is with 4.: 2DMS will indeed account for the fact that "Julius is Whitcomb L. Judson" is not a priori. But this will be false in a 'diagonal' pair of worlds (w, w) just because in (w, w) the inventor of the zip is someone else, that is just because Julius is not who he is in the actual world. And this point is highly open to criticism: it licenses Stalnaker's objection to the semantic interpretation of Chalmers, because in (w, w) "Julius" has changed its meaning, and it licenses our objection according to which 2DMS unduly gives up rigidity.

Impossible worlds semantics The appeal to metaphysically impossible worlds has been widely advertised as a solution to various kind of representation problems. For example, Restall (Restall 1997) proposes an ontologically acceptable interpretation of impossible worlds as sets of possible worlds to deal with logical omniscience. In his setting, $a = b \wedge \diamond_E a \neq b$ is indeed satisfiable, because one can mix up individuals as well as one can confuse worlds, the technical trick being to replace in impossible worlds identity with an equivalence relation to interpret equality. The spirit of this solution is similar to ours, but its implementation has unwelcome consequences, in that it leads to too much impossibility for the impossible worlds. In particular, the semantics makes $\Box(a \neq b \rightarrow a \neq a)$ valid, therefore we cannot but have the last part of 1. violated if we want the first part as a solution to the representation problem. We do not wish here to take a stand as to whether a semantics which makes $a \neq a$ satisfiable is needed to solve the full range of representation problems, including logical omniscience. Nevertheless, this seems far from obvious, and anyway it should not arguably come as part and parcel of an explanation of the

semantic and epistemological problems we are dealing with in this paper.

3.4. Objections to metaphysically impossible worlds

Even if a proponent of 2DMS grants that we should be happy with 1DMS satisfying both 1., 2. and 4., he might argue that the nice thing with 2DMS, namely that it does not appeal to new impossible world, is lost, so that our solution comes at the hight cost of assuming the existence of weird entities. If this criticism amounts to the rejection of the sheer idea of impossible worlds, we cannot but acknowledge that this is the price to pay in order to stay one-dimensional. Nevertheless, the price is not so high as it might first appear; the common objections, as well as Stalnaker's own special ones, against impossible worlds as used to solve the problem of logical omniscience do not apply here:[16]

change of meaning The most standard objection is that the meaning of logical expressions is betrayed in the impossible worlds. For example, nothing prevents one from assuming the existence of a world w where p and $\neg p$ are both true, but in this case, $\neg$ simply cannot be interpreted as the classical negation; in particular, elementary logical laws of course do fail in impossible worlds, and it seems to be that satisfying these laws is a definitional property of the expressions involved. The situation of equality in our setting is far less desperate:[17] let's look at what happens to the usual axioms for identity. Reflexivity, symmetry and transitivity are valid, this follows from $V(w,=)$ being an equivalence relation for every w. Things are slightly more complicated for the indiscernability of identicals: since $V(w,=)$ is a congruence for every w, identicals are substituable *salva veritate* in non-modal formulas. But of course, this does not hold for modal formulas, simply because even though $a = b$, modalities in a formula ϕ can lead to a world in which $a = b$ is no more true. In order to have substituatibility, one needs therefore either to restrict the range of formulas to non-modal ones, or to strengthen the identity conditions, by replacing $a = b$ with $\Box a = b$.

ontological profusion Another common objection is that there are simply no such things as impossible worlds, so that 1DMS would depend on unrealistic ontological assumptions.

First note that there should be nothing wrong in itself with interpreting '=' by an equivalence relation instead of identity on models. Assume $w_0 = \langle Q(w_0), V w_0 \rangle$ represents the actual world and

that $V(w_0, =)$ is the identity on $Q(w_0)$. Adopting 1DMS amounts just to looking at w_0 in a slightly different way. Let f_{N_L} be a function on $Q(w_0)$ such that $f_{N_L}(d) = \{wit_b / b \text{ is a name and } V(w_0, b) = d\} \cup d$. We define $w_0' = \langle Q(w_0'), V(w_0') \rangle$ by setting:

- $Q(w_0') = \bigcup f_{N_L}(Q(w_0))$
- if b is a name and it is defined in W_0, $V(w_0', b) = wit_b$
- let P be a n-ary predicate, $\langle d_1, ..., d_n \rangle$ is in $V(w_0', P)$ iff $\langle d_1', ..., d_n' \rangle$ is in $V(w_0, P)$ where d_i' is the $d \in |W_0|$ such that $d_i' \in f_{N_L}(d)$[18]

It is easy to check that $V(w_0', =)$ will be a congruence on w_0' and that w_0 is the the quotient structure induced by $V(w_0', =)$ on w_0'. That's why we can say that W_0' is just another way to look at W_0.

As to the witnesses wit_{b_i} in general, they are just the object the agent projects from the names b_i to represent the possible worlds; one could take them to be the b_i themselves, or some construction out of them.

arbitrariness of the delineation of possibilities In Stalnaker's philosophical dialogue (Stalnaker 1996), Louis, who is loath to resort to impossible worlds addresses the following complaint to his more liberal companion:"By using the world 'possible' to distinguish between the possible worlds, you are ruining the good old world 'possible'. [...] you owe me an explanation of what distinction between the worlds you are using the word 'possible' to make." We do owe Louis the same explanation with respect to 'metaphysically possible' and 'epistemically possible', but this explanation is straightforwardly read off our formal definitions for 1DMS: for a world to be metaphysically possible, it has to abide by the metaphysical principle that identities between names are necessary, for a world to be epistemically possible, it has to take into account the a priori knowledge embodied in the mastery of synonymy relation of the language.

contagion of impossibility There is a follow up to Louis' objection: impossibilities cannot be assumed to be contained to impossible worlds. Suppose there is a world w such that both P and $\neg P$ are true in w, then, so the argument goes, $'in\ w,\ P'$ is actually true, and $'in\ w,\ \neg P'$ as well, but if we accept that $'in\ w,\ \neg P'$ is logically equivalent to $'\neg in\ w,\ P'$, we will have a true contradiction in the actual world as well. But this strategy does not seem to apply here, because there is nothing wrong with $'in\ w_0,\ a = b'$ and $'in\ w,\ a \neq b'$, because one should not accept the equivalence between $'in\ w,\ a \neq b'$ and $\Diamond_M a \neq b$.

Conclusion

Our ability to conceive of possible state of affairs is grounded in our rigid use of referential terms. According to this basic intuition, rigidity is not a special feature of metaphysically possible worlds: some stronger version of it, namely hyperrigidity, should hold in every possible world, be it a mere epistemic possibility or a mere conceivable state of affairs. As a consequence, we take it that our attitudes towards individuals referred to by proper names are always *de re*: even when I consider as possible that Cicero might not be Marcus Tullius, it is something that I consider as possible of Cicero and of Marcus Tullius. This approach differs radically from the usual ways of making compatible Kripke's theses and the possible world analysis of content. Typically, 2DMS represents the belief that Cicero is not Marcus Tullius by resorting to a possible world in which at least one of the two names has dropped its usual reference. We have argued that this was not satisfactory and that one should be Kripkean all the way long. The semantics we sketched, 1DMS, shows that it is indeed possible, if one grants that conceivable state of affairs that are not metaphysically possible should indeed be modeled by impossible worlds.

Notes

[1] This way of seeing things should not be attributed only to explicit advocates of 2DMS like Chalmers. Contemporary proponents of Hintikkean analysis of belief often assume implicitly the possibility of such a reconciliation of the two theories (See e.g. Aloni M., 2005).

[2] A thorough presentation and defense of these claims, once held by Stalnaker in (Stalnaker 1978), can be found in (Chalmers 2004).

[3] These principles are primarily formulated as conversational principles governing assertions made in certain contexts, but they are suppose to apply also to belief ascriptions.

[4] The context set expresses the background knowledge presupposed at the current stage of the conversation.

[5] "But being put in a situation where we have exactly the same evidence, qualitatively speaking, it could have turned out that Hesperus was not Phosphorus; that is, in a counterfactual world in which 'Hesperus' and 'Phosphorus' were not used in the way that we use them, as names of this planet, but as names of some other objects, one could have had qualitatively identical evidence and concluded that 'Hesperus' and 'Phosphorus' named two different objects (Kripke 1980: 104).

[6] Stephen Yablo, in "Illusions of Possibility" underlines ways in which actuality can figure in the content of modal judgements. In particular he says: "The thing that seems possible - the condition that seems like it could have obtained - can have the concept of actuality in it. This is in fact quite common. People say, for example, "It's cold in here but it could have been colder". Colder than what? Colder than it was actually, of course." (Yablo 200?).

[7] That is, for all w_1, x, y the interpretation of f in (w_1, x) and (w_1, y) is the same

[8] Just because all A operators points to the same world, and $f(x)$ in (w,w') is the same as $f(x)$ in (w,w) by rigidity of f, so that $f(x)$ and $\imath y A y = f(x)$ are everywhere the same.

[9] Technically, this would yield a free logic with an outer domain semantics, objects in the outer domain being those witnesses such that every objects they are identical to is a witness.

[10] Of course, we do not wish to collapse conceivability onto the mere possibility of stating a grammatical sentence. It will be clear that our semantics does not make room for unconceivable states of affairs like one in which Cicero would not be identical to Cicero.

[11] But of course, it can be simply expressed as some kind of knowledge, for example the common knowledge that $Julius = \imath x \ \phi(x)$.

[12] To be precise, all non logical expressions below are variables (standing for constants, definite descriptions or formulas). Hence a schema is valid iff all its instances are valid and it is satisfiable iff one of its instance is satisfiable, and similarly for logical consequence.

[13] There is no claim of technical originality in this paper; but though technically rather trivial, we think the kind of 1DMS we propose should be regarded as a coherent alternative, and maybe the more natural one, as soon as 2DMS comes under scrutiny.

[14] This will imply a further semantic choice as to the treatment of empty singular terms, what this choice will be does not matter here.

[15] With the same proviso as before that a and b exist.

[16] We do not even consider here attempts like Restall's (Restall 1997) to bypass these objections in the larger setting of the problem of logical omniscience.

[17] This is in contradistinction with more 'brutal' treatments of '=' like the one in (Restall 1997).

[18] Existence and uniqueness of d follow trivially from our definition of f_{N_L}.

References

Aloni M.: 2005, Individual Concepts in Modal Predicate Logic,*Journal of Philosophical Logic* **Vol.34 no.1**, pp. 1–64.

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

Chalmers D.J.: 2004, Epistemic two-dimensional semantics, *Philosophical Studies* **Vol.118**, pp. 153–226.

Davies M. and Humberstone L.: 1980, Two notions of necessity, *Philosophical Studies*, **Vol.38**, pp. 1–30.

Evans G.: 1977, Reference and contingency, *The Monist* **Vol.62**, pp. 161–189.

Hintikka J.: 1971, Semantics for propositional attitudes. In Linsky, editor, *Reference and Modality*, Oxford University Press, Oxford, pp. 145–167.

Jackson F.: 1998, *From Metaphysics to Ethics: A Defense of Conceptual Analysis*, Oxford University Press, Oxford.

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

Lewis D.: 1968, Counterpart Theory and Quantified Modal Logic, *Journal of Philosophy* **Vol.65**, pp. 113–126.

28

Restall G.: 1997, Ways Things Cant Be, *Notre Dame Journal of Formal Logic* **Vol.38**, pp. 583-596.

Stalnaker R.: 1978, Assertion, repr. in Stalnaker R., *Context and Content*, pp. 78–95.

Stalnaker R.: 1987, Semantics for Belief, repr. in Stalnaker R., *Context and Content*, pp. 117–129.

Stalnaker R.: 1996, Impossibilities, repr. in Stalnaker R., *Context and Content*, pp. 55–68.

Stalnaker R.: 1999, *Context and Content*, Oxford University Press, Oxford.

Stalnaker R.: 2001, On considering a Possible World as Actual, repr. in Stalnaker R. *Ways a world might be*, pp. 188–200.

Stalnaker R.: 2003, *Ways a world might be*, Oxford University Press, Oxford.

Yablo S.: 200?, Illusions of Possibility, to appear in a volume on 2-D semantics, edited by M. Garcia-Carpintero, Oxford University Press, Oxford.