HOW TO DO THINGS WITH WORLDS: INTENTIONALITY AND THE ONTOLOGY OF MODEL-THEORETIC SEMANTICS

Roger VERGAUWEN

1. *Introduction: intentionality and model-theory*

The relation between logic, philosophy of language, and epistemology has been given content in a new way by the use of *model-theoretic methods* in the study of natural languages and their semantics. The basic goal of model-theoretic semantics is "to specifiy how language connects with the world- in other words to explicate the inherent 'aboutness' of language" (Dowty '81,5). Starting from this assumption, I want first to specify the nature of this relation between language and the world by investigating the concept of *intentionality*. Second, in a discussion of the semantics of de-dicto belief-sentences, it will be shown how a further clarification of intentionality affects the overall structure of the ontology of the model-theoretic semantics in general.

In "Minds, Brains, and Programs" John Searle defines 'intentionality' as "that feature of certain mental states by which they are directed at or *about* objects and states of affairs in the world" (Searle '80,424). Examples of intentional states are 'beliefs', 'intentions', or 'desires'. Searle wants to show that intentionality or 'aboutness', is a product of the causal features of the brain and cannot be programmed into a computer. He sets up a thought-experiment, the 'Chinese-Room-Experiment', to show that a computer can only handle meaningless symbols, and his conclusion is that only human brains can think and produce intentionality. In short, the computer has *only a syntax but no semantics*. The machine cannot properly be said to 'understand' anything of what it is doing, and neither does it understand the language it is using. The issue of understanding is a very central one in semantics and epistemology. According to Searle, the most important difference between machines and human beings is that we know what we are talking about when we are using language. This use is first and

foremost intentional. Linguistic behaviour is fundamentally characterized by its 'aboutness'. I am, however, convinced that the intentional character of language can be accounted for in a formal theory of meaning, such as a model-theoretic semantics. I think that in the case of language this fairly mysterious biological power of the brain may have a logical counterpart. This counterpart will be able to account for the 'aboutness' of language, but it will not need a biological foundation.

In a model-theoretic semantics for natural languages, such as the one developed by the late Richard Montague, a *model* for a language is a quintuple ($\mathfrak{A}$)

(1) $\mathfrak{A}$ = (A,I,J, $\leq$,F) (Montague '73,258)

This model consist of a set of possible individuals (A), a set of possible worlds (I), a set of moments of time (J), an ordering relation ($\leq$), and a function (F) that assigns 'meanings' to the non-logical constants, the 'words', of the language. In this model-theoretic *ontology* it is the function F that will be of interest. The meanings assigned by F are logical types drawn from a higher-order intensional logic that functions as a semantical metalanguage to the natural language under consideration. Though the basic goal of model-theoretic semantics is to specify how language connects with the world, it should be clear that the model as it stands does not account for the *production of intentionality* (or 'aboutness') in language. The meaning-assignment function F merely states that things in the world are to be connected with logical individuals of a certain type in the intensional logic, but, as such, *F presupposes intentionality* – the ability of language and linguistic behaviour to be about something – without providing an explanation for it. In that sense, the model-theoretic semantics is, paradoxically speaking, only a syntax.

2. *Intensions and the grelling-paradox*

To find out more about intentionality, it will be necessary to have a closer look at the concept of meaning used in the model-theoretic approach. 'Meaning' in model-theoretic semantics is, roughly, equated with *intension-with-an-s*. An intension is a function from possible

worlds and moments of time to *extensions* or denotations of the appropriate kind. The intension of a sentence is a function from possible worlds and times to the truth-value of that sentence in these worlds. The intension of a common noun such as 'cat' is a function from possible worlds and times to the set of cats existing in these respective worlds. The extension of 'cat' is a subset of the set of individuals, and is defined by its *characteristic function*. This function belongs to the set of functions from individuals to truth-values and, for a given set of individuals, it gives the value '1' or '0' according as to whether an individual belongs to that set or does not belong to it. In our example this means that speaking about the set of cats in its extension is the same as speaking about the characteristic function of a certain set.

In this theory sameness of intension implies sameness of 'meaning', and so one would expect words with equal intensions to be about the same thing, at least if intentionality would be expressed by what is given in the intensions. This is, however, not so and it could easily be demonstrated by the failure of substitutivity salva veritate of intensionally equal words, such as 'eye-doctor' and 'oculist', in contexts of verbs of propositional attitude ('believe', 'think'...), and the loss of rigid designation in these contexts. Sameness of meaning as sameness of intension would also imply that all tautologies or all contradictions have the same meaning, which is clearly intolerable. I will come back to these problems later on, but all this adds up to show that whatever intentionality is, it is clearly not the same as 'intension'. As will soon become clear there is, however, a link between intensions and intentionality. Intensions are very useful as a first approach to intentionality. They are functions from which the extensions are computed by calculating the value of the function for a possible world as its argument. In my view, this definition of 'intension' is seriously impaired by the existence of a well-known paradox – *the Grelling Paradox* – which can be explained in terms of english adjectives in the following way: an adjective is said to be *autological* if and only if it can truly be applied to itself, e.g. 'short', 'polysyllabic'. Otherwise it is called *heterological*, e.g. 'red', 'monosyllabic'. However, if we want to know whether the adjective 'heterological' itself is heterological or autological we get a paradox. For, if we assume that 'heterological' is autological, then, by definition, it applies to itself. But if

'heterological' applies to itself, it must (by the definition of heterological) be heterological. If, on the other hand, we assume that 'heterological' is heterological, this obviously means that it applies to itself and is therefore autological: therefore, the answer as to where the adjective 'heterological' belongs in the autological-heterological distinction leads to a paradox.

In an article on *incomputability* Hoare and Allison describe this paradox by means of functions and their evaluations by a computer (Hoare and Allison '72, 171 ff.). In the programming language (e.g. LISP) the definition of 'heterological' becomes

(2) Heterological (p) = ¬(p(p)).

In other words: to find out whether an argument (word) 'p' is 'heterological' it must be applied to itself and the answer must be negated. If we want to examine whether this function with itself as an argument – Heterological (heterological) – is true or false, the computer will try to compute '¬(Heterological (heterological))'.

Of course, this will mean that the computer will have to evaluate first the expression between brackets, wich is by definition '¬(Heterological (heterological)'. It is easily seen that the evaluation of 'Heterological' with itself as an argument leads to an infinite, *non-terminating computation*.

In terms of the model-theoretic semantics we could say that the heterological-function as defined in (2) is a *representation of the intension* of the word 'Heterological'. This intension is a function from possible worlds and moments of time to a set of adjectives. From what has been said it follows that the set of adjectives in the extension of 'Heterological' cannot be represented by a characteristic function, as it is *in general* impossible to determine for every adjective whether it is in the set or not. The non-termination phenomenon also seriously impairs the use of the intension-function, because from it the extensions should unambiguously be determined. And this is of course impossible as matters stand now. As the computation is non-terminating, the computer will never give an answer to the original question, as to whether the adjective 'Heterological' belongs to its own extension.

But we might think that this defect can perhaps be remedied. For a function such as the *factorial function* the computation either termi-

nates or, for some argument will not terminate in principle, e.g. 'fac.(-1)', or there would be no answer because it exceeds the integer limit of the machine, though in principle there would be a definite one. An example of this could be 'fac. (98076555431)'. In this case the computer may fail to give the answer, since the numbers involved are too large. This is of course a kind of non-termination, but not in principle, given the possibility to envisage advances in computer technology that would put this calculation within the bounds of practicality. Moreover, it is not always clear that a function that has just been programmed does not terminate (in principle) or will terminate after a few more seconds of computation time. In the heterological-example in (2) the function itself does not express its own non-termination, but in this case it is easy for us to dicover that it is non-terminating.

This suggests that it may be a good idea to try to write a program that will answer the question whether a given function will not terminate or will in principle terminate. Call the function that will determine whether a program terminates '*terminates (f,x)*', and suppose that it gives the answer *true* if the computation f(x) will (in theory) terminate and the answer *false* if it will never terminate. The function 'terminates' itself should be designed always to terminate, otherwise its general usefulness would be seriously impaired. Then, we would have 'terminates fac(-1)' = *false* and 'terminates fac. (9807655543)' = *true*. A solution to the problem of heterologicality could then be offered in the following way, by using the 'terminates' – function to reprogram this problem. A *function* is said to be *autological* if it terminates when applied to itself and delivers the value 'true' and *heterological* otherwise. If it does not terminate when applied to itself it is always heterological. If it does terminate when applied to itself it will be heterological if and only if it delivers the value *false*. In other words:

(3) Hetero (f) = *if* terminates (f,f) *then* ¬f(f) *else true* (Hoare and Allison '72, 173).

It turns out, however, that for a computation of 'hetero' applied to itself – hetero(hetero) – this leads to a contradiction. The conclusion is that it is impossible to program a function *t*, which takes two arguments *f* and *x*, and which terminates for all arguments f and x, and

furthermore takes the value *true* when the application of f to x is a terminating computation and *false* when the computation is non-terminating. This *impossibility result* has important consequences for the model-theoretic semantics. The attempt to define the 'hetero'-function with the help of the 'terminates'-function comes down to trying to adjust or define the intension of 'heterological' in such a way that it *unambiguously* determines a set of adjectives by constructing *a function over intension-functions that will determine the nature of the extension-function*. If it would work, the 'terminates'-function would restore the extension as a set defined by a characteristic function. In the model-theoretical approach this 'terminates'-function can be viewed as the function *F* from the model in (1). If F could be computed it would give an answer to the question whether the computation performed by one of its arguments, the intension-functions, terminates or will in principle never terminate. But unfortunately this seems to be impossible.

Summarizing, we can now say that the *meaning-assignment function F* cannot be constructed so as to fully determine the intensions of the non-logical constants of the language. A complete specification of what these constants (the 'words') are *about*, their *intentionality*, is therefore impossible. The *assignment of a 'meaning' to the 'meanings'* (intensions) cannot be effected through the model, because this would imply the possibility of calculating the 'terminates'-function. Viewing *intensions as programs*, to which I cannot see any obstacles, together with the *self-reference* involved in the computation of extensions from intensions, are the main reasons for this defect.

3. *The incompleteness of semantics: intentionality as a non-standard concept*

In *church's theorem as* explained by Kleene (Kleene '67, 246), uncomputable functions are linked to undecidable predicates. A function is *computable* is there is a (Turing)-machine, T, with index i which, when applied to an argument *a*, will at moment x have computed a value for *a*. For a certain function this machine may be represented by the predicate '$(\exists x)$ T(i,a,x)'. This predicate is decidable if the function which the machine mimics in its computation is computable. However, the formula representing a machine to which

the machine's own index y, considered as the integer which describes the machine table and thence the pattern of behaviour of the machine, is presented as an argument – '$(\exists x)\ T(y,y,x)$' – is undecidable because the corresponding function is uncomputable. The existence of an undecidable, though true, number-theoretic formula is the main idea in *Gödel's incompleteness theorem(s)*. The intuitions to be presented here on the *incompleteness of the model-theoretic semantics* closely parallel the ones in the first Gödel-theorem. I wil be using Gödel's intuitions here mainly as a heuristical device, but there is a clear sense in which the Gödel-theorems apply in the theory of meaning as part of an investigation into the language of mathematics. The incompleteness of semantics can, however, also be motivated *semantically*. At the end of part 2 it was concluded that intentionality cannot be accounted for in the model. The closest we could get to what intentionality within a model-theoretic semantics could be, was the '*terminates*'-function, which is a function over the intension-functions. Unfortunately, this meaning-assignment function, F, could not be computed. However, within the logic we can construct a formula stating that within the model there is no intentionality, viewed as a certain property of the intensions of words. Let this formula be represented informally as:

(4) $\neg$(Meaning (Meaning)).

Formula (4) says that there is not such a thing as 'the meaning of meaning'. In other words: if we would want to construct a machine which, for every linguistic unit would be able to determine the meaning of its intension, then (4) states that such a 'meaning-machine' cannot be constructed, as it would not be able to compute a value for all its arguments. More specifically, formula (4) can be shown to be of the form:

(5) $\neg(\exists x)\ T(y,y,x)$.

In (5), it is easy to see that a kind of self-reference is involved, because in a sense the machine is asked to compute a value with itself as an argument. One might think here of the problems arising with intensions as a consequence of the Grelling-paradox. Formula (4) expresses a true property of word-intensions, but it cannot be validly deduced or 'proved' within the semantics. Loosely speaking, we have

here a typical 'Gödel-situation'. Assuming 'provability' to be an analogue of intentionality within our semantics, which we in fact did in the previous part because it is the only thing we have at our disposal within the metatheory of the intensional logic used, we now ask whether we would be able to prove the formula represented informally by (4), if it is true. If we could prove it, then, by the identification of provability and intentionality, we would establish that this formula 'has' intentionality'. But by what it says about itself as represented in (5), we know that it cannot be proven, or cannot be a representation of what intentionality is. On the other hand, if we would be able to prove its negation, (4) would be false and provable within our semantics, which is intolerable if our semantics is to be *consistent* (otherwise, of course, 'anything goes'). Moreover, proving the negation of (4) would also imply the possibility of calculating the 'terminates'-function and a decision procedure for the formula '$(\exists x)\ T(y,y,x)$'. For all these reasons, there is a formula in the semantics such that neither it nor its negation is provable, though it is true. Such a formula is called *undecidable* and its very existence causes the semantics to be *incomplete* (if it is consistent). The undecidable formula is the one in which the absence of intentionality is stated. But then we might perhaps be forced to conclude that, after all, there is no possibility of representing intentionality within the model-theoretic semantics. This is, however, not so, and the possibility of representing intentionality hinges on the undecidability of the formula representing the absence of intentionality within the model.

The incompleteness of semantics points to another of its characteristics which is its *non-categoricalness*. A theory is said to be categorical if all its models are isomorphic. The undecidable formula that was given by (4) cannot be validly deduced from the formulas and construction principles of the semantics. By the definition of logical validity we would then have the existence of a model in which the formulas of our semantical metalanguage are true but that (undecidable) formula is false. But we know that (4) is true, and therefore the model in which it is false is not isomorphic to the one in which it is true. Such models are called '*non-standard*' or '*non-intended*' and their existence is the reason why the semantics is non-categorical. Put still another way: it can be proved within model-theory that if Σ is a (consistent) set of sentences in a language (L) and σ is a sentence of

this language, then is σ cannot be proven or deduced from Σ, the set Σ united with the negation of σ, (¬σ), is *consistent*. So, we have:

(6a) if $\Sigma \nvdash \sigma$ then $\Sigma \cup \{\neg\sigma\}$ is consistent.

Furthermore, the following can also be shown:

(6b) Every consistent set of sentences has a model.

In the case of the model-theoretic semantics this means that it is possible to add the *negation* of the formula expressed by (4) to the system, and to find a model or interpretation for it (from 6a and 6b). But this means that in the non-standard model so constructed the negation of (4) is provable. By what (4) says, in the non-standard model the *absence of intentionality is denied*. What is considered to be 'intentionality' in the standard model cannot be proved within that model, and therefore intentionality is a *non-standard concept*. It can only be talked about in a model that is non-standard. The semantics is non-categorical, and rightly so. The intended model for a semantics of natural language can be viewed as *the world* in any good sense of the word. But, of course, it is the world as *qualified by a certain definition of intentionality* in which the nature of the correspondence relation between the language and the world is fixed. A non-intended model in this sense is defined by a different specification of intentionality by means of *intentionality-creating verbs or intentionality-predicates*. Basically, they are means of redifining the usual relation of intentionality, thus creating models of a different kind. I will come to this in the next part. The specification of the relation between language and the world (the 'aboutness') can be effected in many ways, but as such this specification never belongs completely to the model under consideration for which it is defined, because it is represented by an undecidable formula such as the one in (4). A specification can, however, be given in a non-standard model. By doing this, we get a kind of 'new' semantics but here also there will be an undecidable, but true, formula, expressing that for this new system there is no intentionality. Intentionality in this system is to be accounted for by the technique described in 6a-b and is not directly definable within it. Coming back now to what has been said about intentionality in the first part of this paper, I think that the 'aboutness' of language, its intentionality, is *simulated* by the incompleteness and the non-catego-

ricalness of the semantics. In our attempt to make it a model of the world, it becomes clear that there is always *something more* that the theory does not fully grasp. In short, if the words and the world would be in a relation of strict isomorphy, words could in a sense not even be said to be *about* anything at all. It is the *logical gap* between both that acts as an analogue for the intentionality of language. More specifically, it is the meaning-relation *itself* that always remains only partially defined, depending on the model under consideration. Intentionality is, then, a property intensions have *by their being framed into a model*. The 'aboutness' of the intensions can only be fixed in a relative way because it is tied to a hierarchy of models. Nevertheless, it is possible to show how this intentionality is an effect of the incompleteness and non-categoricalness of the semantics as a logical theory. In the next part, I will try to show how this general account is applied to the semantics of verbs of propositional attitude, and what its effects are on the ontology of the model presented in (1).

4. *Propositional attitudes, supervaluations, and diagonal functions*

In the introduction it has already been pointed out that the notion of 'belief' can be considered as an example of an intentional state. The semantics of the linguistic expression of this state by means of verbs of propositional attitude has been troubling linguists and philosophers alike for many years. In this part it will be shown how the problems arising in fixing the *semantic objects* of these verbs may lead to a further clarification of a logical concept of intentionality and how this forces upon us a change in the ontology of the model.

In a model-theoretic semantics of the Montague-type sameness of intension implies sameness of meaning and so one would expect words or sentences with the same intensions to be 'about' the same things. From what has been said before, it should be clear that intensions are not enough if one wants to account for intentionality. This becomes particularly clear with verbs of propositional attitude. 'Traditionally', in the typed intensional logic of Montague ('70, '73), these verbs are represented by relations between propositions and (sets of) individual concepts, functions from possible worlds and times to individuals. Therefore, they are assigned the type ((s,t), ((s,e),t)) in

the intensional logic. In this formula *s* indicates the intension of the type immediately following it, and *e* and *t* are the types for individuals and truth-values. Not much is said about the nature of the several intension-functions and the differences between them, but it is possible to give a *formal criterion for intensional equivalence*. This equivalence can be expressed by the following formula:

(7) $\Box(^{\vee}f_{s,a} \longleftrightarrow {}^{\vee}g_{s,a}) \longleftrightarrow (f \longleftrightarrow g)$ (Anderson '84, 363).

This formula says that if two intension-functions $f_{s,a}$ and $g_{s,a}$ have the same extensions ($^{\vee}$) in every possible world ($\Box$), then they are the same function. Any two elements with the same extension in every possible world do therefore have the same 'meaning' *within* the semantics and are called *logically equivalent*. As it stands, the criterion presented in (7) cannot be maintained because it breaks down in contexts with verbs of propositional attitude in their *de-dicto reading*.

Two main problems crop up with these verbs. For an illustration of a first set of problems, consider sentences such as (8) and (9).

(8) Thales of Miletus believed that *two plus two equals four*.
(9) Thales of Miletus believed that *the square root of two is irrational*.

It is clear that the complement sentences in these examples have the same extension in every possible world. They are always true as mathematical tautologies. By the criterion in (7) they should be represented by the *same proposition*. Because they therefore would also have the same meaning, these sentences should be *interchangeable salva veritate*. This is, however, not true. The inference from (8) to (9) is in general impossible with verbs of propositional attitude. It is not because Thales believed that 'two plus two equals four', that he would also believe everything that is logically equivalent to it. Indeed, in the examples given, the first is true while the second is not. Problems such as the ones described are called *logical equivalence problems* and they show, among other things, that intensions are not enough if one wants to account for the intentionality of language. What, by the identity of intension is predicted to be 'the same' certainly is not '*about* the same thing'. *David Lewis* (Lewis '76) has at one time tried to solve this problem by taking advantage of the Fregean principle of *compositionality*. He tries to develop a technical

concept of 'meaning' that makes it possible to distinguish the intensions of the complement sentences in (8-9), by distinguishing words and their intensions within the propositions. These 'structured propositions' allow to a certain extent to account for sentences such as (8) or (9), but they do not allow to account for the impossibility of substituting salva veritate in propositional attitude contexts sentences with the same structure containing synonymous (intensionally equivalent) words. As a specification of this second set of problems there is the *loss of rigid designation* in these contexts. The following examples are relevant:

(10a) Eye-doctors are eye-doctors
(10b) Eye-doctors are oculists.
(11a) Hesperus is Hesperus
(11b) Hesperus is Phosphorus.

If we would use (10a-b) in the context of a sentence such as:

(12) John *believes that...*

it does not follow from the intensional identity of 'eye-doctor' and 'ouclist' that John would believe (10b) if he would believe (10a). The loss of rigid designation is illustrated by (11a-b), when combined with a sentence as

(13) The *ancient Greeks believed that...*

From examples such as the ones given, *Barbara Partee* (Partee '79, 7) concludes that we are forced to recognize the *importance of psychological factors in the model-theoretic semantics*. The differences in truth-value in belief-sentences would, on this view, be due to the fact that in the speaker's mind there is a psychological difference between eventually synonymous words. In the example (10a-b) it is very well possible that John simply does not know the word 'oculist' or that in his own *idiosyncratic interpretation* this word is linked to a different concept. He might think that 'oculists' are members of a bizarre religious sect worshipping the god *Ocul*. In the same vein, it may be true that Hesperus and Phosphorus rigidly designate the same *planet* but to the ancient Greeks, who did not know this, both names were not psychologically equivalent. In an article in this journal I tried to construct a *logical counterpart* for what Partee considers to be

psychological factors, in such a way that they could be fully accounted for within the model-theoretic semantics itself (Vergauwen '84). This approach makes use of the notion of a *supervaluation*, developed by Van Fraassen (Van Fraassen '66). He defines a *valuation* as a function that assigns truth-values to all sentences of the language, and distinguishes *two kinds of valuations over models*. A valuation is called a *classical valuation over a model* if the following conditions are fulfilled: it assigns the value 'true' or 'false' to the simple atomic statements containing no non-referring names such as 'the King of France' or 'Pegasus' in the usual manner, and for complex sentences truth or falsehood is defined compositionally by the values assigned to the less complex sentences. A *supervaluation* is an assignment of truth values according to which certain propositions are assigned 'classical' truth-values ('true' or 'false') and the remaining propositions are assigned 'true', 'false', or a '*truth-value gap*' indicated by a dash (-), on the basis of what the classical truth tables plus the given partial assignment of 'true' and 'false' forces on one. Specifically, one considers the set of all classical valuations that assign to the given proposition the agreed upon values; for any other proposition, if those valuations all assign it the same value, we assign it that value, but if those valuations do not all assign it the same value, because some of them make it true and others make it false, we assign it a *truth-value gap*. The idea behind all this is the following. Van Fraassen tries to explore the consequences for formal semantics and logic of the fact that in some circumstances syntactically well-formed sentences may be neither true nor false. Specifically, they are sentences containing non-referring names, such as the ones mentioned. Sentences containing such names are evaluated differently in different classical valuations and are therefore assigned a truth-value gap in the supervaluation constructed over these valuations, and hence these sentences are neither true nor false in the model under consideration.

Coming back now to the problems noted with verbs of propositional attitude, let us make it clear that linguistic communication is largely dependent on an already present *common basis of knowledge and belief*. For people to communicate effectively it is necessary that there be a large part of common knowledge, *including knowledge about the meaning of most words*. It is highly probable that within a linguistic community the 'knowledge-systems' and even the 'belief-systems' of

the language-users run parallel to a certain extent: "If someone's whole belief-structure, including beliefs about what many words mean, differs radically from mine, I cannot hope to describe his or her belief in my language. A compositional semantics may be impossible without the assumption of an homogeneous interpretation system (both the model theory and the interpretation into it)" (Partee '79,8).

In model-theoretic semantics common nouns such as 'eye-doctor' or 'oculist' are represented as predicates over individuals (or 'individual concepts' for that matter). These sets of individuals are assigned to those words within the model as their extensions in the possible worlds. Now, one would expect the extensions of 'eye-doctor' and 'oculist' not to be disjoint, because they are synonymous words. Let this situation be represented by the *classical or general valuations* (v_g) over the model. On the other hand, it can never be excluded that someone's belief-structure contains for some words or sentences an interpretation that is different from 'the usual denotations' of these elements. Or, there may even be no denotation at all, if it is not known what the referent is. That is why it is necessary to add to the general valuations another type of valuation, called *idiosyncratic valuations* (v_i). These are supposed to represent whatever assignments of extensions are different from the general valuations in certain individuals.

For a predicate such as 'eye-doctor' this may be visualized by the (hypothetical) 'meaning diagram' (14).

(14) *eye-doctor (x)*

	v_{g_1}	v_{g_2}	.....	v_i	v_s
0_1	1	1		1	1
0_2	1	1		1	1
.	.	.		.	.
.	.	.		.	.
.	.	.		.	.
0_n	0	0		0	0

In (14), the sets of objects that are assigned in the different possible worlds to the predicate are indicated by $(0_1 \ldots 0_n)$. If the predicate is

true of these sets we assign it the value 'true' (1) and 'false' (0) otherwise, as would be the case if there are no 'eye-doctors' in a particular world. The v_i-values of an element need not be different from the v_g-values, as is clear from (14). We could, now, set up a diagram for a predicate such as 'oculist (x)'. The coextensiveness of 'oculist' and 'eye-doctor' would be guaranteed by the identity of assignments in the v_g-values. Supposing, however, that there would be a difference in the v_i-value of 'oculist (x)' in comparison to the v_i-value of 'eye-doctor (x)' would lead by the definition of a supervaluation to v_s-values represented by a truth-value gap ('–'). Problems of logical equivalence and loss of rigid designation with verbs of propositional attitude could then be solved in the following way. I proposed that with these verbs the *supervaluated meanings* (the v_s-values) of the linguistic expressions should be taken into consideration. Sentences (10a-b) can be formalized by means of an *implication*. The truth-table of this connective in a logic using supervaluations is given by (15).

(15) (P ⟶ Q)

P	T	F	–
Q			
T	T	F	–
F	T	T	T
–	T	–	T/–

(Mc Cawley '81, 245)

Now it is possible to calculate the value of (10a-b) in contexts such as (12). The inference from (10a) to (10b), when combined with a sentence as 'John *believes that*...', can be represented as the value of the supervaluated implication between (10a) and (10b). It turns out that this inference ends up in a truth-value gap (Vergauwen '84, 58 f).

It is the very existence of these truth-value gaps that blocks the substitutability of (10b) for (10a) with verbs of propositional attitude. The reason for this is that there is a difference in v_s-value between (10a) and (10b). For the loss of rigid designation an analogous argument could be given, though I will not do that here. Meanwhile, it is not yet clear *how all this is related to the possibility of producing intentionality within the model-theory* due to the incompleteness of

semantics and the existence of non-standard models, and how this forces us to *change the ontology of the model*. In my opinion, the function described by verbs of propositional attitude as 'believe' is a very peculiar one, comparable to the 'terminates'-function from part 2. It is comparable in the sense that it is *not effectively calculable*. In 2, the function F, which was to assign intensions to the words, was formulated as a 'terminates'-function and was supposed to supply us with an assignment of intentions as a final way of connecting the words to the objects in the world. As such, this function was in a way not strictly a part of the semantics itself, unless one would consider it to be a representation of the *truth-predicate* for the language. This can very well be defended, but one should also take into account that this predicate is both a word of the (natural) language, and as such it may be a means of expressing the aboutness of the linguistic elements (because 'to be true *of* something', is a way to say that something is 'about' something else) and an expression used in the model-theoretic interpretation of the semantical metalanguage, i.e. the intensional logic.

The type of function I want now to consider seems to operate fully within the semantics, as it is linked to the class of verbs of propositional attitude. The principle upon which it rests can be found in Cantor's idea of a *diagonal function*. To illustrate this idea, let us imagine the set of all one-place number-theoretic functions over the integers. We shall say that a function is *effectively calculable* if there exists a definite algorithm that enables us to compute the functional value corresponding to any given value of its arguments. Let us assume that such an algorithm can be expressed as a set of instructions and imagine all such sets ordered in a certain way (Davis '73, xvii). With each positive integer i, there is associated in the list of instructions, the *i-th* set of instructions in this list, which we will call E_i. The function associated in this way with E_i, we will call $f_i(x)$, and E_i tells us how to compute the values of some function. (Davis, ibid.).

Now, let us define the following number-theoretic function:

(16) $g(x) = f_x(x) + 1$

This function g(x) is a perfectly good one. Its value for a given integer, x, is obtained by finding *the x-th set of instructions* E_x, then

applying it to the number x as an argument, and finally increasing this result by 1. Now we have:

(17) For no value of i is it the case that $g(x) = f_i(x)$.

This means that however good the function g(x) is, it can never be in the list of instructions E_i. Because, suppose that $g(x) = f_{i_o}(x)$ for some integer i_o. Then, by (16)

(18) $f_{i_o}(x) = f_x(x) + 1$.

(18) would hold for all values of x. In particular, this equation would have to hold for $x = i_o$, yielding:

(19) $f_{i_o}(i_o) = f_{i_o}(i_o) + 1$

But of course this is a contradiction, as no number equals itself plus one. Now, from the manner of choice of the E_i, the functions $f_i(x)$ were to include *all* effectively calculable functions, which yields that *g(x) is not effectively calculable*. It is easily seen that in this 'diagonal method' there is an element of *self-reference*, which is one of the reasons for the function's not being effectively calculable. Let us now apply this method to the model-theoretic framework of interpretation in the analysis of language. Suppose we would tabulate the meanings of the words of a language as the sequences of values of these words in the different possible worlds (PW) of the model or interpretation. The words themselves are identified with functions *f(x)* and numbered successively so as to make it clear that different words represent different functions. So we get $f_1(x)$, $f_2(x)$ and so on. The extensions of the words are determined in the possible worlds according to the valuations V_n over the model. The extension of a word is always to be the value of the intension-function for a possible world and a valuation as its arguments. As will soon become clear in the treatment of verbs of propositional attitude, *we have to include among the possible extensions something like the 'meaning' of the word itself*. We want to be able to talk about the 'meaning' of the word *as a separate entity*. Therefore, for every word there has to be included a possible world for which the extension-value of that word is its meaning. In the formalism, this is effected by including as part of the interpretation of the words a *representation of the set of its denotations (extensions) in the model*. This implies the possibility of *quantifying over the model*

itself, which is not possible in Montague's ontology as given in (1). In his model-theoretic approach quantification over sets of possible worlds that make up a model is not allowed, as there are no variables over possible worlds as such. The following diagram may help to visualize this situation thus far:

(20)	PW_0/V_0	PW_1/V_1	PW_2/V_2 ..... PW_{n-2}/V_i	PW_{n-1}/V_s	PW_n/V_n
words					
$f_0(x)$	$f_0(PW_0/V_0)$*	$f_0(PW_1/V_1)$	$f_0(PW_2/V_2)$ ..... $f_0(PW_{n-2}/V_i)$	$f_0(PW_{n-1}/V_s)$	$f_0(PW_n/V_n)$
$f_1(x)$	$f_1(PW_0/V_0)$	$f_1(PW_1/V_1)$*	$f_1(PW_2/V_2)$ ..... $f_1(PW_{n-2}/V_i)$	$f_1(PW_{n-1}/V_s)$	$f_1(PW_n/V_n)$
$f_2(x)$	----------	----------	$f_2(PW_2/V_2)$* ..... ----------	----------	----------
.					
.					
.					
$f_{n-2}(x)$	----------	----------	---------- ..... $f_{n-2}(PW_{n-2}/V_i)$*	----------	----------
$f_{n-1}(x)$	----------	----------	---------- ..... ----------	$f_{n-1}(PW_{n-1}/V_s)$*	----------
$f_n(x)$	----------	----------	---------- ..... ----------	----------	$f_n(PW_n/V_n)$*

In (20), the *extensions* are the rows containing the function-value of the words in the possible worlds and the valuations. Notice that in the n-th world in the interpretation of the n-th word, the extension is marked with an asterisk. This is done to make it clear that in this world the *meaning itself, as the set of denotations (extensions) for this word*, is to be the extension. If we would give the example of a common noun such as 'eye-doctor', we could say that its extension in the various possible worlds consists of a set of individuals, except in one world where the extension is the *set of all the sets that make up its extension in the possible worlds of the model.* This is the set of denotations of the word viewed as a whole, and may therefore be said to define its 'meaning'. In other words, in some possible world the *macrocosm* of that part of the model that has to do with the interpretation of a linguistic element is *projected* within the *microcosm* of one possible world.

Coming back now to what has been said before about verbs of propositional attitude, there is at least one thing that should be clear, namely that these verbs are able to *influence the (usual) denotations of elements in their scope*. By this, I simply mean that in context with these verbs, words may come to get new denotations or denotations

not normally assigned to them. As has been shown, the supervaluated meanings (v_s-values) of elements may be different though 'classically' there are no differences. This is of course due to the presence of idiosyncratic valuations, v_i, that have to be taken into account with verbs of propositional attitude. Loosely speaking, it looks as if *these verbs generate relations between words and words or between words and things that did not exist before in the model under consideration.* Therefore, we would want to say that these verbs are *intentionality-creating verbs*. With this in mind, let us give a formal elaboration of the intuitions just presented. Given the diagram (20) expressing the denotation assignment, we propose to identify the function corresponding to verbs of propositional attitude as a *function over this assignment* in the following way:

(21) $g(x) = f_x(x)$ and R

What does this function say? It instructs us to take the 'x-th set of instructions', say the x-th word, then to apply it to the x-th argument, – in this case the x-th possible world – and then to *RESHUFFLE (R)* the value of the function on this argument. In other words, the function g(x) always picks out that possible world where the *representation of the meaning (set of extensions) of the word functions as its extension* (that is what '$f_x(x)$' instructs us to do) and reorders or changes this representation in a certain way. This means that the extension-assignment is to a certain extent changed. *Reshuffling* this assignment is an operation effected in one possible world *on the representation of the extensions in the possible worlds of the model.* But as this representation is a representation of the extensions assigned to the word by the model, it is necessary to change also the extension-assignments in the other worlds of the model, because there should be a strict identity between the extensions over the model and their representation in one possible world. So, what really happens as a result of the R-operation, is that the model may be completely changed in accordance with the requirements of the meaning-relation as expressed by the verb of propositional attitude. From the point of view of *calculability*, however, we must say that the function corresponding to these verbs is *not effectively calculable*, for almost the same reasons as the number-theoretic function (16). Suppose that the function (21) could express the meaning of words in propositional

attitude contexts in general. Then, it would be the case that g(x) equals $f_{i_0}(x)$ for some word i_0. But, if this is so, by the definition (21), '$f_{i_0}(x) = f_x(x)$ and R' – with *x* a possible world –, would hold for all values of x, and particularly for that possible world for which the value of the function is (the meaning of) the word itself. This would yield:

(22) $f_{i_0}(i_0) = f_{i_0}(i_0)$ and R.

This is impossible, *because no word's meaning equals itself and at the same time something else induced by the R-operation*. The $f_{i_0}(i_0)$ – values are the ones indicated in (20) by means of an asterisk. Due to the R-operation they are changed in several ways, but as they are identical with the set of extensions in all possible worlds, this identity is destroyed by the action of 'R'. Therefore, by (22) a contradiction is reached, and the function g(x) of (21), as a function over the meaning assignment in the model and a representation of the class of verbs of propositional attitude, defines a kind of *diagonal operation*, which, metaphorically speaking, brings us out of the usual meaning-assignment.

5. *Conclusion: non-standard meanings and model-theoretic ontology*

The identification of verbs of propositional attitude with 'non-computable' or 'not-effectively calculable' functions provides us now with a link to what has been said in part 3 on the nature of intentionality and its representability within a model-theoretic framework. There, the formula (4) expressing the absence of intentionality within the (standard) model was found to be true but unprovable. In fact, (4) turned out to be an undecidable formula. In the same way as the undecidability of (4) was established, it could be shown that the predicates corresponding to the representing function (21) of *verbs of propositional attitude* are undecidable (Kleene '67, 245) and that within the semantics there is a formula corresponding to these that is itself undecidable. What this formula would express is a true property of word – meanings, namely that *no word's meaning equals itself and at the same time something else induced by the R-operation*. This was the conclusion we reached at the end of part 4.

The foregoing is true in the standard-model, but by (6a-b) we are able to add the *negation* of the formula expressing this true property of meanings to our semantics and find a model for this new system. Consequently, the formula representing the verbs of propositional attitude is true in the standard model, but *false in a non-standard* one. What is true in the standard-model is that no word's meaning has the property just mentioned, but in the non-standard model there are supposed to be objects satisfying the negation of the formula expressing this property. These objects are, then, the *non-standard meanings of linguistic elements*.

Looking at it from another point of view, we could say that verbs of propositional attitude redifine models and meanings in all kinds of ways. They are elements that 'pull the plug out of the semantic bathtub', and then fill it up again with meanings and intentionality-relations that may be very different from the usual ones.

That is why these verbs may be called *intentionality-creating verbs*. They provide new links between things and words, and, on the theoretical side, give rise to a *hierarchy of models each non-standard from within a previous one*.

I think that *truth-value gaps* as introduced in part 4 illustrate well the idea that from within one model some meanings are non-standard, for instance in the sense that the absence of identity between such equivalent words as 'eye-doctor' and 'oculist' in belief-contexts is indicated by a truth-value gap. This truth-value gap in not the expression of the lack of truth-value altogether, but rather that the items in question have *unknown truth-values* (Haack '74, 58). Truth-value gaps can be considered as the 'don't-know'-answer of the model, when confronted with an intentionality-relation that is not its own.

I will not have much to say about the nature and formal representation of the non-standard meanings that are introduced in the semantics. I think the problems involved with these concepts, philosophical as well as conceptual and formal, will not be easily solved. I would, however, want to suggest to consider these non-standard meanings as *ordered pairs consisting of the denotations assigned by the non-standard model and the ones assigned by the standard-model*.

So defined, non-standard meanings may be used not only to describe the semantics of verbs of propositional attitude, but also of

such verbs as 'dream' or 'imagine', which are closely related to them as they are also intentionality-creating verbs. Consider the following sentence:

(23) I dreamt that *I was Brigitte Bardot and that I kissed me*. (Lakoff '70, 639)

In this sentence, there is an identity in the 'dream-world' – in which the usual relation of intentionality does not hold – between myself and the well-known rigid designator Brigitte Bardot. At the same time, however, a complete identity is denied, because otherwise I would not be able 'to kiss me'. The interpretation of (23) seems to require both the usual denotation of 'I' and 'brigitte Bardot' and at the same time a different one in which there is an identity of both individuals. From the point of view of the standard-model the individual defined by the model created by the verb 'to dream' could be called a *diagonal individual*, by virtue of the function by which it is defined. This function requires the individual denoted by 'I' to have both its usual denotation and at the same time some other one defined by the rigid designator 'Brigitte Bardot'. This is clearly a *'non-standard individual'*. It may be said that the individual denoted by 'I' in the non-standard model consists of the *ordered pair* of the usual denotations of two individuals, and is therefore in a sense *composite*. In the same vein we can say that the sentence serving as a complement to the verb 'to dream' has no truth-value in any of the possible worlds from the standard-model, but it may well be true in the world identified by the intentionality-creating verb. The world in which the truth of the complement sentence is guaranteed is evidently non-standard and is called a *diagonal world*. One can immediately feel the problem of 'counterpart-relations' and identifiability of individuals involved in this approach. As I said, I will not go into that there, but I am convinced that these problems can be solved and are not more difficult than 'more traditional' problems in model-theoretic semantics, such as the definition of 'possible world'.

The approach presented here implies a treatment of propositional-attitude verbs that is, in a sense, *metalinguistic*. The semantic objects of these verbs are not merely intensions of words or propositions, but they are *the models themselves* as a representation of the meaning. Furthermore, with these verbs the meaning is 'folded back' upon itself

by the operations and functions described and this allows us to change them and to switch from one model to another.

Finally, all this forces us to modify the ontology of the model in (1) in the following way:

(24) $\mathfrak{A} = (A,I,J,\leqslant,F,\mathfrak{A}^*)$

This modification means that *the model contains a representation of itself* ($\mathfrak{A}^*$). To those objecting to the principle of including a representation of the model within itself we would want to say, paraphrasing William James's 'little old lady' that "it is models all the way down". The overall picture is one of a kind of 'Looking-Glass-Semantics' in which certain elements function in the same way as Alice's bottle of magical potion in a model-theoretic Wonderland. Here, *self-reference* is not any longer a 'private vice' but a 'public virtue', the more as it enables us to clarify the important epistemological and semantic concept of *intentionality*.

Katholieke Universiteit Leuven Roger VERGAUWEN

Research Assistant at the National Fund for Scientific Research
Hoger Instituut voor Wijsbegeerte
Kardinaal Mercierplein 2
3000 Leuven
Belgium

REFERENCES

Anderson, A.
1984 "General intensional logic". In: Gabay D. and Guenthner (eds.), 355-385.
Bäuerle, R., Egli, U., Von Stechow, A. (eds.)
1979 Semantics from different points of view. (Berlin-Heidelberg-New York)
Davidson, D., Harman, G.
1972 Semantics of natural languages (Dordrecht).
Davis, M.
1982 Computability and Unsolvability (New York).
Dowty, D., Wall. R., Peters, S.
1981 Introduction to Montague semantics (Dordrecht).
Gabay, D., Guenthner, F. (eds.)
1984 Handbook of philosophical logic, vol. II (Dordrecht).

Haack, S.
1974 Deviant logic (London-New York).
Hoare, C., Allison, D.
1972 "Incomputability". Computing Surveys, Vol 4, 3, 169-178.
Kleene, S.
1967 Mathematical logic (New York-London-Sydney).
Lakoff, G.
1972 "Linguistics and natural logic". In: Davidson and Harman (eds.), 545-656.
Lewis, D.
1976 "General semantics". In: Partee (ed.), 1-50.
Mc Cawley, J.D.
1981 Everything that linguists have always wanted to know about logic, but were ashamed to ask (Chicago-Oxford).
Montague, R.
1970 "Universal grammar". In: Thomason (ed.), 222-246.
Montague, R.
1973 "The proper treatment of quantification in ordinary English". In: Thomason (ed.) 247-270.
Partee, B.
1976 Montague-Grammar (London-New York).
Partee, B.
1979 "Semantics, Mathematics, or Psychology". In: Bäuerle, R., Egli, U., Von Stechow, A. (eds.), 1-14.
Searle, J.
1980 "Minds, Brains, and Programs": The behavioral and Brain Sciences 3, 417-457.
Thomason, R.H. (ed.)
1974 Formal Philosophy: selected papers of Richard Montague. Edited and with an introduction by Richmond H. Thomason (New Haven-London).
Van Fraassen, B.
1966 "Singular terms, truth-value gaps and free logic". Journal of Philosophy 63, 481-495.
Vergauwen, R.
1984 "Propositional Attitudes in model-theoretic semantics". Logique et Analyse 105, 39-61.