Title	Rigid designation and Frege's puzzle
Sub Title	
Author	Tancredi, Christopher
Publisher	Centre for Advanced Research on Logic and Sensibility The Global Centers of Excellence Program, Keio University
Publication year	2011
Jtitle	CARLS series of advanced study of logic and sensibility Vol.4, (2010.) ,p.185- 196
JaLC DOI	
Abstract	
Notes	Part 3 : Cognition and Language
Genre	Research Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO12002001-20110331-0185

21 Rigid Designation and Frege's Puzzle

Christopher Tancredi[1,2]

[1] Institute for Cultural and Linguistic Studies, Keio University
[2] Centre for Advanced Research on Logic and Sensibility (CARLS), Keio University

I. Introduction

In *Naming and Necessity*, Kripke argues that names in natural language are typically interpreted as rigid designators. What this means is that the contribution of a name in a given language to semantic interpretation is an individual, and that this contribution is constant across possible worlds for that language. This view of names was motivated by considerations of metaphysical modality and what Soames (2002) refers to as the modal profile of a sentence. The same view, however, faces significant and well-known challenges when it comes to statements of propositional attitude. The basic problem is the following. If the sole contribution of a name is the individual it refers to, then coreferential names should be substitutable everywhere, including in the complement clause of a propositional attitude statement. Thus, the theory leads to the expectation that the sentences in the following pairs should be semantically equivalent, given that Hesperus is Phosphorus and Clark Kent is Superman.

(1) a. Hesperus is Hesperus
b. Hesperus is Phosphorus

(2) a. Lois Lane believes that Superman can fly

b. Lois Lane believes that Clark Kent can fly

The problem is, of course, that in the first case we seem to be capable of learning something from the second sentence that we cannot learn from the first, and in the second case that our intuitions stemming from knowledge of the Superman story tell us that the (a) sentence is true while the (b) sentence is false. The problem the theory faces, then, is that of explaining why we have such different intuitions about such pairs of sentences that are predicted to be semantically identical.

There have been many attempts to solve these problems, falling into two basic categories. On the one hand are solutions that attempt to re-introduce a semantic distinction into the two sentences. Theories based on Interpreted Logical Forms such as those of Ludlow and Larson (1993) as well as those that introduce modes of reference such as Salmon (1986) fall into this category. On the other hand are solutions that accept the semantic equivalence of the two sentences and attempt to explain our intuitions pragmatically. Notable among these approaches is Soames (2002). My goal in this paper is to motivate a third approach, one that builds on the basic insights of Kripke's theory of rigid designation but does not accept the conclusion that sentences like the (a) and (b) sentences above are semantically equivalent. This view looks inherently untenable at first glance. However, the untenability I will show stems entirely from a constrained view of the role that models play in interpretation. While the comments I make will apply largely to both types of example, for reasons of space I will restrict discussion to simple identity statements of the first type.

II. Direct Reference

On a direct reference theory, the individuals that figure in the semantic values of expressions are taken to be real individuals. In particular, this means that names refer directly, so that the interpretation of a name of an existing individual is that individual him- or her-self. Within a model theoretic semantics, this means that the individuals of a direct reference model minimally consist of the actual individuals of the world. The primary advantage of such an approach is that it makes the issue of connecting expressions with

real world individuals a trivial matter. This is because the interpretation of a natural language expression on this view will directly involve real world individuals, not mere representations of them, and so once the interpretation is determined there is no further question about how to connect that interpretation with the real world.

Since there is presumably a fact of the matter about who is who in the real world, this fact about identity infects the model, making it impossible for the model to contain Hesperus and Phosphorus as distinct individuals given that in fact Hesperus and Phosphorus are one and the same. While the necessary identity of Hesperus and Phosphorus follows from a direct reference theory, a model whose individuals are themselves model-individuals rather than actual individuals is not restricted in the same way. In particular, if the individuals of a model are themselves model-individuals, nothing requires that a model identify Hesperus and Phosphorus as a single individual rather than as two, regardless of what the real world turns out to be like. It is this fact that I will take advantage of here to motivate a new view of identity statements and attitudes. The basic idea I will argue for is that model theory allows for and in fact requires use of multiple models if it is to be adequate to the tasks of interpretation.

III. Model Theory

Model-theoretic semantics provides the basis for formalizing the semantics of natural language. As its name implies, model theory is concerned with how to model meaning in natural language. A model will be sufficient for this purpose if it provides all of the tools needed for giving adequate semantic interpretations to all sentences uttered by all speakers of the language it models.

For the purposes of this paper, I take a model to be a tuple <D,W,F> consisting of a set of individuals D, a set of worlds W encoding relations among (at most) the individuals in D, and a valuation function F. The model is used to give an interpretation to expressions in a language L. The valuation function F associates the basic expressions of L with extensions in each of the worlds in W – names with individuals, one-place predicates with sets of individuals, n-place predicates with n-tuples of individuals, etc. While it is

picturesque to view the worlds of the model like we view the actual world we live in, filled with real people smelling real flowers and with real stars exploding in real supernovae, for the purposes of model theory this is not a requirement. All that a world has to do in model theory is to make it possible to determine the extensions of basic expressions in a language. While model theory is in many respects non-intuitive, it provides a useful tool for explaining aspects of meaning related to reference, truth, entailment and modality among other things and in this sense serves an indispensable role in developing a semantics of natural language. The sketch just given of model theory ignores several things that are required in a fully adequate account, minimally including a set of times and also plausibly a set of events. These limitations reflect the focus of this paper, not substantive claims about model theory.

To make this paper equally accessible to linguists and philosophers, I employ a somewhat non-standard version of model theory that marries two standards in the literature. In particular, I divide the labor of the interpretation function in standard model theories into two parts, a compositional part and a lexical part. The compositional part is handled by a Heim and Kratzer (1998) style interpretation, based on function application and predicate abstraction. This approach takes the meanings of basic and derived predicates to be functions, not sets, and has become a widely shared standard for semantic interpretation over the past decade. The lexical part incorporates standard set theoretic interpretations in the meanings of lexical items, and will connect these with the functions needed for Heim and Kratzer style composition. This part incorporates standard model theoretic interpretations into Heim and Kratzer style functions, and makes reference to the valuation function F from the model. To illustrate, consider the simple sentence *John left*, and assume that the sentence is to be interpreted with respect to a model M and a world w. The compositional part of the interpretation will yield the following:

(3) $[\![\text{John left}]\!]^{M, w} = 1$ iff $[\![\text{left}]\!]^{M, w} ([\![\text{John}]\!]^{M, w}) = 1$

The lexical part of the interpretation will then tell us how to interpret the predicate *left* and the name *John*. I assume the following as lexical entries for these items:

(4) $[\![\text{ left }]\!]^{M, w} = \lambda x \in D_e . x \in F(\text{left})(w)$
$[\![\text{ John }]\!]^{M, w} =$ $F(\text{John})(w)$ iff $F(\text{John})(w) \in D$, undefined otherwise

Together these two parts combine to tell us that the sentence *John left* is true iff $F(\text{John})(w) \in F(\text{left})(w)$, i.e. iff the individual in D picked out by F in w by the name *John* is a member of the set of individuals in D picked out by F in w by the predicate *left*.

There are two general versions of model theory that differ in their conception of the relation between individuals and worlds. On the Lewis view, no two worlds contain the same individuals. This means that for no two distinct worlds w and w' is it ever the case that $F(\text{John})(w) = F(\text{John})(w')$. The best we can do on such a model for relating individuals across worlds is to determine what individual in one world is most like some individual in another world. Lewis proposes a counterpart relation for accomplishing this. The Kripke view of possible worlds differs radically from the Lewis view. For Kripke, individuals typically exist in multiple worlds. Worlds can differ in two ways: in the individuals they contain, and in the properties and relations that hold among individuals. Among other things, this makes it possible in principle for an identity such as $F(\text{John})(w) = F(\text{John})(w')$ to hold even when $w \neq w'$. It is on the basis of this view of possible worlds that Kripke proposes that names should be analyzed as rigid designators. For the remainder of this paper I will be adopting the Kripke view of worlds. I will also be adopting a modified version of rigid designation, one in which the individual denoted by a name is an individual of a model (not directly an individual in the real world) and that with respect to any single model a name designates rigidly.

An often implicit assumption made in work on model-theoretic semantics is that speakers share a common language, say English or Japanese. A model is then developed for interpreting that language. This assumption puts severe constraints on the uses to which model theory can be put, too severe I argue. In particular, it makes it impossible for the model to encode idiosyncrasies of individual speakers. This artificial limitation on the way that model theory is put to use makes it impossible to model both understanding and evaluation of what a speaker says in those cases in which the speaker takes herself

to say something true but the hearer evaluates what she said as false. It is important to note that this is not an inherent limitation on model theory itself but only a consequence of the way model theory has been used to model natural language interpretation. Adopting the I-language perspective of Chomsky (1986) gives us a different target for interpretation – the separate I-languages of individual speakers. This perspective makes it possible to introduce multiple models for interpretation, in particular including separate models for understanding and for evaluation, a modification in the use of model theory that I have argued in Tancredi (2007) is needed.

In addition to the assumption of a common language, it is also commonly assumed that once a language is fixed, interpretation of an expression in that language makes reference to exactly one model. The initial choice of model might be arbitrary, of course. However, once made it is assumed that no additional models enter into interpretation. I argue in contrast that propositional attitude attributions cannot be adequately interpreted by restriction to a single model. Rather, attitude predicates must be allowed to introduce a distinct model with respect to which the proposition it embeds is interpreted. The model introduced serves a role very similar to that played by models used for understanding (rather than evaluating) other speakers, and is required for much the same reasons. The main difference is that in the case of attitude predicates the model is taken to model understanding of the matrix subject rather than of the speaker.

Expanding the ways in which model theory is used to model natural language places model theory in a much more central role in all aspects of interpretation. In particular, the proposed revisions make it possible for models to play many if not all of the roles that Fregean senses were invented to deal with. This raises the possibility of once and for all dispensing with Fregean senses in semantics, a desirable goal if it can be reached. Though I make no pretense of showing that senses can be eliminated entirely from semantics, I do show that they are unnecessary in explaining our intuitions in two areas which initially were used to motivate their introduction, namely in understanding identity statements and in interpreting propositional attitude attributions.

IV. A Puzzle for Model Theory

In this section I examine Frege's well-known puzzle of identity from the perspective of model theory as outlined above together with the assumption that names are rigid designators. This puzzle is not standardly cast as a puzzle for model theory in this way, and taking Frege's puzzle as a puzzle for Kripke's rigid designation theory of proper names is clearly anachronistic. As we will see, however, casting the puzzle in this way both gives a clarity to the puzzle itself and puts into a clear light the challenges this puzzle poses to standard model theoretic accounts of semantic interpretation.

1. Frege's Puzzle of Identity

Frege's puzzle of identity stemmed from his attempt to spell out a compositional theory of interpretation. One of the basic premises of the theory he developed is that at least part of the meaning of a name is the individual the name picks out. The puzzle arises if we take this view to the Millian extreme of taking the referent of a name to exhaust its meaning. With respect to such a view, the puzzle of identity can best be stated as a question: how can an identity statement of the form "A is A" differ in cognitive significance from one of the form "A is B" when the latter statement is true? To see why this puzzle is a puzzle for model theory, consider the model-theoretic interpretations of the these two sentences, keeping the model and world of evaluation constant and taking "is" to denote the identity relation among individuals:

(5) a. $[\![\text{A is A}]\!]^{M, w} = 1$ iff $F(A)(w) = F(A)(w)$
b. $[\![\text{A is B}]\!]^{M, w} = 1$ iff $F(A)(w) = F(B)(w)$

The puzzle comes in when we ask what values F assigns to A and B in w. Suppose that $F(A)(w) = a$, for some individual a in the domain D of M. If the statement "A is B" is true with respect to M and w, then it follows that $F(B)(w) = a$ as well, and thus that the interpretations of the two sentences are identical:

(6) $[\![\text{A is A}]\!]^{M, w} = [\![\text{A is B}]\!]^{M, w} = 1$ iff $a=a$

Furthermore, if we follow Kripke in treating names as rigid designators, then it will follow that for any distinct worlds w and w' in M, F(A)(w) = F(A)(w'), and similarly for B, so that the identity given in (6) holds in every world of the model. And yet the statement "A is B" is one that can easily be called into question, doubted, or not known to be true while the statement "A is A" cannot be.

Note that we do get a distinct interpretation for the two sentences in the case in which "A is B" is false with respect to M and w. In this case it follows that for some individual *b* distinct from *a*, F(B)(w) = *b*, giving us the following interpretations:

(7) a. $[\![\text{A is A}]\!]^{M, w} = 1$ iff $a=a$
b. $[\![\text{A is B}]\!]^{M, w} = 1$ iff $a=b$

However, this distinction is of little help. It makes it possible for a person to accept one of the statements and doubt the other, but only when A and B are used to refer to distinct individuals. This is not puzzling in the least, but nor does it help us with the case in which A and B happen to refer to the same individual.

Frege's puzzle is an unavoidable consequence of our assumptions (i) that the interpretation of a name is the individual it names, (ii) that names are rigid designators, and (iii) that a single model is used for interpretation. The core of a model is the set of individuals it contains, D. If D contains two individuals, say *a* and *c*, then it follows from the definition of a set that $a \neq c$, a non-identity that holds in every world of the model. We may well have multiple names for some of these individuals, for example calling *a* either A or B while calling *c* C. If this is the case, "A is B" will be true, but this does not reflect any identity between distinct individuals. At the level of individuals such a statement merely asserts that $a=a$, something that once again holds in every world of the model. "A is C" in this case would be false, since it asserts that $a=c$, an impossibility if a and c are distinct elements of D. Thus if "A" and "B" are both names for *a* in a given model, it follows that the interpretation of "A is A" within that model will be identical to the interpretation of "A is B". It follows from these considerations that within standard model theoretic treatments of semantics, whatever distinctions we recognize in cognitive significance between these two statements cannot

come from a distinction in their model theoretic interpretation. That is, that "A is A" can have such a different cognitive significance from "A is B" cannot follow from the model theoretic meanings of the two statements alone. And yet model theory as conceived above has nothing else to offer as a basis for the distinctions.

2. Sense and Reference: Frege's Solution

Frege proposed to account for our different attitudes toward distinct but synonymous semantic truths by appealing to a distinction between sense and reference. A sense for Frege is essentially an instruction for picking out a reference. On this view, the statement "A is A" differs from the statement "A is C" only at the level of sense, and not at the level of reference. We can formalize the distinction by taking "is" to denote a relation between senses, not directly between referents, as below:

(8) $[\![\text{ is }]\!] = \lambda x \lambda y . \iota z(x \text{ picks out } z) = \iota z(y \text{ picks out } z)$

Here, x and y range over senses and z over referents. By including senses in the meanings of identity statements we introduce a distinction that could be taken to correlate with our intuitions about the differences between the two identity statements under consideration. For while it is obvious that $\iota z(A_s$ picks out z) = $\iota z(A_s$ picks out z) (where A_s is the putative sense of the name A), it is not immediately obvious from inspection alone that $\iota z(A_s$ picks out z) = $\iota z(C_s$ picks out z). If grasping this identity in this form constitutes part of understanding the identity statement "A is C", then this could plausibly form the basis for an explanation of why we can have such different attitudes to pairs of true identity statements. Or so the explanation goes.

Whatever advantages this theory may have elsewhere, as Kripke (1972) has shown in eviscerating detail, as a theory of the meaning of proper names it fails miserably. The problem in a nutshell is that there is nothing that can serve as the sense of a name without thereby making obviously incorrect predictions about the interpretations of sentences containing names. In order for the Fregean view to have any plausibility for definite descriptions within model theory, it is necessary to relativize reference determination to worlds. Doing so makes it possible to account for intuitions about sentences like "Imagine the US president were a republican." In particular it enables us to

understand this statement as telling us to imagine what the world would have to look like in order for what we are being asked to imagine to be true. We have two distinct ways we can understand this sentence. In the one, we find the current president, namely Barack Obama, and imagine that he is a republican. In this understanding, the instructions to identify an individual as the US president are followed in the actual world, though the property of being a republican is imagined of holding of that person in other possible worlds. On the second way of understanding the sentence, we imagine a situation in which the identity of the US president is not fixed once and for all but in which whoever happens to be the US president in an imagined world is a republican. The instructions encoded in the sense of the expression "the US president" are in this case followed separately in each world.

If this same distinction is applied to proper names and we follow Frege in taking proper names to have a sense that determines their reference, this view gives clearly counterintuitive results. Proper names contrast with definite descriptions in this respect. If Frege's solution were to extend to proper names, it would lead to two wrong predictions. The first is that a sentence like "Imagine Barack Obama were a republican" should be ambiguous in the same way that the sentence "Imagine the US president were a republican" is. Of the two ways of understanding this latter sentence, however, only the first seems to find an analogy in the former sentence. Second, attributing to an individual named A the property of being picked out by the sense A_s of the name should result in a necessary truth. The problem is that there is no independent way of specifying what A_s could be that makes this prediction both testable and true. If A_s is taken to be identifiable only as the sense of the name A, then it is unlikely that any counterexamples can be given which show the prediction to be false. However, in that case the theory also becomes unfalsifiable as a theory of the meaning of names. If, in contrast, we accept that there are other ways of identifying A_s besides by using the name A, e.g. as the individual having property P, then the prediction becomes testable but always turns out to be false. For whatever property P may be, the statement "A has property P" is predicted not only to be true but to be necessarily true. Under analysis the sentence becomes equivalent to "The person who has property P has property P", a trivially true statement. However, sentences of the sort "A has property P" are never understood as necessary truths when A is a proper name. This, of course, is part of Kripke's objection

to analyzing proper names as having senses.

V. Toward A New Solution to Frege's Puzzle

We have seen that Frege's puzzle poses a challenge to Kripke's view of names as rigid designators, but also that Kripke's objections to treating names as having senses is a valid one, leaving us without any analysis of the original identity statements. I propose that the problem that such statements highlight is a problem in the way in which models are used for interpretation. If we use a single model to model interpretation of the two sentences in (1), and if the names contained therein are treated as rigid designators, then there is no way of avoiding the problematic conclusion that the two sentences are semantically identical despite intuitions to the contrary. However, models as a tool for interpretation do not need to be so restricted in their use. In particular, if we take interpretation to be with respect not to a single model but with respect to multiple models simultaneously, then the problem sketched above can be avoided. The key to avoiding the problem is to use a Kripke style of relating individuals across worlds within a model, but a Lewis style of relating individuals across models. The concept of rigid designation will then be limited in its application to a single model, so identity statements will only be necessary within a single model, not across models.

With multiple models at our disposal, the identity statement "Hesperus is Hesperus" comes out as true not only in the actual world, and not only in every world of a single model, but in every world of *every* model. This sets it apart from the statement "Hesperus is Phosphorus". If this latter statement is true of one world of a model it is true of every world of that model. Similarly, if it is false of one world of a model it is false of every world of that model. However, it is perfectly plausible for the statement to be true in the worlds of one model and false in those of another, and the fact that both kinds of models exist gives us a basis for distinguishing these two statements. We need only expand the way we interpret statements so as to encompass not only the possibility that a given individual might have different properties, but also the possibility that the individuals we have available might be different from what we take them to be. And introducing multiple models accomplishes exactly that.

Bibliography

Chomsky, Noam (1986) *Knowledge of Language*: *Its Nature, Origin and Use*, Prager.

Heim, Irene, and Angellika Kratzer (1998) *Semantics in Generative Grammar*, Blackwell.

Kripke, Saul (1972) *Naming and Necessity*, Princeton University Press.

Ludlow, Peter, and Richard Larson (1993) "Interpreted Logical Forms,"*Synthese*.

Salmon (1986) *Frege's Puzzle*, MIT Press.

Soames (2002) *Beyond Rigidity: The Unfinished Semantic Agenda of* Naming and Necessity, Oxford University Press.

Tancredi, Christopher (2007) "A Multi-Model Modal Analysis of I-Semantics: Part 2, Identity and Attitudes, in Yukio Otsu (ed.) *Proceedings of the Tokyo Conference on Psycholinguistics 8*, 2007, pp. 21–55.