The Fallacies of the New Theory of Reference: Some Afterthoughts[1]

GABRIEL SANDU

1. Introduction

The *New Theory of Reference* is a view according to which there is a subclass of expressions in natural language which are genuine naming devices in that they function as rigid designators, that is, they designate the same object in all possible worlds in which that object exists. The view has been mainly developed by Saul Kripke in his *Naming and Necessity* (Kripke is explicit though about not being after a theory) but elements from it may be found already in the work of Ruth Barcan Marcus in her talk "Modalities in Intensional Languages," delivered in 1962 at a session of the *Boston Colloquium for the Philosophy of Science*. In this paper I will reassess some of the claims made in Hintikka and Sandu, "The Fallacies of the New Theory of Reference" (1995). In that paper we denied the need for a class of basic expressions which function as rigid designators and claimed that the rigidity of those expressions can be expressed by using quantifiers. In the present paper I will qualify some of these assertions.

2. Kripke: Naming and Necessity

Kripke's famous lectures on *Naming and Necessity* (NN) were given in Princeton 1970, then published *verbatim* in Davidson and Harman (1972), and finally published by Harvard University Press as a book in 1980. The latter contains an Introduction in which Kripke tells us that he reached the main ideas of the monograph around 1963–64 based on his earlier

[1] I am greatly indebted to Joseph Almog for suggestions to improve the paper.

work in the model theory of modal logic. The earlier work refers to Kripke (1963), an article in which he developed a model-theoretical semantics for a first-order modal predicate language (with no individual constants but only individual variables). At the beginning of that article, we are told that:

> The authors closest to the present theory appear to be Hintikka and Kanger. The present treatment of quantification, however, is unique as far as I know, although it derives some inspiration from acquaintance with the very different methods of Prior and Hintikka. (Kripke 1963, 83, fn. 1.)

What is unique about Kripke's treatment of quantification in Kripke (1963)? It is the quantificational structure it imposes on a set of possible worlds (and the corresponding accessibility relation). It is such that:

- The Tarski-type notion of satisfaction of a formula is now generalized to a possible world, an interpretation of the non-logical vocabulary and an assignment to its free variables.
- Every possible world is endowed with its own domain of individuals which is the range of the quantifiers occurring in the formula (obeying the constraint: if an object exists in a possible world w, and w' is a distinct world accessible from w, then that object exists also in w'.)
- The individual assigned to a free variable does not depend on a possible world but is picked up, once and for all, from the union of the domains of the possible worlds ("rigid" interpretation of free variables.)

This semantic interpretation renders valid Leibniz's law of identity $\forall x \forall y(x = y \rightarrow Nec\ x = y)$. The language, however, does not contain individual constants and thus the semantic interpretation does not tell us anything about them, even less so about the interpretation of names in natural language.

In the Introduction to *NN* (1980) Kripke tells us that little by little (1963–64) he came to be convinced that names in natural language also function as rigid designators and that the necessity of identities holds for them too. The point about rigidity that Kripke emphasizes is that "we have a direct intuition of the rigidity of names, exhibited in our understanding

of the truth conditions of particular sentences." (Kripke 1980, 14) There are two kinds of such sentences that Kripke considers.

One of them consists of *simple sentence* like:

(i) Aristotle was fond of dogs.

There is general agreement that (i) is true if and only if a certain philosopher we call "Aristotle" was fond of dogs. But for Kripke our understanding of (i) requires more: we have to be able to recognize "the conditions under which a *counterfactual course of history*, resembling the actual course in some respects but not in others, would be correctly (partially) described by (i)" (Kripke 1980, 6). And that happens if and only if the same aforementioned man would have been fond of dogs, had the situation obtained." (Ibid.)

The other kind of sentences Kripke considers are *counterfactual sentences*:

> In the monograph I argued that the truth conditions of 'It might have been the case that Aristotle was fond of dogs' conform to the rigidity theory: no proof that some other person other than Aristotle might have been both fond of dogs and the greatest philosopher of antiquity is relevant to the truth of the quoted statement. (Kripke 1980, 12–13)

The intuition behind the two kinds of sentences considered by Kripke is that once a proper name, say "Nixon," names a particular person, it would continue to do so in all counterfactual scenarios in which that person exists. That led Kripke to develop his doctrine of proper names as rigid designators. As he tells us in the Introduction:

> … I imagined a hypothetical formal language in which a rigid designator 'a' is introduced with the ceremony, 'Let 'a' (rigidly) denote the unique object that has property F, when talking about any situation, actual or counterfactual'. It seems clearly that if a speaker did introduce a designator into language that way, then in virtue of his very linguistic act, he would be in a position to say 'I know that Fa', but nevertheless 'Fa' would express a contingent truth (provided that F is not an essential property of the unique object that possesses it.) First, this showed that epistemic questions should be separated from

> questions of necessity and contingency, and that to fix a reference is not to give a synonym (NN, 14).

Kripke's connection between names and modal questions in natural language was an important insight and it is difficult not agree with him that one of the main reasons for Russell's proposal of a theory of names (the so-called *descriptive theory*) incompatible with our intuitions of rigidity was his failure to consider modal questions.

3. The fallacies of the new theory of reference

In Sandu and Hintikka (1995) we argued that, contrary to what the proponents of the New Theory of Reference, including Kripke, hold, there is neither class of expressions (singular terms) which function as rigid designators nor primitive semantic phenomena of rigidity in natural language. We made our point by using, not the modalities of necessity and possibility, as Marcus and Kripke did, but epistemic notions like knowledge and belief whose logic Hintikka had analyzed in his *Knowledge and Belief* (1962). Let us shortly recall the basic steps.

In a first step, we rehearsed the well-known distinction between *de dicto* vs *de re* knowledge which seems to require two uses of certain singular terms (definite descriptions). Here is one of the examples we used.

In the *de dicto* case, someone, say a, may know something, e.g., that b is S, abbreviated by $S(b)$, of whoever is or may be referred to by the singular term "b." For instance, Stefan may know something about Marie Antoinette's lover, whoever he might have been, for instance that he was not French. We represent such knowledge in the logical notation by "$K_a S(b)$," where "b" stands for the description "Marie Antoinette's lover." The model-theoretic import of the truth of "$K_a S(b)$" is that in all the scenarios compatible with what a knows, it is the case that $S(b)$. But given that Stefan does not know who the gentleman in question is, the term "b" ("Marie Antoinette's lover") will pick out different individuals in the different scenarios compatible with everything Stefan knows.

On the other hand, in the *de re* case, a may know something about the individual who in fact is b, without knowing that he is b. For instance, Stefan may know some fact or other

about Count von Fersen, who in fact was Marie Antoinette's lover, even if Stefan does not know this fact about him. The decisive step is to observe that the truth of the knowledge statement requires the phrase "Marie Antoinette's lover" to pick out the same gentleman (viz. Count von Fersen) in all the scenarios admitted by Stefan's knowledge. In general, knowledge "of the individual who in fact is b" cannot be expressed by a statement of the form "$K_a S(b)$" unless "b" picks out the same individual in all the scenarios compatible with what a knows. In such case the term "b" designates whatever it designates necessarily, and it might seem that, in order to express *de re* knowledge, we must have at our disposal "rigid designators" referring to whatever they refer to necessarily. Furthermore, this rigid reference cannot be mediated by any contingent definite description. For such a description can always in principle refer to distinct individuals in different possible scenarios. We took the protagonists of the New Theory of Reference to identify their rigidly referential singular terms with proper names. However, this is not the strategy we endorsed.

In a second step, we expressed the "rigidity" of a definite description as a particular kind of *de re* modal attitude. We then observed that the same technique can be applied to express the rigidity of proper names. Finally, we pointed out, rehearsing some of Hintikka's earlier arguments in Hintikka (1969), that for quantifiers to perform this job, they must be interpreted referentially (objectually), which in turn presupposes a mechanism of cross-identification of individuals as denizens of various possible situations (worlds). We exemplified all these claims using a toy logical language. We recall again the main stages.

In a first stage, we consider sentences of the form:

(1) $\exists x N S(x)$

(2) $\exists x K_a S(x)$

where "N" stands for the necessity operator and "K_a" for the epistemic operator "a knows that." On the referential interpretation of quantifiers, the truth of (2) in a possible world w requires that there be an individual in w which belongs to the extension of S in all epistemic a-alternatives to w. Similar

truth conditions can be formulated for (1). Thus in both of these sentences, one is saying that something is true of one and the same individual in a range of different possibilities. In (1), the relevant possibilia are all the states of affairs or courses of events that are being considered possible. In (2) they are all the possibilities left open by what a knows.

In a second stage we express the *de re* interpretation of our earlier examples, using variants of (1) and (2). We first consider the *de re* interpretation of "$K_aS(b)$" in which a knows something about the individual who in fact is b, without knowing that he is b. This is rendered in our logical notation by:

(3) $\exists x(x = b \wedge K_aS(x))$;

and the analogue *de re* interpretation of "$NS(b)$" is expressed by:

(4) $\exists x(x = b \wedge NS(x))$.

Neither (3) nor (4) requires that "b" ("Marie Antoinette's lover") pick up the same individual in all the relevant alternatives, but only that the individual which is the actual referent of "b" belong to the extension of "S" in all these alternatives. In other words, the truth of both (3) and (4) is consistent with "b" picking up different individuals in various possible scenarios.

In a final stage, we consider the particular *de re* interpretation of "$K_aS(b)$" according to which a also knows who b is:

(5) $\exists xK_a(x = b \wedge S(x))$.

The corresponding *de re* interpretation of "$NS(b)$" is similarly expressed by:

(6) $\exists xN(x = b \wedge S(x))$.

We can actually abstract from the claim "$S(b)$" and express the rigidity of "b" simply by:

(7) $\exists xK_a(b = x)$.

And likewise, we can express the rigidity of "b" in alethic contexts by:

(8) $\exists xN(x = b)$.

The truth of (5)–(8) forces "b" to refer to one and the same individual in all the relevant alternative worlds, including the actual one (we ignore here some problems concerning the non-existence of individuals). In other words, "b" acts as a "rigid designator," something that we also alternatively expressed as "a knows who b is."

Finally, we realized that the same "rigidifying" strategy works independently of whether "b" stands for names or definite descriptions. This led us to conclude that there is no need to assume any class of singular terms in natural language which act as "rigid designators." We expressed this in the paper in the following way:

> …as soon as we have quantifiers at our disposal, we do not need any other kind of direct representability. In sum, the right slogan of modal logicians should be: We do it with quantifiers. And this dispensability seems to invalidate all arguments for the need of rigid designators or anything remotely like them in natural or formal languages. (Hintikka and Sandu 1995, 252–253.)

Let me emphasize two points about our approach in the paper. One of them, which we often repeated, was that the strategy of imposing the rigidity of a singular term by an outside quantifier works because we interpreted quantifiers "referentially." That is, quantified formulae are interpreted with respect to a possible world and an assignment, the latter assigning individuals (from the joint domain of discourse) to the free variables which occur in the corresponding open subformulae. Thus, recalling our earlier example (we assume here that if an individual exists in a possible world, then it also exists in all its relevant alternatives):

> (9) $\exists x K_a(b = x)$ is true in a possible world w with respect to the assignment g if and only if there is an individual $\beta \in dom(w)$ such that $b = x$ is true in every a-alternative w' with respect to the assignment $g\left(\frac{x}{\beta}\right)$ if and only if there is an individual $\beta \in dom(w)$ such that the individual who is the semantic value of "b" in w' is β. (Ibid., 249)

We observe that the interpretation of the constant "b" inherits its "rigidity" (i.e., constancy of its semantic value in all the a-alternative possible worlds) from the "rigidity" of the varia-

ble "x" induced by the referential interpretation of the existential quantifier "$\exists x$" which binds it. There is nothing new here, as we also acknowledged in the paper: this line of reasoning has been countenanced much earlier by Kripke himself in Kripke (1963). In that paper he treats quantifiers in alethic contexts in a referential way; and in Kripke (1976), he makes the distinction between *de re* and *de dicto* interpretations of definite descriptions and observes that there is way of expressing *de re* belief by using quantifiers (ibid., 374):

(10) $\exists x(x = b^*$ and Jones believes that x is an airdale).

Here "b^*" is a definite description. (10) is essentially the same as our example (3) above. Kripke actually uses this formulation for a language which does not contain explicit scope indicators. One would think that such a language is a fragment of our natural language, and thereby does not contain quantifiers and variables. I will say something about this below.

The second point to be emphasized about our approach in the paper is that the strategy we followed to impose rigidity as a particular kind of *de re* epistemic attitude works, obviously, only for modal contexts. I will return to this issue below.

4. Criteria of cross-identification

Although the strategy we followed to impose rigidity in our paper relies on a referential treatment of quantifiers in modal contexts due to Kripke himself, it is not the strategy he finally endorses with respect to the rigidity of names. I will say something about this in the next section. For now, let me shortly comment on another major philosophical disagreement between his treatment and ours, technicalities aside. Its source lies in the requirement of an individual to be a denizen of several possible worlds. Or, we thought in the paper, echoing some of Hintikka's earlier work, such a requirement presupposes criteria of cross-identification:

> As a slogan, we may perhaps put it, quantifying in presupposes that criteria of cross-identification have been given. These criteria cannot themselves be expressed by quantifiers. For in order to do so, we must be able to compare the denizens of any two

> scenarios ("possible worlds") for identity. (Hintikka and Sandu 1995, 249)
>
> Bound variables do not, in any literal sense, refer to anything at all. The rigidity and directness they exhibit is not a matter of reference but of criteria of cross identity. (Ibid., 253)

The requirement poses no problem for Kripke for whom possible worlds "are little more than the miniworlds of school probabilities blown large" (Kripke 1980, 18). We recall in this context Kripke's well-known example with two dice being thrown. There are 36 possible outcomes, that is, 36 states of the dice that Kripke takes to be 36 possible worlds. One of them is the state (die A, 6; die B, 5); another one is (die A, 5; die B, 6), etc. These possible worlds are abstract, not complex physical entities, and there is no need for some further criteria to compare e.g., die A, 6 in the first world with dies A, 5 in the second world. All in all, for Kripke, philosophical questions like "Which die is that?" simply do not make sense, for, as he observes, the states of the dices are simply *given*. (Ibid., 17)

The requirement of criteria of cross-identification in my paper with Hintikka, on the other side, amplified some of Hintikka's ideas in the late sixties (which finally go back to Carnap's "individual concepts") and was motivated by the way we understand the truth-conditions of certain belief sentences. It is well known from the rich industry of epistemic puzzles that singular terms in such sentences do not seem to behave "rigidly" and this, in turn, seems to have something to do with the modes of identification of individuals. Whether the latter is somehow related to the question of the substitutivity of names in belief contexts, as Hintikka thought in his earlier work, a view we endorsed in the paper, is a difficult matter, one which I will not deal with here. My main concern is more modest, viz., to reassess the claim we made in the paper to the effect that the rigidity of singular terms can be expressed and thereby eliminated if we have quantifiers (and identity) at our disposal. Whether, in addition, something like criteria of cross-identification is needed or presupposed seems to me a secondary matter relative to this concern, although, I have to say, thinking about the role played in Kripke's account by individual essences, inclines

me to believe we were after something here. In any case, as I hope to make it clear below, I now think criteria of cross-identification are a secondary matter to questions of reference.

5. Rigidity, scope, and modal embedding

The fact that rigidity in the sense of constancy of designation can be expressed in formal languages with the help of quantifiers does not mean it is the correct way to capture the notion of rigidity for certain singular terms in natural languages. And it is this view which is in focus in *Naming and Necessity*. For those languages the mechanism consisting of quantifiers, variables, and binding, all in all, the "method of the variable," simply does not exist. Thus, it appears that the conclusion we drew, namely that it leads to the "dispensability of rigid designators or anything remotely like them in natural or formal languages" is not fully supported by the arguments we presented as they stand.

To be more precise, as I see it, there are two ways to counter our conclusion in the paper. Firstly, there is the claim that the expressibility of rigidity of singular terms with the help of quantifiers does not work for natural languages for the reason I just mentioned in the preceding paragraph. Secondly, there is the further claim that even if we had available scope distinctions, the expressibility strategy would not work simply because rigidity in natural language does not reduce to them, that is, is not a matter of scope distinctions.

I think that the first point can be easily taken care of: the scope mechanism can also be applied, although in a different format, to natural language to enforce rigidity. For instance, Dummett held the view that natural language has a convention according to which a name, in the context of any sentence, should be read with a large scope including all modal operators. The same idea, although in a different form and not applied to rigidity, appears in Hintikka's earlier work on the game-theoretical semantics (GTS) for natural language (Hintikka and Kulas 1985; Hintikka and Sandu 1991). GTS associates with a fragment of discourse a semantical game played by two players, Myself and Nature. Quantifiers, more generally logical expressions, and names prompt moves by

one of the players. A proper name prompts a move by Myself who chooses the referent of the name from the universe of discourse. Modal and intensional concepts are handled by combining game-theoretical semantics with possible worlds semantics. To take an example, the rule (G. knows that) looks like this:

- If the game has reached a sentence of the form "b knows that X" and a world w_1, then Nature may choose an epistemic b-alternative w_2 to w_1. The game is then continued with respect to X and w_2.

A specificity of natural languages, due to the lack of scope indicators, is that game rules must be complemented by a set of ordering principles which govern their order of application (cf. Hintikka and Kulas 1985, section 8). More importantly for the present purpose, the game rule for names, (G.name), has priority over many other game rules applicable to the constituents of the same clause. This amounts, in the traditional jargon, to proper names having "broader scope" over many other expressions in the same clause. True enough, the issue of the priority of proper names over the game rule for intensional operators has never been, to the best of my knowledge, systematically addressed in the GTS literature. My point in bringing it up is only to show that GTS has the resources to handle it. This way of handling it also shows, incidentally, that GTS assumes a convention about natural language according to which names have larger scope than many logical expressions and operators, including modal ones.

The "larger scope" view of rigidity in natural language has, however, been dismissed by Kripke, as somehow incoherent. As I mentioned in section 3, this view, held, among others, by both by Dummett and Hintikka, eliminates rigidity only in sentences with modal operators. In this connection, Kripke observes against Dummett that rigidity appears and makes sense not only in sentences with modal operators but also in simple sentences like "Aristotle was fond of dogs" (cf. our discussion in section 2). In other words, rigidity is a doctrine about the truth conditions of all kinds of sentences, simple and modal ones. Kripke acknowledges that the thesis of the rigidity of names in simple sentences can be expressed as

a "wide scope" phenomenon, that is, he agrees that that view is equivalent (ignoring complications arising from the possible nonexistence of an object) to the thesis that if a modal operator governs a simple sentence containing a name, the two readings with large and small scopes are equivalent (Kripke 1980, 12, fn. 15). But this equivalence, Kripke continues, "goes against the doctrine that natural language has a convention according to which only large scope reading is allowed. In fact, the equivalence makes sense only for a language where both readings are admissible" (ibid.). To conclude, the strategy we followed to eliminate rigidity in Hintikka and Sandu (1995) works only for sentences with modal embeddings. Rigidity, however, is a thesis about all kinds of sentences, including simple ones.

Perhaps I should add, commenting on the conclusion, that I believe Hintikka has never reached a definitive opinion on these matters. For instance, in Hintikka (1996), he reconsiders the difference between *de dicto* and *de re* epistemic attitudes. But now, somehow surprisingly, he uses the distinction to argue for the need for "rigid designators" in the language:

> …we need two kinds of singular terms. We need terms which pick up the same individual in all possible worlds; and terms which designate different individuals in different possible worlds. Constants proper serve the former purpose; ordinary (improper) serve the latter. Our improper constants are obviously related closely to Russell's logical proper names and to Kripke's 'rigid designators'. (Hintikka 1996, 122.)

Before closing the section, let me point out that the existence of rigidity in simple sentences (recall Kripke's example "Aristotle was fond of dogs") which shows its priority over modal embeddings, does not show in my opinion that there is no connection between naming and necessity, as claimed, e.g., in Almog (1986). As I observed in section 2, following Kripke, the rigidity of "Aristotle" in "Aristotle was fond of dogs" is manifest in the way we understand the truth conditions of this sentence in *counterfactual* situations.

6. Rigidity, quantifiers, and reference

It follows from what we said in the previous section that even independently of modal and attitudinal embeddings, the reduction of rigid reference to the objectual interpretation of quantifiers and quantifier scope is off the target, given that even in extensional fragments, reference and rigidity do not have to do with quantifier treatment, for the same reason they do not have to do with modal operators either. And when I say this, I have in mind natural languages. That is, reference concerns simple locutions in "Nixon is blue" or "John loves Mary," whereas quantification is semantically and logically posterior to the treatment of such simple nouns and predicates. If this is so, then the question of how to read and deal with quantifers is logically independent from the question of the semantical and logical analysis of proper names, which is prior to it. In other words, we should be free to interpret a quantifier objectually or substitutionally or blown it away altogether with no variables, with no constraints imposed by the interpretation of simple nouns and predicates. That is, there should be reference without objectual quantifiers and independently of the quantifying in into alethic, belief, or knowledge embeddings. The definability of the rigid reference of singular terms is a model-theoretical notion (in the sense of constant designation across a class of possible worlds) which may have a role to play in formal languages. I believe we were right about it in the case of formal languages with modal operators. For natural languages though, the model-theoretical expressibility is out of place and does not show the dispensability or eliminability of rigidity for a class of expressions or other.

University of Helsinki

References

Almog, Joseph (1986). "Naming without Necessity." *The Journal of Philosophy*, Vol. 83, No. 4, 210–242.

Davidson, Donald, and Gilbert Harman (eds.) (1972). *Semantics of Natural Language*. Dordrecht: D. Reidel Publishing Company.

Hintikka, Jaakko (1969). "On the Logic of Perception." In J. Hintikka, *Models for modalities*, Dordrecht: D. Reidel.

Hintikka, Jaakko (1996). "World Lines and Their Role in Epistemic Logic." In P. I. Bystrov and V. N. Sadovsky (eds.), *Philosophical Logic and Logical Philosophy*- Dordrecht: Kluwer, 121–137.

Hintikka, Jaakko, and Jack Kulas (1985). *Anaphora and Definite Descriptions*. Dordrecht: D. Reidel Publishing Company.

Hintikka, Jaakko, and Gabriel Sandu (1991). *On the Methodology of Linguistics*. Oxford and Cambridge, MA: Basil Blackwell.

Hintikka, Jaakko, and Gabriel Sandu (1995). "The Fallacies of the New Theory of Reference." *Synthese*, Vol. 104, No. 2 (Aug., 1995), 245–283.

Kripke, Saul (1963). "Semantical Considerations on Modal Logic." *Acta Philosophica Fennica* 16, 83–94.

Kripke, Saul (1976). "Is There a Problem about Substitutional Quantification?" In G. Evans and John McDowell (eds.), *Truth and Meaning: Essays in Semantics*. Oxford: Clarendon Press, 325–419.

Kripke, Saul (1980). *Naming and Necessity*. Oxford: Basil Blackwell.