A FREE LOGIC FOR FICTIONALISM[1]

MIRCEA DUMITRU

Faculty of Philosophy
University of Bucharest
Romanian Academy, Institute of Philosophy and Psychology

Abstract: In *Reference without Referents,* Mark Sainsbury aims to provide an account of reference that honours the common-sense view that sentences containing empty names like "Vulcan" and "Santa Claus" are entirely intelligible, and that many such sentences – "Vulcan doesn't exist", "Many children believe that Santa Claus will give them presents at Christmas", etc. – are literally true. Sainsbury's account endorses the Davidsonian program in the theory of meaning, and combines this with a commitment to Negative Free Logic, which holds that all simple sentences containing empty names are false. In this paper, I pose a number of problems for this account. In particular, I question the ability of Negative Free Logic to make appropriate sense of the truth of familiar sentences containing empty names, including negative existential claims like "Vulcan doesn't exist".

1. INTRODUCTION

The problem of empty or non-referring names – names that lack a referent, even though they feature in apparently meaningful discourse - continues to be one of the most hotly debated problems in philosophy of language. Following the explosion of work that attended Kripke's arguments for a direct reference account of names, there is now something approaching orthodoxy about the semantic contribution that names make to propositions and truth conditions. This orthodoxy – if truth be told, an uneasy orthodoxy whose core support is found in the US – is Millianism. According to Millianism, what names and other directly referential expressions semantically contribute to propositions are their referents (nothing more, nothing less); sentences that contain names express structured propositions, often called singular or Russellian propositions that embody such referents as constituents. Softer versions of Millianism allow that names contribute something more as well, say a mode of presentation. But of course the problem of empty names is troubling for Millians of any kind, soft or hard, since Millians appear committed to the view that empty names are unable to make a semantic contribution so that sentences containing empty names fail to say anything, let alone anything *true.*

There is now a lively industry devoted to finding Millianism-friendly solutions to this quandary. Some Millians insist on distinguishing what is expressed and what is implicated, a distinction wielded in support of the view that what we see as meaningful and even true concerns what is implicated rather than expressed (e.g., Taylor 2000). Others appeal to gappy propositions – proposition-like entities expressed by sentences containing empty names that can fail to be true because of the gaps, and become true when negated (see especially Braun 1993).[2] But these solutions all have well-documented problems. That might suggest the wisdom of a return to some version of descriptivism, and, indeed, neo-descriptivism has had a minor resurgence in some parts of the world. But neo-descriptivist accounts of names have their own familiar suite of problems, and are anathema to many.[3] In short, while the problem of empty names continues to be a thorn in the side of Millianism, Millianism continues to survive, even thrive.

In *Reference without Referents (RWR),* Mark Sainsbury takes a very different approach. His avowed aim is to describe a picture of reference that honours the common-sense view that names like "Hamlet" and "Vulcan" are indeed empty, that sentences containing such names are entirely intelligible, and that many such sentences – "Vulcan doesn't exist", "Many children believe that Santa Claus will give them presents at Christmas", and so on – are literally true. Sainsbury makes it clear that he thinks that there is something categorically flawed in Millianism's idea of singular and general structured propositions that form the content of sentences (see *R WR*, 2.2)[4]. But he thinks a return to descriptivism is not viable. The present paper gives a brief account of Sainsbury's alternative approach to these issues before presenting some problems.

2. *RWR*

The alternative that Sainsbury favors combines the following two fundamental ideas. There is, to begin with, his endorsement of the Davidsonian program in the theory of meaning, filtered through McDowell's interpretation, and this is accompanied by a commitment to negative free logic (NFL), which he takes it to be the most appropriate logical tool for coping with semantic problems in natural language. In short, Sainsbury's approach (R[eference]W[ithout]R [eferents]) is Davidsonian truth-theoretic semantics plus NFL. Now, it may seem odd that in developing RWR Sainsbury should embrace a Davidsonian position. After all, the, main issue here is that of reference, and it is well known that Davidson's attitude toward the role that reference plays when it comes to explaining the relation between

---

[2] There is even the view that the problem is exaggerated, because (most) allegedly empty names stand for abstract entities. (Something like this view is held by Nathan Salmon, himself a foremost Millian, although Salmon combines it with "the" doctrine of gappy propositions; see Salmon 1998).

[3] For one line of support for a version of descriptivism, see Jackson 1998. For a critique, see, for example, Soames 2002.

[4] Like a number of other philosophers, Sainsbury uses the name "the-theory of direct reference" in preference to "Millianism".

language and reality is one of down-sizing: for Davidson, reference is a derived rather than a primitive relation. Sainsbury gives his own reasons for being dissatisfied with this aspect of Davidson's views (in chapter 7, Sainsbury appeals to the role of non-linguistic, reference), but in the next section we give some reasons for thinking the matter goes rather deeper.

So, the framework for RWR is configured by truth-theoretic semantics plus NFL. An important part of the motivation for supporting such a combination is that they will secure for Sainsbury what he is looking for, viz. a semantic position which eventually avoids the Fregean-Millian polarity that defines the current orthodoxy in the semantics of natural language. What is the specific contribution that each theoretic assumption makes to the story that Sainsbury is telling us? Let's begin with Sainsbury's Davidsonianism. The program of truth-theoretic semantics for natural languages, initiated through the classic paper "Truth and Meaning" (1967), is one of two very important programs that Davidson founded in the field of the theory of meaning. (The other program is the project of radical interpretation, which shows how a theory of meaning should be corroborated by evidence from human behaviour.) Davidson's profound idea was that a Tarski-style axiomatic truth theory for a given language can serve as a compositional theory of meaning for the same language. To be adequate, such a (finitely axiomatizable) truth theory should allow us to infer for every (indexical-free) sentence s of the object language a biconditional (T-theorem) of the form

(T) s is true iff *p*

where s is a name or structural description of an object language sentence and p stands for a metalanguage sentence that gives the *condition* under which s is true. Davidson thinks that a theory of this kind shows how we form and understand complex expressions on the basis of our understanding of the constituent primitive expressions and of the rules of combination.

In agreeing with this fundamental approach, Sainsbury sides with Davidson and against a tradition which identifies meanings with special entities that can be grasped in our understanding. (See *RWR,* pp. 52–59). Davidson famously says that meanings as entities "have no demonstrated use" in a theory of meaning (Davidson 1967). This applies no less to model-theoretic approaches to meanings, for what is distinctive in such approaches is the spelling out of meanings in terms of semantic values. Given his belief in the intelligibility of sentences containing empty names, it is scarcely surprising that Sainsbury should prefer a Davidsonian position. Model theory is not fit to treat the meaningful contribution of referring expressions that actually refer to nothing and so can have no semantic value in themselves. Davidson's position, by contrast, can be adapted to deal with the intelligibility of such expressions.

Here, in condensed form, is Sainsbury's idea:

> I suggest that we should see the axioms for referring expressions in the same way: these expressions are associated with a reference condition which may or

may not be satisfied, just as a sentence is associated with a truth condition which may or may not be satisfied. Roughly speaking, in model theory meanings are entities, whereas in truth theory they are conditions. *(RWR, p. 53)*

On Sainsbury's implementation of this suggestion, the meaning-specification can be done by the subsidiary clauses of a Davidsonian truth-theory, with compositional rules generating (ideally, homophonic) truth conditions of entire sentences. The subsidiary clauses for names are given in the form "(x) (`a'` refers to x iff x = a)"[5], where language-users who have been initiated into the practice of using "a" are presumed to know enough to understand the condition of reference "= a", whether or not "a" has a referent. This can sound close to descriptivism, but here Sainsbury endorses an important theme from McDowell: names can very well have senses (in Sainsbury's terms, associated reference conditions) despite these lacking descriptive content. Sainsbury takes the senses/reference conditions for names to be both non-descriptive and at the same time singular in what they (purport to) pick out. He thinks this allows names to retain some of the important features Millians claim for them, in particular rigidity. McDowell too thought that the senses of names were singular. But for McDowell they are singular because they are *de re* or object-dependent. Sainsbury accepts the intelligibility of genuinely non-referring names, and so he rejects this second McDowellian theme (*RWR*, p. 85).

Names are not the only potentially non-referring referring expression that can be ascribed meaning in a way that meshes with such an extended Davidsonian approach. In an excellent chapter on "Pronouns: Anaphora and Demonstration", Sainsbury extends the account to pronouns. Consider the following brief example of how he handles empty demonstratives. In the case of someone who wrongly thinks that there is a little green man in her field of view and utters "That little green man is bald", we can ascribe content to her utterance by describing what she says in a certain canonical scene/content way: conditionalising on the scene, we can say: "Hallucinating a little green man, her utterance was true iff he was bald" (this gives a conditional truth condition), or, more simply, "Hallucinating a little green man, she said that he was bald" (*RWR*, pp. 166–168). Sainsbury thinks that RWR, unlike Millianism, has no trouble accommodating such failed thoughts:

There are as many singular thoughts on the present view as on a [Millian] view. But on this view, unlike [Millianism], a failed singular thought or utterance can be a genuine thought; or an utterance which succeeds in saying something: these have, if not truth conditions, at least conditional truth conditions, and can be reported in the standard scene/content way. (*R WR*, p. 169)

But how, in failing to refer to reality, can failed thoughts and utterances feature in truth-oriented discourse, including arguments? This brings us to the foundational part of RWR concerned with the logical principles that govern a language containing non-referring referring expressions. What RWR needs is a kind of logic that will allow, in contradistinction to classical first order logic with identity, expressions deemed intuitively intelligible even if they lack referents in their domain of

---

[5] They can't be given in the form " 'a' refers to a", since Sainsbury's preferred logic, NFL, counts any instance in which "a" is an empty referring expression as false.

interpretation. It is not surprising, then, that Sainsbury adopts a free logic as his standard logic, where a logical system *LF* is a free logic iff *LF* is free of existential presuppositions with respect to both the singular and general terms of the language of *LF*, and the quantifiers of *LF* are the standard quantifiers whose variables, of quantification ranging over existing things and nothing else. The issue is what kind of free logic.

3. *A CRUSH COURSE IN FREE LOGIC*[6]

Logic free of existential presuppositions is a branch of philosophical logic which has been developed in the last forty years. Existential presuppositions are linked with singular and general terms. Accordingly, the concept of a free logic was understood as 'logic free of existential presuppositions with respect to its singular and general terms'. Standard first order logic with '=' (FOL=) is almost fully free with respect to its general terms or predicates. There is only one exception, though, namely universal terms or predicates like 'Px ∨ ∼Px' or '$x = x$'. In (FOL=) '$(\exists x)(Px ∨ ∼Px)$' and '$(\exists x)(x = x)$' are valid. We can read the latter as 'something exists', and this seems to express a truth of ontology rather than a truth of logic.

The main concern of free logic has been existential presuppositions with respect to singular terms. For in standard FOL= we have for every singular term t and variable v the valid formula: $\models (\exists v)(v = t)$. And due to the ontological commitment of singular terms, in FOL= we have rules for quantifiers, such as existential introduction ($\exists$I) and universal elimination ($\forall$E), which are not sound if the terms do not refer to actual existing things.

Against this motivational background, an adequate definition of free logic has to include three components.[7] Thus, a logical system L_F is a free logic if and only if (iff)

(1) L_F is free of existential presuppositions with respect to the singular terms of the language of L_F;

(2) L_F is free of existential presuppositions with respect to the general terms of the language of L_F; and

(3) the quantifiers of L_F have existential import.

It is appropriate to speak about a family of systems of free logic. The distinctive feature of those systems is the fact that singular terms which are empty or non-denoting, provided they do not refer to existing things, have a legitimate place in this family of logic systems. Moreover, the theorems of a free logic system are valid even if the singular terms which occur in them are empty.

There are three types of free logic systems. The criterion according to which we can distinguish between those types is whether or not elementary sentences containing empty singular terms are true or false or else lack any truth-value at all.

---

[6] A rich resource for the topic is (Morscher and Hieke, 2001), which my exposition in this section is based upon.

[7] In giving this compact presentation of free logic systems I draw on the excellent systematic and historic synopsis one can get in (Morscher and Simons, 2001).

(FL-) A logical system L_{F-} is a *negative free logic* iff L_{F-} is a free logic and every atomic sentence of L_{F-} containing at least one empty singular term is false.

(FL+) A logical system L_{F+} is a *positive free logic* iff L_{F+} is a free logic and there is at least one true atomic sentence of L_{F+} containing at least one empty singular term.

(FLn) A logical system L_{Fn} is a *neutral free logic* iff L_{Fn} is a free logic and every atomic sentence of L_{Fn} containing at least one empty singular term has no truth-value at all.

Alongside with those three types of free logic systems, the following three semantic approaches that have been developed for free logic systems are now well-entrenched:

(S1) Semantics with a *partial interpretation function* and a *total valuation function*.

(S2) Semantics with an *inner* and an *outer domain*: this uses a *total interpretation function* and a *total valuation function*.

(S3) *Supervaluation semantics*: this type of semantics uses a *partial* and a *total interpretation function* and a *total* and *two partial valuation functions*.

SEMANTIC SYSTEMS OF FREE LOGIC

	Semantics with a partial interpretation function and a total valuation function	Inner and outer domain	Supervaluation
Interpretation function	Partial	Total	Partial & Total
Valuation function	Total	Total	Total & Two partial valuation functions

Each type of semantic system specifies its own type of models *M*. As usual, a model *M* consists in a domain *D* and in an interpretation function *I*, which is associated with a valuation function *V*. *I* is always defined on the set of descriptive symbols, i.e. non-logical predicates and individual constants of the language of that free logic system. What is distinctive for the semantics for free logic is that *I* is not supposed to assign an *existing* object to each individual constant. *I* therefore assigns to some individual constant *t* of L_F either *a non-existing object* or *no object at all*; in the second case *I(t)* remains undefined, and *I* therefore is a partial function. The valuation functions *V* which are based on the interpretation functions *I* are always defined on the set of closed formulae of L_F. They can be either total or partial.

(S1) Semantics with a partial interpretation function and a total valuation function

An M^{pitv}-model is an ordered pair. It comprises a possibly empty domain D and a partial function I^{pitv}, i.e. $M^{pitv} = (D, I^{pitv})$, such that

(1) for every individual constant t of the language of L_F: either I^{pitv} does not assign anything at all to t and $I^{pitv}(t)$ thereby remains undefined or $I^{pitv}(t) \in D$;

(2) for every n-place predicate P^n of L_F: $I^{pitv}(P^n) \subseteq D^n$;

(3) for every object $d \in D$ there is an individual constant t of the language of L_F such that $I^{pitv}(t) = d$. [The interpretation function I^{pitv} of an M^{pitv}-model provides a 'full' (or complete) interpretation of the associated domain D.]

Next we define truth and falsehood in a model M^{pitv} for every closed formula A of the language of L_F. We do this by defining a total valuation function V^{pitv} from the set of closed formulae of L_F into the set $\{T,F\}$ of truth-values as follows:

(1) $V^{pitv}(P^n t_1, t_2, \ldots, t_n) = T$ iff for every t_i $(1 \leq i \leq n)$: $I^{pitv}(t_i)$ is defined and $<I^{pitv}(t_1), I^{pitv}(t_2), \ldots, I^{pitv}(t_n)> \in I^{pitv}(P^n)$;

(2) $V^{pitv}(t_1 = t_2) = T$ iff $I^{pitv}(t_1)$ is defined and $I^{pitv}(t_2)$ is defined and $I^{pitv}(t_1) = I^{pitv}(t_2)$.

(3) $V^{pitv}(E!t) = T$ iff $I^{pitv}(t)$ is defined.

(4) $V^{pitv}(\sim A) = T$ iff $V^{pitv}(A) \neq T$;

(5) $V^{pitv}(A \rightarrow B) = T$ iff $V^{pitv}(A) \neq T$ or $V^{pitv}(B) = T$ or both;

(6) $V^{pitv}(\forall vA) = T$ iff for every individual constant t: if $I^{pitv}(t)$ is defined then $V^{pitv}(A(t/v)) = T$.

(7) $V^{pitv}(A) = F$ iff $V^{pitv}(A) \neq T$.

It is worth noticing that the interpretation of quantifiers as shown in clause (6) above is substitutional. Hence, it is obligatory that the interpretation functions of the models provide a complete interpretation. The semantic concepts of validity, logical consequence, and satisfiability are defined in the usual way.

The system L_{F-} of *negative free logic* is adequate, i.e. sound and complete, with respect to the semantics with a partial interpretation function and a total valuation function. However, by a change of clauses (1) and (2) above in the definition of the valuation function V^{pitv} we can adapt M^{pitv}-models in such a way that they can be used for proving the adequacy of systems of positive free logic (in the way done by Hughes Leblanc and Robert K. Meyer).[8]

(S2) Semantics with an inner and an outer domain

We define an M^{iod}-model as a triple: $M^{iod} = (D_o, D_i, I^{iod})$. D_o and D_i are two disjoint and possibly empty sets of objects. D_o is called the *outer domain*, and D_i is called the *inner domain*, whose union is non-empty:

---

[8] See, for instance, (Leblanc and Meyer, 1970).

(i) $D_o \cap D_i = \emptyset$
(ii) $D_o \cup D_i \neq \emptyset$.

We define D as the union: $D = D_o \cup D_i$.

The *interpretation function I^{iod}* is a total function which is defined thus:

(1) for every individual constant t of L_F, $I^{iod}(t) \in D$;
(2) for every n-place predicate P^n of L_F, $I^{iod}(P^n) \subseteq D^n$;
(3) for every object $d \in D_i$ there is an individual constant t of L_F such that $I^{iod}(t) = d$.

The valuation function V^{iod} is also total and it assigns a truth-value, i.e. T or F, to each closed formula of L_F relative to an M^{iod}-model. V^{iod} is defined recursively as follows:

(4) $V^{iod}(P^n t_1, t_2, \ldots, t_n) = $ T iff $<I^{iod}(t_1), I^{iod}(t_2), \ldots, I^{iod}(t_n)> \in I^{iod}(P^n)$;
(5) $V^{iod}(t_1 = t_2) = $ T iff $I^{iod}(t_1) = I^{iod}(t_2)$.
(6) $V^{iod}(E!t) = $ T iff $I^{iod}(t) \in D_i$.
(7) $V^{iod}(\sim A) = $ T iff $V^{iod}(A) \neq $ T;
(8) $V^{iod}(A \rightarrow B) = $ T iff $V^{iod}(A) \neq $ T or $V^{iod}(B) = $ T or both;
(9) $V^{iod}(\forall v A) = $ T iff for every individual constant t: if $I^{iod}(t) \in D_i$ then $V^{iod}(A(t/v)) = $ T.
(10) $V^{iod}(A) = $ F iff $V^{iod}(A) \neq $ T.

M^{iod}-models are used mainly for *positive free logic*. L_{F+} is adequate with respect to M^{iod}-models (as it was proved by Hughes Leblanc and Richmond Thomason).[9]

(S3) Supervaluation semantics

Meinongianism is not appealing for everybody, which makes the inner/outer domain semantics not a favorite for all. Then the question arises how to develop a semantics appropriate for a positive free logic without using the inner/outer domain semantics? *Supervaluation semantics* comes as a compelling solution to this. It starts with models of the same type as in the first approach; however, atomic sentences containing empty singular terms are allowed to be truth-valueless. But this would result in a rejection of classical laws of logic, and in order to avoid this effect, the models are 'completed'. This way, the truth-value gaps which result in the first part of the valuation process are removed.

This way we build a new type of models: $M^{sv} = (D, I^{sv})$. D is again a possibly empty set of objects and I^{sv} is a partial interpretation function like I^{pitv}. As in the case of M^{pitv}-models, the conditions we impose here on I^{sv} are the same as the conditions we imposed earlier on I^{pitv}. Thus:

---

[9] See, for instance, (Leblanc and Thomason, 1968).

14

(1) for every individual constant t of the language of L_F: either I^{sv} does not assign anything at all to t and $I^{sv}(t)$ thereby remains undefined or $I^{sv}(t) \in D$;

(2) for every n-place predicate P^n of L_F: $I^{sv}(P^n) \subseteq D^n$;

(3) for every object $d \in D$ there is an individual constant t of the language of L_F such that $I^{sv}(t) = d$.

However, unlike V^{pitv}, the valuation function V^{sv} associated with M^{sv}-models is also a partial function (like I^{sv}) and *its domain is restricted to atomic formulae* of L_F. Consequently, V^{sv} is a partial function from closed atomic formulae of L_F into the set {T,F} of truth-values; it is defined as follows:

(1a) If for any t_i ($1 \leq i \leq n$), $I^{sv}(t_i)$ is defined, then $V^{sv}(P^n t_1,t_2,\ldots,t_n) = $ T iff $<I^{sv}(t_1),I^{sv}(t_2),\ldots,I^{sv}(t_n)> \in I^{sv}(P^n)$;

(1b) If for at least one t_i ($1 \leq i \leq n$), $I^{sv}(t_i)$ is undefined, then $V^{sv}(P^n t_1,t_2,\ldots,t_n)$ is undefined.

(2a) If both $I^{sv}(t_1)$ and $I^{sv}(t_2)$ are defined, then $V^{sv}(t_1 = t_2) = $ T iff $I^{sv}(t_1) = I^{sv}(t_2)$.

(2b) If either $I^{sv}(t_1)$ or $I^{sv}(t_2)$ is undefined but the other is defined then $V^{sv}(t_1 = t_2) = $ F.

(2c) If neither $I^{sv}(t_1)$ nor $I^{sv}(t_2)$ is defined, then $V^{sv}(t_1 = t_2)$ is undefined.

(3) $V^{sv}(E!t) = $ T iff $I^{sv}(t)$ is defined, and $V^{sv}(E!t) = $ F iff $I^{sv}(t)$ is undefined.

We define now the concept of a *completion* (i.e. a complete supermodel) of an M^{sv}-model:

$M^{csv} = (D',I^{csv})$ is a *completion* of $M^{sv} = (D,I^{sv})$ iff

(1) $D' \neq \emptyset$;

(2) $D \subseteq D'$;

(3) for every n-place predicate P^n: $I^{sv}(P^n) \subseteq I^{csv}(P^n)$;

(4) for every individual constant t: if $I^{sv}(t)$ is defined then $I^{csv}(t) = I^{sv}(t)$;

(5) for every individual constant t: $I^{csv}(t) \in D'$.

(1) through (4) say that M^{csv} is a supermodel of M^{sv}, and clause (5) says that I^{csv} is a total function and M^{csv} is, accordingly, 'complete'.

Now, from the point of view of an M^{sv}-model of which M^{csv} is a completion, the valuation function V^{csv} of an M^{csv}-model is a total function from all the closed formulae of L_F into the set {T,F} of truth-values. V^{csv} therefore depends on V^{sv}. It is defined as follows:

(1) If A is a closed atomic formula of L_F and $V^{sv}(A)$ is defined, then $V^{csv}(A) = V^{sv}(A)$.

(2) If A is a closed atomic formula of L_F and $V^{sv}(A)$ is undefined, then $V^{csv}(A)$ is determined independently of V^{sv} in the usual way for complete models as follows:

(2a) If A is a closed atomic formula of the form $(P^n t_1,t_2,\ldots,t_n)$, then
$V^{csv}(P^n t_1,t_2,\ldots,t_n) = $ T if $<I^{csv}(t_1),I^{csv}(t_2),\ldots,I^{csv}(t_n)> \in I^{csv}(P^n)$, and
$V^{csv}(P^n t_1,t_2,\ldots,t_n) = $ F if $<I^{csv}(t_1),I^{csv}(t_2),\ldots,I^{csv}(t_n)> \notin I^{csv}(P^n)$.

(2b) If A is a closed atomic formula of the form $t_1 = t_2$, then $V^{csv}(t_1 = t_2) = \text{T}$ iff $I^{csv}(t_1) = I^{csv}(t_2)$.

(2c) If A is a closed atomic formula of the form E!t, then $V^{sv}(\text{E!}t)$ is always defined. Hence, clause (1) will do the job, i.e. for each individual constant t: $V^{csv}(\text{E!}t) = V^{sv}(\text{E!}t)$.

(3) $V^{csv}(\sim A) = \text{T}$ iff $V^{csv}(A) = \text{F}$;

(4) $V^{csv}(A \rightarrow B) = \text{T}$ iff $V^{csv}(A) = \text{F}$ or $V^{csv}(B) = \text{T}$ or both;

(5) $V^{csv}(\forall vA) = \text{T}$ iff for every individual constant t: if $I^{csv}(\text{E!}t) = \text{T}$ then $V^{csv}(A(t/v)) = \text{T}$.

Next, we define the *supervaluation $S(M^{sv})$* as a partial function from closed formulae of L_F into the set {T,F} of truth-values as follows:

(1) $S(M^{sv})(A) = \text{T}$ iff $V^{csv}(A) = \text{T}$ for every completion M^{csv} of M^{sv}.

(2) $S(M^{sv})(A) = \text{F}$ iff $V^{csv}(A) = \text{F}$ for every completion M^{csv} of M^{sv}.

(3) $S(M^{sv})(A)$ is undefined otherwise, i.e. iff $V^{csv}(A) = \text{T}$ for at least one completion M^{csv} of M^{sv} and $V^{csv'}(A) = \text{F}$ for at least one completion $M^{csv'}$ of M^{sv}.

Lastly, we define logical consequence in terms of supervaluations in the following way: a closed wff of L_F is *logically supertrue* iff for all M^{sv}-models M^{sv}: $S(M^{sv})(A) = \text{T}$.

A closed formula B of L_F is a *logical consequence* of a class C of closed formulae of L_F iff for all M^{sv}-models M^{sv}: if $S(M^{sv})(A) = \text{T}$ for each $A \in C$, then $S(M^{sv})(B) = \text{T}$.

A set C of closed formulae of L_F is *supersatisfiable* iff there is at least one model M^{sv} such that $S(M^{sv})(B) = \text{T}$ for each $B \in C$.

Bas van Fraassen has used supervaluation semantics in order to prove soundness and completeness of *positive free logic with =*.[10]

4. RWR (CONT'D)

We have already mentioned that Sainsbury opts for *negative free logic* (NFL). This is a system in which elementary sentences containing empty singular terms are always false. (In *positive free logic,* by contrast, they are sometimes true, while in *neutral* or *Fregean free logic* they always lack a truth-value.) Given Sainsbury's choice of a Davidsonian meaning theory, we should not expect the model theory for NFL to tell us about the meaning of expressions; it should, however, tell us about logical consequence and validity, and ought to do so by taking proper account of the fact that some referring expressions refer to nothing whatsoever, not even to non-real objects of some kind. In particular, then, our *intended* model-theoretic semantics better not involve both an inner and an outer domain. Fortunately, this can be done. NFL can be shown to be sound and complete with respect to its intended semantics, one that involves *partial* interpretation functions and *total* valuation functions.

---

[10] See (van Fraassen, 1966).

Like Sainsbury, we think that there is much to be said on behalf of NFL, even though we think Sainsbury overestimates its role when it comes to explaining our intuitions about truth values. Sainsbury himself is appreciative about NFL as involving a minimal departure from classical logic. Among other arguments for NFL, he is keen to appreciate the virtues of the resulting account of the truth conditions of negative existentials. This account he describes as "dazzlingly straightforward" (p. 195). According to NFL, simple predications like "Vulcan exists" and "Hamlet exists" are false because the names they embed are non-referring. Hence their negations, "Vulcan doesn't exist" and "Hamlet doesn't exist", come out as true – just what is wanted. (The moves here parallel the moves used by Millian advocates of gappy propositions.)

This is a brief sketch of some of the main features of Sainsbury's RWR. Sainsbury goes on to argue that RWR can not only be applied to proper names and anaphoric pronouns and demonstratives, but also to plural names and even to singular and plural definite descriptions. In the penultimate chapter, he argues that it provides a simple and promising account of fiction. (We have already seen how RWR tackles negative existential like "Hamlet doesn't exist") And the final chapter extends the theory to the case of mental, reference. The result is an excellent book, innovative, written with care for detail, yet fully willing to acknowledge that RWR as it stands can't be the final word, that problems remain. We next turn to some of these problems, focusing mainly on the book's *leitmotif* (the central case of empty names and the semantic problems they present). More specifically, we will consider some methodological issues to do with the choice of a Davidsonian framework for understanding how names can be meaningful in the absence of things to name. But we will also consider the choice of NFL as the logical vehicle for assigning truth values to sentences containing empty referring expressions, and particular problems that arise from the way this choice plays out for RWR.

5. *PROBLEMS FOR RWR*

We begin with the question, already foreshadowed, to what extent a Davidsonian framework is the best framework for Sainsbury's purposes. For Davidson, the theory of radical interpretation shows that reference is a derivative, indeed an instrumental, relation. On this approach, there are no underlying referential facts that determine the right truth conditions for our sentences. Sainsbury challenges the derivativeness of reference in his last chapter (he thinks non-linguistic animals may have a capacity for non-linguistic reference), although he thinks that "the success of this challenge is not essential to the official theses of RWR" (p. 59). But there is a sense in which a view like RWR sits unhappily with the thought that reference is derivative, that the assignment of truth conditions in the interests of proper interpretation is primary. This is especially so given the adoption of NFL, for when it comes to NFL, knowing and grasping (conditional) truth conditions is not even enough for knowing whether an expression has a referent. Knowing the latter requires us to know whether anything is determined as the referent of the expression. It requires us to know a basic referential fact.

This brings us to a second question. RWR does not advocate an official theory of how reference is determined (for example in the case of names). Like some familiar versions of Millianism, it seems to regard this as a kind of pre-semantic matter. Its official semantic theory holds that referring expressions, including ones that fail to secure referents, are associated with reference conditions rather than referents, and that these conditions are public. But this leaves it unclear just what is publicly known when speakers and hearers grasp reference conditions, and how what is publicly known feeds into certain crucial abilities they have with names, such as the ability to work out (perhaps with the help of others) who is referred to with a name, and whether a name really has a referent.

It is arguable that these are not accidental abilities, but part and parcel of the reason why we have name-using practices: such abilities are needed if name-using practices are to fulfil their role of allowing the acquisition and sharing of information about the referents of names. (Sainsbury's chapter on names suggests that he agrees with the centrality of these features of name-using practices.) Descriptivism, say as understood by Frank Jackson, places great stress on such features, which he regards as an essential part of the public profile of names and part of what makes names the public currency they are. Millianism understands the public nature of names – what we assert as of right when using a name, and what we grasp in acquiring the ability to use a name in a very different way. Sainsbury rightly points to problems with such views, but one wishes that RWR had more to say about how public knowledge of semantic reference conditions is connected to features of name-using practices highlighted by other traditions. Until this is done, adherents of these other traditions may well suspect that Sainsbury's account of the nature and role of reference conditions will in the end reduce to something closer to their own account of the semantics of referring expressions.

These methodological points aside, let us return to the actual details of the way in which RWR approaches the issue of empty referring expressions. Sainsbury's avowed goal is to make a convincing case for the claim that RWR offers the right framework for understanding the semantic behavior of all such terms, no matter what the context. How successful is his defense of RWR? By focusing on some of its notable successes, Sainsbury certainly establishes that RWR gets us a long way. One such success is the account of the truth conditions of negative existentials. A second success concerns fiction. As the chapter on fiction and existence reminds us, RWR has nothing to fear from truths like "Holmes is a detective". The fact that NFL counts this fictional sentence as strictly false is irrelevant, for to the extent that we regard the sentence as true, it should be taken as (implicitly) prefixed with an "in the fiction" operator. Because this leaves the fictional name shielded by an intensional operator, NFL can no longer be used to show that the sentence is false. Sainsbury acknowledges that extra-fictional truths such as "Tony Blair admires Coriolanus" pose more of a problem, and he sees these as requiring more work. (He adamantly rejects views that appeal to special fictional objects in order to cope with such examples.)[11]

---

[11] A famous example is Kripke's view of such contexts in Kripke 1973.

We think, however, that the difficulties run rather deeper than Sainsbury makes out. As the problem of extra-fictional truths like "Tony Blair admires Coriolanus" shows, what makes it so hard to get by on the thin ontological diet recommended by RWR is that much of what we want to say using fictional and other seemingly empty referring expressions seems to carry a straightforward commitment to non-existent objects. Sainsbury hopes that he can find analyses that isolate such terms behind appropriate intensional operators, but as a general strategy this seems dubious. The problem even arises in the case of the scene content style of reports that we mentioned earlier:

(1) Hallucinating a little green man, she said that he was bald.

We think that such constructions prove too much, for they are prone to a curious kind of leak: reporters can exploit elements of the scene in order to say things that go beyond mere content. Suppose, for example, that the reporter wishes to make explicit the way the reported speaker executed her demonstration. He can do this by saying:

(2) Hallucinating a little green man in the corner of the garden, she pointed at him, and said that he was bald.

This way of describing the reported speaker's contribution seems no less appropriate (indeed, it is more informative, and in certain cases such reporting may be necessary, as when the reported speaker hallucinated too many objects and the reporter wants to signal how the speaker discriminated "them"). But Sainsbury's acceptance of NFL means he would count a report like (2) as false since it incorporates a simple relational claim – "She pointed at him" – that involves an empty term. We think that Sainsbury's account thus fails to capture a sense in which, by uttering (1) and (2), our reporter shows himself to be entering into commitments that, at another level, he repudiates.

We think that a version of this problem even affects Sainsbury's "dazzlingly straightforward" solution to the problem of true negative existentials. Recall that a sentence like "Vulcan doesn't exist" comes out as true on NFL because the simple sentence "Vulcan exists" is false, "Vulcan" being non-referring. But as Sainsbury acknowledges, there are problems with this kind of explanation. One is that a statement like "Vulcan is not a planet" should then also count as true (and for exactly the same reason); and this seems wrong. Note that we can't retreat to the view that our sense of falsity in this case can be explained in terms of the falsity of the internally negated claim, "Vulcan is a non-planet". That would suggest that we should hear "Vulcan is non-existent" as similarly false, which we don't.

But there is what seems to us a deeper problem. Sainsbury's account is an account of the simplest kind of negative existential (cases of the form "*N* doesn't exist"). Negative existentials are often much more complex, and often in ways that threaten the role of NFL.

Suppose, for example, that we utter the simple negative existential

(3) Vulcan does not exist,

a sentence that Sainsbury counts as true. But our utterance of (3) may well not exist in isolation. It may prompt a request for clarification. ("What do you mean it doesn't exist? Is it a *fictional* planet, maybe?") That is why we might add:

(4) Vulcan is one of the many failed posits of 19th century science: it was a planet posited to explain certain astronomical phenomena, but something other than the movement of Vulcan provided the correct explanation.

Note, however, that (4) is false from the point of view of NFL, since it contains as a conjunct a simple affirmative statement involving an empty name ("Vulcan was a planet posited to explain certain astronomical phenomena"). Yet (4) is surely no more contentious than (3). It simply provides more information, clarifying or perhaps explaining what makes (3) true. In some sense, its truth is what grounds the truth of (3). (Note that NFL's reasons for classifying (4) as false go even further. For (4) contains apparent quantification over non-existent things [failed posits of 19th century science], and for NFL all quantification is over existent things; there is nothing else to quantify over. Assuming that such quantification is objectual rather than substitutional, the apparent truth of claims like (4) yields another challenge to NFL.)

"(4)" is not the only way we might clarify a negative existential like (3). At other times we might wish to clarify just what we are talking about by offering an alternative way of identifying our target object:

(5) Vulcan – you know, the planet described in that interesting paper by Le Verrier that we have been reading – does not exist.

Suppose, then, that "(5)" is true, that it correctly provides an alternative way of identifying Vulcan. But how should we understand the logical role of the relative clause? We see just two options. The first is to read (5) as a simple conjunction:

(6) Vulcan is the planet described in that interesting paper by Le Verrier that we have been reading, and it [Vulcan] does not exist.

The second is to treat the relative clause as part of a complex definite description that also features the name "Vulcan":

(7) The thing x such that (x = Vulcan and x is the planet described in that interesting paper by Le Verrier that we have been reading) does not exist.

Which reading is right? Given that "(5)" is true, the correct reading should be true as well. "(7)" seems promising, for the description it uses is non-referring, which makes "(7)" true by the lights of NFL. But note that this reading of "(5)" is quite unintended, since "(3)" by itself entails "(7)" and any other negative existential of the form "The x [x = Vulcan and (φ)] does not exist". For according to NFL if "Vulcan does not exist" is true, that is because the name "Vulcan" is empty. As a result, any compound description "The x [x = Vulcan and φ]" will be empty as well, making the sentence "The x [x = Vulcan and φ] doesn't exist" also true. But clearly "(3)" may be true without "(5)", or variants of "(5)", being true (we might not have been reading any paper by Le Verrier). Hence "(7)" can't be the right reading of "(5)".

That leaves us with "(6)" as the only viable reading of "(5)". And now the problem is clear. "(6)", and hence "(5)", is false according to NFL, since one of its conjuncts is false. In short, even if NFL does well on simple negative existentials like "(3)", it does poorly on complex negative existentials like "(5)".

Just how to understand negative existentials in the light of such problems remains a deep challenge for any theory that tries to be ontologically austere in the manner of RWR. (They are no less a problem for Millian attempts to deal with empty names, of course. Our aim has been to show that RWR's reliance on NFL means that it doesn't easily escape them.) We think that much more should be made of the intuition that in uttering such statements we speculatively engage with the commitments of those whose commitments we don't share: in some sense, a speaker who engages in "Vulcan" talk – even when her aim is to disown such talk – is going along with the thought that there is such a thing. A related suggestion has been made by Wiggins (although only in the context of simple negative existentials), and is sympathetically discussed by Sainsbury (*RWR*, pp. 198ff). Applied more widely, such a suggestion may get us close to some kind of pretense theory. But Sainsbury adamantly turns his face against the kind of pretense theory that Evans, for example, offers us (he rejects the latter's refusal to count a name like "Vulcan" as genuinely, intelligible rather than merely quasi-intelligible) and he rejects the Millianism of Ken Walton's version of pretense theory[12]. It remains unclear, therefore, how RWR should go from here.

Whatever our view on this debate, one can't help but be impressed by RWR's single-minded focus, and, in particular, by the way it doggedly sticks to the view that names like "Hamlet", "Vulcan", and "Pegasus" are genuinely non-referring, and then tries to understand the semantic behavior of such terms in a way that invokes nothing more than the semantic machinery and ontology needed for ordinary terms. Sainsbury should be congratulated for articulating this framework as clearly and honestly as he has, and for pushing it as far as he has.

6. *WHAT'S NEXT?*

There we have some suggestions from Kit Fine's *Semantic Relationism* (*SR*). The phenomenon that we have here is that in which the intended referent of a name does not exist, or that in which two uses of a name are taken to be one (which is "confused" reference). Here, I am interested in the first aspect: when one uses a name, one intends a referent which does not exist.

This is a case of "defective" semantics – in which the requirements laid down for the use of an expression cannot be met. Fine suggests here that in such cases a "backup" semantics comes into effect.

---

[12] See Evans 1982, ch. 10, and Walton 1990, especially chs. 10 and 1. Walton thinks that speakers merely *pretend* to affirm propositions when they knowingly use fictional names in declarative sentences, an idea also found in Kripke 1973.

" … given that the original requirements cannot be met, they are replaced with suitably related requirements which can be met. Thus instead of failed reference to an ordinary object, we have successful reference to an intentional object". (*SR*, p. 126)

How are we going to describe the functioning of a defective semantics? We need to distinguish between the case in which an unsuccessful attempt is made to lay down a semantic requirement and the case in which no attempt is even made.

Here we might appeal, to begin with, to the notion of a *putative* requirement. This putative requirement that *S* may succeed. Then, it results in a genuine requirement that *S*. But the putative requirement may fail. Then, in that case, it is subject to a "back-up" requirement that *S'*, where *S'* and *S* are suitable related.

The whole idea is encapsulated now into this: "If the putative requirement that *S* fails then it will not be a genuine requirement that *S* and usually it will not even be the case that *S*. It will, therefore, be impossible for the speaker to know that *S*, although he may know that it is a putative semantic requirement that *S* and he may also know, if he is sufficiently enlightened, that it is a genuine semantic requirement that *S'*." (*SR*, p. 127)

REFERENCES

Braun, D., 1993, "Empty Names", *Nous*, vol. 27, no. 4, pp. 449–469.
Davidson, D., 1967, "Truth and Meaning", *Synthese*, vol. 17, no. 1, pp. 304–323.
Evans, G., 1982, *The Varieties of Reference*, Clarendon Press, Oxford.
Fine, Kit, 2007, 2009, *Semantic Relationism*, Wiley-Blackwell.
Jackson, F., 1998, "Reference and Descriptions Revisited", *Philosophical Perspectives*, vol. 12, pp. 201–218.
Kripke, S., 2013, *Reference and Existence: The John Locke Lectures for 1973, Oxford*.
Morscher, Edgar & Alexander Hieke (eds.), 2001, *New Essays in Free Logic. In Honour of Karel Lambert*, Kluwer Academic Publishers.
Saynsbury, R., 2005, *Reference Without Referents*, Clarendon Press, Oxford.
Salmon, N., 1998, "Nonexistence", *Nous*, vol. 32, no. 3, pp. 277–319.
Soames, S., 2002, *Beyond Rigidity*, Oxford University Press, Oxford.
Taylor, K., 2000, "Emptiness Without Compromise", in Anthony Everett and Thomas Hofweber (eds.), *Empty Names, Fiction, and the Puzzles of Non-Existence*, CSLI Publications, Stanford, pp. 17–36.
Walton, K., 1990, *Mimesis as Make-Believe*, Harvard University Press, Cambridge, Mass.