Substitution Puzzles and Substitutional Semantics

Bartosz Wieckowski

Institut für Philosophie &
Zentrum für Logik, Wissenschaftstheorie und Wissenschaftsgeschichte
Universität Rostock

Wilhelm-Schickard-Institut für Informatik
Universität Tübingen

bartosz.wieckowski@uni-tuebingen.de

Abstract

Substitution failure is usually said to occur when a change from one *co-referential* name to another affects the truth value of the sentence. Taking the idea seriously that *non-referring names* like 'Superman' and 'Clark Kent' are indeed *names* and that they do *not refer* (to anything whatsoever), it is argued that the usual characterization does not only give rise to theories which inflate ontology but that it falls short of capturing substitution puzzles about constructions which involve non-referring terms. The paper suggests a more general characterization of substitution failure and proposes an account of (ignorant and enlightened) anti-substitution intuitions which does not invoke referents at all. The account is made formally precise in terms of *associative substitutional semantics*.

1 Introduction

Typically, substitution failure is characterized as follows:

(SF) Substitution failure occurs when a change from one *co-referential* name to another affects the truth value of the sentence.

Accordingly, in a sentence like (1) a substitution of 'Mark Twain' by 'Samuel Clemens' does not appear to be truth preserving, even though both names are supposed to be co-referential.

(1) Ann believes that Mark Twain is a writer.
(1*) Ann believes that Samuel Clemens is a writer.

As first observed in Saul (1997), anti-substitution intuitions are by no means restricted to standard opacity inducing constructions, e.g., to attitude reporting sentences like (1), since they seem to be triggered also in the case of "simple sentences" (i.e., sentences

Grønn, Atle (ed.): *Proceedings of SuB12*, Oslo: ILOS 2008 (ISBN 978-82-92800-00-3), 645–662.

which do not contain standard opacity producing expressions like 'believes') such as (2), when the names 'Clark Kent' and 'Superman' are exchanged.

(2) Clark Kent went into the phone booth and Superman came out.
(2*) Superman went into the phone booth and Clark Kent came out.

In agreement with (SF), attempts to explain our intuitions of substitution-resistance typically appeal to the notion of reference. This is the case with Millian (or referentialist) approaches, Fregean approaches, and, as far as I can see, with the rest.[1]

From the intuitive point of view, though, (SF) cannot be sensibly applied to (2) and (2*), as 'Clark Kent' and 'Superman' are *non-referring* terms, and, therefore, just cannot be co-referential. So it seems that reference guided approaches can do justice to our anti-substitution intuitions only at the price of violating our anti-denotation intuitions. What is more, by introducing the referents which are needed for referential theories to take off the ground, reference guided theorists inflate ontology with (more or less perplexing) referents for non-referring terms (i.e., fictional objects or abstract surrogates for them).

Furthermore, the expressions 'Clark Kent' and 'Superman' are intuitively classified as *names*—just like 'Mark Twain' and 'Samuel Clemens'—rather than as (disguised) definite descriptions. Similarly, the first conjunct of (2), e.g., 'Clark Kent went into the phone booth' seems to have the logical form and the truth conditions of an atomic sentence—just like 'Mark Twain went into the phone booth'—rather than those of a complex existential quantification. It seems, thus, that Russell's familiar analysis of sentences in which non-referring terms occur can avoid ontological inflation only by defying intuitions about the logical shape of expressions.[2]

The situation for the reference guided theorist is problematic also with respect to attitude-reporting clauses such as (1) and (1*), even though no non-referring names are contained therein. The problem here is that on a semantical analysis of these sentences in terms of a—philosophically interpreted applied—denotational semantics (e.g., in terms of counterpart theory or in terms of a standard possible worlds semantics for a first-order intensional language) the names 'Mark Twain' and 'Samuel Clemens' do not refer to their actual referent, but to an individual representative of some sort which is contained in the objectual domain of the intended model (e.g., a *possibile*, that is, a

[1]For a recent overview on the positions and the literature on substitution issues see, e.g., Saul (2007) and the references therein. It should be mentioned that reference guided theorists must not be confused with Millians who take it that the sole semantic function of a name is to refer to its bearer. All Millians are reference guided theorists, but the converse does not hold. Fregeans, e.g., are not referentialists, as they think that what is relevant for the truth value of opacity inducing constructions like (1) and (1*) is not the (customary) referent of the names but their sense. However, they are reference guided, since for them (1) and (1*) differ in truth value just because 'Mark Twain' and 'Samuel Clemens' *refer* to different senses within these sentences.

[2]Some theorists take sentences like (2) and (2*) which contain non-referring names to be false (notably, Russell) or, alternatively, to lack a truth value. But it seems that we can safely assume that those who take the transition from (2) to (2*) to be puzzling will not share this view, as they take the puzzle to be triggered by the very fact that the former seems to be true whereas the latter seems false.

possible but non-existing object, or an actualistically acceptable surrogate of the actual referent). In this case our intuitions of denotation are violated again, but now the violation does not consist in the fact that we are forced to foist reference upon non-referring terms, but in that we are invited to take such terms to refer to the surrogates contained in the intended domain of discourse rather than to the intuitively correct referents. (Note that a shift of reference of this kind does not occur when the sentences embedded in (1) and (1*) are considered in isolation.) Again, the referential perspective gives rise to ontological inflation and to various additional problems concerning the newly introduced referents (e.g., to problems with the identity criteria for *possibilia* or with trans-world identity). Of course, the situation for the Fregean is exactly analogous to the situation of the applied denotational semanticist, since for her, as has been noted already, the names in (1) and (1*) refer to the senses of these names rather than to their intuitive referent.

Substitution puzzles which involve non-referring terms strongly suggest that the referential conception of substitution failure is too narrow; (SF) simply fails to capture the puzzle about (2) and (2*). Such cases indicate, I think, that the reference of the exchanged terms does not play any essential role in the formulation of substitution puzzles. By the same token, the truth of identity statements like ($2^\star$) seems not to be relevant for the creation of substitution puzzles either, as identity is a relation that obtains between objects (here the referents of the exchanged terms), and there simply are none in this case.

($2^\star$) Clark Kent = Superman.

Moreover, a satisfactory account of anti-substitution intuitions should preferably not give rise to the above mentioned metaphysical problems engendered by referential accounts of intensional constructions such as (1) and (1*).

Now it seems that our anti-substitution intuitions are indeed correct in cases in which we are ignorant about the synonymy of the names exchanged. And, perhaps, they are also correct in cases in which we know that they are synonymous. In this paper I shall suggest a characterization of the notion of substitution failure which does not appeal to the reference of the exchanged terms and outline a semantic account of ignorant and enlightened anti-substitution intuitions from a non-referential perspective on the issues. This explanation will be made formally precise in terms of the framework of what I call *associative substitutional semantics*. The distinctive philosophical feature of this semantics is that it supports a view of the relation between language and the world according to which a sentence can be true without being true about something! The associative framework is, thus, in opposition to the exclusively referential (or denotational) conception of semantics which, in a sense, is dominating since Tarski:

> "We shall understand by *semantics* the totality of considerations concerning those concepts which, roughly speaking, express certain connexions between expressions of a language and the objects and states of affairs *referred* to by

these expressions."[3]

The structure of the remainder of the paper is as follows: Section 2 suggests a non-referential account of substitution failure. In Section 3 the formal framework of associative semantics is presented, and Section 4 addresses a couple of substitution puzzles from a non-referential perspective.

2 A more general conception of substitution failure

On the picture of the relation between language and the world on which the account of the puzzles to be given is based, every referring name has a referent to which it refers and a sense which it expresses, where the sense serves to determine the referent of the name. In contrast to the Fregean conception of sense, we take the sense of a name to be captured by the *nominal definition* of the name. For referring names this definition involves a list of descriptions which contains all the information which may be taken to be captured also by the real definition of its bearer.[4] In case the name does not refer, the nominal definition contains all the information which is associated with that name in the relevant piece of fiction (or discourse). For instance, the nominal definition of the referring name 'Hesperus' will be a list 'the brightest object visible in the evening sky, the brightest object visible in the morning sky, a planet, ...'. Importantly, as on the present account 'Phosphorus' will have the same nominal definition as 'Hesperus', our conception of sense differs crucially from the Fregean.[5]

In addition to the referential portion of the picture, a referring name does also have a *sense-extension* associated with it, which, again, is determined by the nominal definition of the name. Intuitively, the sense-extension of a (referring or non-referring) name contains all the information which is compatible with the nominal definition for that name, whereas the sense-extension of an elementary predicate (independently, of whether it has a referential extension or not) contains all the information associated with that predicate. More precisely, the sense-extension of a name is, in effect, a set of elementary (or pure) atomic sentences which contain that name and which are compatible with its sense. The sense-extension of a predicate, on the other hand, is a consistent set of atomic sentences which contain that predicate. The sense-extension of the name 'Hesperus', e.g., will be {..., Hesperus is a planet, ..., Hammurabi likes Hesperus, ...}. Non-referring names do only have a sense and a sense-extension but no referents. We say that a referring name refers to its bearer and that it *reflects* its sense-extension. A

[3] Tarski (1983), p. 401; my emphasis. (There are several variants of this characterization of semantics in Tarski's writings.)

[4] We take it that the nominal definition of a name is more fundamental than the real definition of the bearer in that we assume that the bearer is defined by a definition of its name. An account of the nominal definitions of numerals and the real definitions of numbers along these lines is suggested in Fine (2002), p. 16.

[5] The nominal definition of 'Hesperus' and 'Phosphorus' will be also that of 'Venus'. I discuss the semantical behaviour of homonyms, like, e.g., 'Venus (planet)' and 'Venus (goddes)' in more detail in Wieckowski (2008).

non-refering name does only reflect. The picture for elementary predicates like '... is a planet' is essentially analogous.[6]

We call the totality of the sense-extensions of the names and the predicates of the language the *level of sense*. The elementary atomic sentences contained in the sense-extensions are not true, they serve to define truth. An elementary atomic sentence is said to be true with respect to the level of sense exactly if the sense-extensions of all the terms which occur in it match with respect to that sentence. The models of the associative substitutional framework (see Section 3) are intended to represent the level of sense of the language.[7] Its *level of reference*, by contrast, can be taken to be represented by the customary denotational first-order models. And the notion of truth with respect to that level can be explained along the familiar lines of denotational first-order semantics. The language which will be discussed in what follows will be a (in the relevant cases intensional) first-order fragment of English.[8] (All this will be made more precise in the next section.)

In view of the narrowness of (SF) and the metaphysical problems with reference guided theories, I suggest that we look to the level of sense in order to explain anti-substitution intuitions. The fact that sense-extensions are sets of sentences will commit us only to an ontology to which we seem to be already committed.[9] And the fact that every name reflects a sense-extension, whereas not every name refers to a referent, will guarantee a greater generality of the reflection-guided account of substitution failure. So we replace (SF) with the following characterization:

> (SFg) Substitution failure occurs when a change from one name to another—both being *substitutionally identical*—affects the truth value of the sentence.

In contrast to the theories which support (SF), this proposal explains the synonymy of names in terms of their substitutional identity rather than (as most reference-guided

[6]We may distinguish, e.g., between the sense-extensions which are objectively associated with the names and predicates (in the same way in which referring names and referential predicates are associated with their referents and referential extensions, respectively), those portions of the objective sense-extensions which a subject associates with these terms, and those which are inter-subjectively associated with them. Unless indicated otherwise, we take sense-extensions to be objective sense-extensions in what follows.

[7]Although we shall assume in what follows that the elements of sense-extensions are sentence types (of a first-order fragment of English), this assumption is not mandatory. Alternatively, granted sufficient syntactic structure, one might take these elements to be, e.g., structured propositions of some language-like system of (neo-Fregean) concepts (see, e.g., Peacocke (1999), pp. 126-127), or, in the case of subjective sense-extensions, formulae of a person's *lingua mentis* (see, e.g., Fodor (1975)).

[8]I discuss the picture outlined above at greater length in Wieckowski (2008). There and in Wieckowski (ms) it is argued that fictional and modal truths are truths with respect to the level of sense rather than truths with respect to the level of reference, and that only sentences which do neither contain non-referring terms, nor intensional operators can be sensibly evaluated with respect to the level of reference.

[9]Of course, when we commit ourselves to something like the alternatives mentioned in footnote 7 this will no longer be so. Note in this connection that we do not introduce (object-language) referents of any sort (i.e., no ficitional referents, no *possibilia*, nor abstract individual representatives and so on). Note also that an ontology of aspects or modes of personification of referents (for discussion see Saul (2007), chapter 2) is not introduced either. This will be seen more clearly in the analyses of Section 4.

approaches have it) their co-reference. We say, e.g., that the statement of substitutional identity ($2^\dagger$) is true with respect to the level of sense just in case every elementary atomic sentence B which is just like sentence A except for containing an occurrence of ‘Clark Kent’ at all or some places where A contains ‘Superman’ the following holds: A is true with respect to the level of sense just in case B is true with respect to that level.

($2^\dagger$) Superman $\ddot{=}$ Clark Kent.

3 A fine-grained substitutional semantics

We shall now provide the picture sketched in the previous section with a formal underpinning.[10]

1. The substitutional language $\mathcal{L}$. Following Kripke (1976), we distinguish the substitutional language proper from its base language. The alphabet of the *base language* $\mathcal{L}_0$ contains nominal substitutional constants a, b, c, ... (metavariables: α, β, γ, ...), pure n-ary predicates F^n, G^n, H^n, ... with $n \geq 1$ (metavariables: φ^n, χ^n, ψ^n, ...), and the substitutional identity predicate $\ddot{=}$. (The symbols of the first two categories can appear with subscripts.) We let $\mathcal{C}$ be the set of all nominal constants of $\mathcal{L}_0$ and $\mathcal{P}$ the set of all pure predicates of that language. We do not count $\ddot{=}$ as a pure predicate.

The notion of a sentence of the base language is defined in the usual inductive manner giving us sentences of the form $\varphi^n\alpha_1...\alpha_n$ (pure atomic sentences) and $\alpha_1 \ddot{=} \alpha_2$ (where the constants need not be distinct). We let Atm be the set of pure atomic sentences of $\mathcal{L}_0$. Moreover, we define the sets $Atm(\alpha)$ and $Atm(\varphi^n)$ as follows:

$Atm(\alpha) =_{df} \{A \in Atm$: A contains at least one occurrence of the nominal constant $\alpha\}$.

$Atm(\varphi^n) =_{df} \{A \in Atm$: A contains an occurrence of the predicate $\varphi^n\}$.

The *extended language* $\mathcal{L}$ (i.e., the substitutional language proper) extends the alphabet of $\mathcal{L}_0$ with nominal substitutional variables $x, y, z, ...$, the universal substitutional quantifier symbol Π, the logical connectives $\neg$ (negation) and $\wedge$ (conjunction), and with parentheses. We let $\mathcal{V}$ be the denumerable set of nominal variables and we let the set of nominal terms of $\mathcal{L}$ be the union of $\mathcal{C}$ and $\mathcal{V}$. We let $o, o_1, ..., o_n, p, p_1, ..., p_n, ...$ be variables ranging over nominal terms. (Calling the terms “nominal” we deviate from the terminology of “individual” terms, as the semantics will not employ individual domains.)

Atomic formulae of $\mathcal{L}$ have the shape of either $\varphi^n o_1...o_n$ (pure atomic formulae) or $o_1 \ddot{=} o_2$ (substitutional identity formulae; the terms need not be distinct). The set of $\mathcal{L}$-formulae (metavariables: $A, B, C, ...$) comprises atomic formulae and formulae of the forms $\neg A$

[10] This section overlaps with the presentations of the associative framework in Wieckowski (2008) and Wieckowski (ms). There I also discuss the differences between this framework and standard substitutional (or truth-value) semantics (cf. Leblanc (1976)).

(negations), $A \wedge B$ (conjunctions), $(\Pi x)A$ (substitutionally quantified universal formulae) and also formulae which are composed from defined connectives $A \vee B$ (disjunctions), $A \rightarrow B$ (conditionals), $A \leftrightarrow B$ (biconditionals), and $(\Sigma x)A$ (substitutionally quantified existential formulae).

2. Semantics. An *associative substitutional model* is a triple $\mathcal{I} = \langle \mathcal{C}, \mathcal{P}, v \rangle$, where $\mathcal{C}$ is a non-empty substitution class of nominal constants of $\mathcal{L}_0$ and $\mathcal{P}$ is the set of pure predicates of $\mathcal{L}_0$. Recall that $\mathcal{P}$ does not contain $\ddot{=}$! The assignment v is defined as follows: $v : \mathcal{C} \rightarrow \wp(Atm)$ such that $v(\alpha) \subseteq Atm(\alpha)$ and $v : \mathcal{P} \rightarrow \wp(Atm)$ such that $v(\varphi^n) \subseteq Atm(\varphi^n)$. We call the semantic values $v(\alpha)$ and $v(\varphi^n)$ *associates*.

When $\mathcal{I} = \langle \mathcal{C}, \mathcal{P}, v \rangle$ is a model and $\mathcal{V}$ is the set of nominal variables, the assignment to nominal variables σ based on that model is a mapping: $\sigma : \mathcal{V} \rightarrow \mathcal{C}$. Thus for any $x \in \mathcal{V}$, $\sigma(x) = \alpha$ where α is a nominal constant in $\mathcal{C}$ of that model. When σ and τ are two nominal variable assignments, σ and τ are *x-variants* just in case for all nominal variables y except at most x, $\tau(y) = \sigma(y)$.

Let $\mathcal{I} = \langle \mathcal{C}, \mathcal{P}, v \rangle$ be a model and σ an assignment in $\mathcal{C}$. Then for any nominal term o the *term value* of o with respect to v and σ, $v_\sigma(o)$, is defined as follows:

$$v_\sigma(o) = \begin{cases} v(o) & \text{if } o \text{ is a nominal constant} \\ v(\sigma(o)) & \text{if } o \text{ is a nominal variable.} \end{cases}$$

We now define truth in a model $\mathcal{I} = \langle \mathcal{C}, \mathcal{P}, v \rangle$ with respect to some nominal variable assignment σ as follows:

1. $\mathcal{I}_\sigma \models \varphi^n o_1 ... o_n$ iff (i) if $o_1, ..., o_n$ are nominal constants, then $o_1, ..., o_n \in \mathcal{C}$ and if they are nominal variables, then $\sigma(o_1), ..., \sigma(o_n) \in \mathcal{C}$ and (ii) $\varphi^n o_1 ... o_n \in v_\sigma(o_1) \cap ... \cap v_\sigma(o_n) \cap v(\varphi^n)$; otherwise $\mathcal{I}_\sigma \not\models \varphi^n o_1 ... o_n$.
2. $\mathcal{I}_\sigma \models o_1 \ddot{=} o_2$ iff for all pure atomic formulae B_1 and B_2 where B_2 is like B_1 except for containing occurrences of the term o_2 at one or all places where B_1 contains the term o_1: $\mathcal{I}_\sigma \models B_1$ iff $\mathcal{I}_\sigma \models B_2$.
3. $\mathcal{I}_\sigma \models \neg A$ iff $\mathcal{I}_\sigma \not\models A$.
4. $\mathcal{I}_\sigma \models A \wedge B$ iff $\mathcal{I}_\sigma \models A$ and $\mathcal{I}_\sigma \models B$.
5. $\mathcal{I}_\sigma \models (\Pi x)A$ iff for every x-variant τ of σ: $\mathcal{I}_\tau \models A$.[11]

The clauses for disjunctions, conditionals, biconditionals, and substitutionally quantified existential formulae are given in the obvious way. For example, the clause for the latter is: $\mathcal{I}_\sigma \models (\Sigma x)A$ iff for some x-variant τ of σ: $\mathcal{I}_\tau \models A$.

A sentence A of $\mathcal{L}$ is *true in* a model $\mathcal{I} = \langle \mathcal{C}, \mathcal{P}, v \rangle$ (in symbols: $\mathcal{I} \models A$) iff it is true in that model under all assignments to the nominal variables.

[11] Associative truth clauses exclusively for sentences of $\mathcal{L}$ are offered in Wieckowski (ms). There the clauses for substitutional quantifications take a form which is familiar from truth-value semantics.

3. Admissible associative models. Admissible associative models represent the level of sense and the above truth conditions are intended to capture the notion of truth with respect to that level. Associative models are admissible just in case they satisfy certain constraints imposed upon the valuation function v. Intuitively, these constraints are governed by nominal definitions for the constants in the substitution class $\mathcal{C}$ and by meaning postulates for the predicates in $\mathcal{P}$.

To explain the notion of an admissible model we introduce a couple of auxiliary notions by making the following stipulations:

> (A1) Every nominal constant $\alpha \in \mathcal{C}$ has a (possibly empty) *defining associate*, $v_{def}(\alpha)$, associated with it. This is the set of all sentences from $Atm(\alpha)$ which we call *defining of* α. (This associate captures the nominal definition of the name symbolized by α.)
>
> (A2) For every nominal constant $\alpha \in \mathcal{C}$ which has a defining associate, $Def(\alpha) \subseteq \mathcal{P}$ is the set of all the pure predicates occurring in the sentences in $v_{def}(\alpha)$. If $Def(\alpha) = \{\varphi, \chi, \psi, ...\}$, we say that $\varphi, \chi, \psi, ...$ are the *defining predicates* of α.
>
> (A3) Every pure predicate φ in $\mathcal{P}$ has a *meaning postulate*, $Mp(\varphi)$, associated with it. $Mp(\varphi)$ is a sentence of $\mathcal{L}$. We put $Mp =_{def} \{Mp(\varphi) : \varphi \in \mathcal{P}\}$. A meaning postulate for φ determines which predicates are consequential upon φ with respect to some nominal constant, and which predicates conform to φ with respect to it: (a) A predicate $\chi \in \mathcal{P}$ is *(definitionally) consequential upon* φ with respect to a constant $\alpha \in \mathcal{C}$ just in case, if $\varphi...\alpha... \in v_{def}(\alpha)$, then $\chi...\alpha...$ is Mp-derivable from $v_{def}(\alpha)$; (b) A predicate $\chi \in \mathcal{P}$ *(definitionally) conforms to* φ with respect to α just in case, if $\varphi...\alpha... \in v_{def}(\alpha)$, then $\chi...\alpha...$ is Mp-consistent with $v_{def}(\alpha)$.

(We assume that the relavant notion of derivability and consistency is that of a purely syntactical meaning calculus (of meaning rules) which captures the postulates in Mp. The sole syntactical function of that calculus is to fill the associates with pure atomic sentences in a systematic way.)

Let $\mathcal{I} = \langle \mathcal{C}, \mathcal{P}, v \rangle$ be a model and α any nominal constant in $\mathcal{C}$. An *assignment to a constant* α is said to be *admissible* just in case it satisfies the following conditions:

> (B1) The resulting associate $v(\alpha)$ for α contains the *defining associate* of α, i.e., $v_{def}(\alpha)$.
>
> (B2) The resulting associate $v(\alpha)$ for α contains the *consequential associate* of α, $v_{cns}(\alpha)$, i.e., the set of all sentences from $Atm(\alpha)$ which are derivable from $v_{def}(\alpha)$ in view of Mp.

We call the set $v_{chr}(\alpha) = v_{def}(\alpha) \cup v_{cns}(\alpha)$ the *characteristic associate* of α. (So an assignment to a constant α is admissible just in case $v_{chr}(\alpha) \subseteq v(\alpha)$.) We call the set $Chr(\alpha)$ of predicates which occur in the elements of $v_{chr}(\alpha)$ the set of *characteristic predicates* for α.

When $\mathcal{I} = \langle \mathcal{C}, \mathcal{P}, v \rangle$ is a model and α any constant in $\mathcal{C}$, we say that an *assignment to a predicate* $\varphi \in \mathcal{P}$ is *admissible* just in case the resulting predicate associate $v(\varphi)$ for φ contains all the sentences from $Atm(\varphi)$ that are contained in the characteristic associates $v_{chr}(\alpha)$ for any $\alpha \in \mathcal{C}$.

We call a model $\mathcal{I} = \langle \mathcal{C}, \mathcal{P}, v \rangle$ *admissible* just in case it satisfies the conditions on admissible assignments to nominal constants and pure predicates just stated.

When $\mathcal{I} = \langle \mathcal{C}, \mathcal{P}, v \rangle$ is a model and α a constant in $\mathcal{C}$, we define the *(characteristically) conforming associate* of α, $v_{cnf_{chr}}(\alpha) \subseteq v(\alpha)$ to be the set of sentences form $Atm(\alpha)$ which are derivationally consistent with $v_{chr}(\alpha)$ given Mp. We call the predicates which occur in the sentences from $v_{cnf_{chr}}(\alpha)$ and which are not already contained in $Chr(\alpha)$ the *(characteristically) conforming predicates* of α. These predicates form the set $Cnf_{chr}(\alpha)$.

There is a subclass of admissible models in which atomic formulae can come out true which are composed from conforming predicates. These models will be the intended ones considered in later sections. We call an admissible model $\mathcal{I} = \langle \mathcal{C}, \mathcal{P}, v \rangle$ *diversifying* when and only when it satisfies the conditions that for some $\alpha \in \mathcal{C}$ and some $\varphi^n \in \mathcal{P}$: (a) $v_{chr}(\alpha)$ is such that $v_{chr}(\alpha) \subseteq v(\alpha)$ and $v_{chr}(\alpha) \neq v(\alpha)$; and (b) $v(\varphi^n)$ is such that $\varphi^n \notin Chr(\alpha)$.

For any admissible model $\mathcal{I} = \langle \mathcal{C}, \mathcal{P}, v \rangle$, we require that the set of the associates for all the pure predicates in $\mathcal{P}$, which we call the *spectrum of predicates* of the model, be Mp-consistent. In contrast, the associate of a nominal constant α in an admissible model, $v(\alpha) = v_{chr}(\alpha) \cup v_{cnf_{chr}}(\alpha)$, will not be a consistent set in view of Mp. The reason is that the members of $v_{cnf_{chr}}(\alpha)$ are only required to be consistent with the characteristic associate and need not be consistent with each other. We call the union of the associates for all nominal constants of an admissible model the *spectrum of constants*.

An example: {Hesperus is a planet} is a subset of the characteristic portion of the actual sense-extension of 'Hesperus', and {Hammurabi likes Hesperus, Hesperus is smaller than Jupiter, Jupiter is smaller than Hesperus} is a subset of the characteristically conforming sense-extension of this name. The actual sense-extension of the predicate '... is smaller than ...' contains {Hesperus is smaller than Jupiter} as a subset but not {Jupiter is smaller than Hesperus}.

4 Addressing the puzzles at the level of sense

We say that an agent is *enlightened* with respect to the synonymy of a pair of names, when she is in a position to assent to a statement of substitutional identity which is composed from these names, and that she is *ignorant* otherwise. In what follows we shall first consider puzzle cases, in which we are ignorant about the synonymy of a pair of names, as they seem to be less problematic than those in which we are enlightened.

4.1 Ignorant anti-substitution intuitions

With the exception of (1) and (7) every example we shall consider in what follows is taken from Saul (2007).

1. Phone booth. Consider the following triad:

> (2) *C*lark Kent went *i*nto the phone booth and *S*uperman came *o*ut. (In symbols: $Ic \wedge Os$.)
>
> (2*) *S*uperman went *i*nto the phone booth and *C*lark Kent came *o*ut. ($Is \wedge Oc$.)
>
> (2†) *S*uperman $\doteqdot$ *C*lark Kent. ($s \doteqdot c$.)

In the ignorant case we take (2) to be true and (2*) to be false, because we do not know that (2†) is true.

We can model such cases in terms of *restricted associative models*. These models represent the portion of the level of sense which is accessible to an ignorant agent a. To obtain a restricted model, $\mathcal{I}_{\mathsf{a}} = \langle \mathcal{C}, \mathcal{P}, v_{\mathsf{a}} \rangle$, from an unrestricted one, $\mathcal{I} = \langle \mathcal{C}, \mathcal{P}, v \rangle$, we relativize the valuation v to a. (Like remarks apply to the portion $v_{\mathsf{a}}(\varphi^n)$ of the sense-extension of a predicate φ^n that is accessible to the agent.)

A simple (fragment of a) restricted model in which the sentences from the phone booth triad receive their correct ignorant truth values is the following one:

> $\mathcal{I}_{\mathsf{a}} = \langle \mathcal{C}, \mathcal{P}, v_{\mathsf{a}} \rangle$ where: $\mathcal{C} = \{c, s\}$; $\mathcal{P} = \{I, O\}$; $v_{\mathsf{a}}(c) = \{Ic\}$; $v_{\mathsf{a}}(s) = \{Os\}$; $v_{\mathsf{a}}(I) = \{Ic\}$; $v_{\mathsf{a}}(O) = \{Os\}$.

In this model we have: $\mathcal{I}_{\mathsf{a}} \models Ic \wedge Os$; $\mathcal{I}_{\mathsf{a}} \not\models Is \wedge Oc$; $\mathcal{I}_{\mathsf{a}} \not\models Ic \wedge Oc$; $\mathcal{I}_{\mathsf{a}} \not\models Is \wedge Os$. We also have: $\mathcal{I}_{\mathsf{a}} \not\models s \doteqdot c$. In the unrestricted model $\mathcal{I} = \langle \mathcal{C}, \mathcal{P}, v \rangle$ which represents the objective level of sense and in the enlightened models all these sentences will come out true. Here, the (relevant fragment of the) enlightened model will be:

> $\mathcal{I}_{\mathsf{a}} = \langle \mathcal{C}, \mathcal{P}, v_{\mathsf{a}} \rangle$ where: $\mathcal{C} = \{c, s\}$; $\mathcal{P} = \{I, O\}$; $v_{\mathsf{a}}(c) = \{Ic, Oc\}$; $v_{\mathsf{a}}(s) = \{Os, Is\}$; $v_{\mathsf{a}}(I) = \{Ic, Oc\}$; $v_{\mathsf{a}}(O) = \{Os, Is\}$.

2. Building leaping. As we do not know that ($2^\dagger$) is true, we take (3) to be true and (3*) to be false.

(3) *S*uperman *l*eaps more tall buildings than *C*lark Kent. (*Lsc*.)

(3*) *C*lark Kent *l*eaps more tall buildings than *S*uperman. (*Lcs*.)

This situation can be modeled as follows:

$\mathcal{I}_{\mathsf{a}} = \langle \mathcal{C}, \mathcal{P}, v_{\mathsf{a}} \rangle$ where: $\mathcal{C} = \{s, c\}$; $\mathcal{P} = \{L\}$; $v_{\mathsf{a}}(s) = \{Lsc\}$; $v_{\mathsf{a}}(c) = \{Lsc\}$; $v_{\mathsf{a}}(L) = \{Lsc\}$.

We have: $\mathcal{I}_{\mathsf{a}} \models Lsc$; $\mathcal{I}_{\mathsf{a}} \not\models Lcs$; $\mathcal{I}_{\mathsf{a}} \not\models Lss$; $\mathcal{I}_{\mathsf{a}} \not\models Lcc$. We also obtain: $\mathcal{I}_{\mathsf{a}} \not\models s \ddot{=} c$. In the unrestricted model and in the enlightened models the $\ddot{=}$-sentence will be true, but the above leaping sentences will be false. In this case, a simple fragment of the enlightened model will be:

$\mathcal{I}_{\mathsf{a}} = \langle \mathcal{C}, \mathcal{P}, v_{\mathsf{a}} \rangle$ where: $\mathcal{C} = \{s, c\}$; $\mathcal{P} = \{L, K, T\}$; $v_{\mathsf{a}}(s) = \{Lsc, Lcs, Ks, Tsc, Tcs, Tss\}$; $v_{\mathsf{a}}(c) = \{Lsc, Lcs, Kc, Tsc, Tcs, Tcc\}$; $v_{\mathsf{a}}(L) = \{\ \}$; $v_{\mathsf{a}}(K) = \{Ks, Kc\}$; $v_{\mathsf{a}}(T) = \{Tsc, Tcs, Tss, Tcc\}$.

Here we may take K to symbolize '... is a Kryptonian' and T to symbolize '... is as tall as ...'.

3. Love life. As a last example of a simple sentence with names we shall consider the case in which we do not know that ($2^\dagger$) is true, and assume (4) to be true and (4*) to be false.

(4) *L*ois *s*lept with *S*uperman but she didn't *s*leep with *C*lark Kent. ($Sls \wedge \neg Slc$.)

(4*) *L*ois *s*lept with *C*lark Kent but she didn't *s*leep with *S*uperman. ($Slc \wedge \neg Sls$.)

A restricted model of this situation will be:

$\mathcal{I}_{\mathsf{a}} = \langle \mathcal{C}, \mathcal{P}, v_{\mathsf{a}} \rangle$ where: $\mathcal{C} = \{l, s, c\}$; $\mathcal{P} = \{S\}$; $v_{\mathsf{a}}(l) = \{Sls\}$; $v_{\mathsf{a}}(s) = \{Sls\}$; $v_{\mathsf{a}}(c) = \{\ \}$; $v_{\mathsf{a}}(S) = \{Sls\}$.

We have: $\mathcal{I}_{\mathsf{a}} \models Sls \wedge \neg Slc$; $\mathcal{I}_{\mathsf{a}} \not\models Slc \wedge \neg Sls$; $\mathcal{I}_{\mathsf{a}} \not\models Sls \wedge \neg Sls$; $\mathcal{I}_{\mathsf{a}} \not\models Slc \wedge \neg Slc$. And we have: $\mathcal{I}_{\mathsf{a}} \not\models s \ddot{=} c$. In unrestricted model and in the enlightened ones the $\ddot{=}$-sentence will be true, but the rest of the sentences considered will come out false. We shall now consider a couple of non-name cases discussed in Saul (2007).

4. Definite descriptions. We take names and definite descriptions to be singular terms rather than devices of quantification and formalize them, accordingly, as nominal constants rather than, like on the familiar Russellian analysis, as complex existential quantifications. (There is no metaphysical motivation for such an analysis, when we look to the level of sense.) We treat definite descriptions semantically in the same way like names. Thus the nominal constants which symbolize them, can be contained in substitution classes, can be assigned associates and so on.

The sense-extensions of definite descriptions (and, accordingly, the associates of the nominal constants which represent them) differ from names in but one respect: The defining portion of the sense-extension of a definite description contains a *single* pure atomic sentence. Consider, e.g., the description 'the worst-dressed superhero' (in symbols: h). The defining portion of the sense-extension of this description is {The worst-dressed superhero is the worst-dressed superhero}. As sense-extensions do only contain pure atomic sentences, the 'is' in the sentence contained in it is a copula rather than an identity predicate. Accordingly, we take this sentence to be composed from the singular term 'the worst-dressed superhero' and the 1-place predicate '... is the worst-dressed superhero' and take its formalization to receive the logical form of a pure atomic sentence of $\mathcal{L}$, i.e., $\varphi\alpha$.[12] The meaning postulate for '... is the worst-dressed superhero' will be to the effect that it implies, e.g., the predicates '... is worst-dressed' and '... is a superhero'.

Let us assume that according to the fiction, Superman is the worst-dressed superhero and Clark Kent is the shyest reporter. Our ignorance about the truth of ($2^{\dagger}$) allows us to assume that (5) is true and (5*) false.

> (5) The shyest *r*eporter went *i*nto the phone booth and the worst-dressed super*h*ero came *o*ut. ($Ir \wedge Oh$.)
>
> (5*) The worst-dressed super*h*ero went *i*nto the phone booth and the shyest *r*eporter came *o*ut. ($Ih \wedge Or$.)

An ignorant model which captures this situation is the following one:

> $\mathcal{I}_{\mathsf{a}} = \langle \mathcal{C}, \mathcal{P}, v_{\mathsf{a}} \rangle$ where: $\mathcal{C} = \{r, h\}$; $\mathcal{P} = \{R, H, I, O\}$; $v_{\mathsf{a}}(r) = \{Rr, Ir\}$; $v_{\mathsf{a}}(h) = \{Hh, Oh\}$; $v_{\mathsf{a}}(R) = \{Rr\}$; $v_{\mathsf{a}}(H) = \{Hh\}$; $v_{\mathsf{a}}(I) = \{Ir\}$; $v_{\mathsf{a}}(O) = \{Oh\}$.

We have: $\mathcal{I}_{\mathsf{a}} \models Ir \wedge Oh$; $\mathcal{I}_{\mathsf{a}} \not\models Ih \wedge Or$; $\mathcal{I}_{\mathsf{a}} \not\models Ir \wedge Or$; $\mathcal{I}_{\mathsf{a}} \not\models Ih \wedge Oh$. And we also obtain: $\mathcal{I}_{\mathsf{a}} \models Rr$; $\mathcal{I}_{\mathsf{a}} \models Hh$; $\mathcal{I}_{\mathsf{a}} \models r \ddot{=} r$; $\mathcal{I}_{\mathsf{a}} \models h \ddot{=} h$; $\mathcal{I}_{\mathsf{a}} \not\models h \ddot{=} r$. In the unrestricted model and in the enlightened models all sentences considered will be true.[13]

5. Indexicals. We assume (oversimplifying greatly) that, at the level of sense, indexicals recieve an anaphoric treatment. The idea is, roughly, to substitute the indexical expressions in the original sentence by the singular terms to which they are anaphorically

[12] There is room for such a construal of 'The worst-dressed superhero is the worst-dressed superhero', as there is a perfectly good sense in which Lois Lane might suffer from the fact that the worst-dressed superhero is the worst-dressed superhero, without suffering from the fact that he is self-identical.

[13] A terminological remark: It seems to me to be philosophically misleading to call non-referring expressions like 'the worst-dressed superhero' "definite descriptions" as this suggests that there is a bearer this expression describes as possessing such-and-such identifying traits. Of course, this terminology is inappropriate for the level of sense in general, since we do not describe anything when we reflect the sense-extensions of singular terms and predicates. Nevertheless, we shall stick to the usual terminological convention. Note in this connection that it would be odd to assume that a non-referring definite description like 'the worst-dressed superhero' describes something after all, e.g., an intentional object. For, arguably, abstract objects (like these) do not wear cloth and so on. (This does also apply when non-referring names or referring names which occur in intensional contexts are taken to refer to abstract objects.)

linked and then to evaluate the resulting substitution instance instead of the original indexical sentence. Let us assume, for example, that according to the fiction, Lois Lane sees Superman flying through the sky and utters (7).

(7) I adore him.

In this case 'I' will be linked to 'Lois Lane' and 'him' to 'Superman', and (7) will be evaluated in terms of (7′).

(7′) Lois Lane adores Superman.

Next consider sentence (8).

(8) He hit Lex Luther more times than he did.

We assume that the fiction links the first occurrence of 'he' to 'the be-caped hero in the Superhero Book of Mug Shots', and the second to 'the shy reporter lurking in the corner of the room'. We then evaluate (8) as (8′).

(8′) The be-caped hero in the Superhero Book of Mug Shots hit Lex Luther more times than the shy reporter lurking in the corner of the room did.

In the appropriate restricted model (8) will be true, but it will be false in the unrestricted model and in the enlightened ones. (The situation here is similar to that of (3).)

6. Quantification. Let us consider a further example presented in Saul (2007). According to the fiction, Clark Kent's shy colleagues *A*rt and *B*art are sitting together with *C*lark in the conference room. All of them bemoan their lack of dates for an upcoming ball. Sentence (9), thus, seems to be true of this scenario.

(9) Nobody *i*n the conference roo*m* is *s*uccessful with woman.

In the appropriate ignorant model $\neg(\Sigma x)(Ixm \wedge Sx)$ will be true. But it will be false in the unrestricted model and the enlightened ones, as in view of the relevant piece of fiction they will ensure the truth of something like '*S*uperman is a *h*eart-throb' and of 'Superman $\doteqdot$ Clark Kent'. Here the meaning postulate for '... is a *h*eart-throb' will ensure that '... is *s*uccessful with woman' is implied.

An ignorant model which captures this situation is the following one:

$\mathcal{I}_{\mathsf{a}} = \langle \mathcal{C}, \mathcal{P}, v_{\mathsf{a}} \rangle$ where: $\mathcal{C} = \{a, b, c, s, m\}$; $\mathcal{P} = \{I, S, H\}$; $v_{\mathsf{a}}(a) = \{Iam, Sa\}$; $v_{\mathsf{a}}(b) = \{Ibm, Sb\}$; $v_{\mathsf{a}}(c) = \{Icm, Sc\}$; $v_{\mathsf{a}}(s) = \{Hs, Ss\}$; $v_{\mathsf{a}}(m) = \{Iam, Ibm, Icm\}$; $v_{\mathsf{a}}(I) = \{Iam, Ibm, Icm\}$; $v_{\mathsf{a}}(S) = \{Ss\}$; $v_{\mathsf{a}}(H) = \{Hs\}$.

An enlightened model will be as follows:

$\mathcal{I}_{\mathsf{a}} = \langle \mathcal{C}, \mathcal{P}, v_{\mathsf{a}} \rangle$ where: $\mathcal{C} = \{a, b, c, s, m\}$; $\mathcal{P} = \{I, S, H\}$; $v_{\mathsf{a}}(a) = \{Iam, Sa\}$; $v_{\mathsf{a}}(b) = \{Ibm, Sb\}$; $v_{\mathsf{a}}(c) = \{Icm, Sc, Hc\}$; $v_{\mathsf{a}}(s) = \{Ism, Ss, Hs\}$; $v_{\mathsf{a}}(m) = \{Iam, Ibm, Icm, Ism\}$; $v_{\mathsf{a}}(I) = \{Iam, Ibm, Icm, Ism\}$; $v_{\mathsf{a}}(S) = \{Sc, Ss\}$; $v_{\mathsf{a}}(H) = \{Hc, Hs\}$.

7. Attitude Reports. Let us now consider a case in which our reluctance to substitute is triggered by the opacity inducing attitude verb 'believes'.

(1) Ann believes that *M*ark Twain is a *w*riter. ($\mathrm{B}_a(Wm)$.)

(1*) Ann believes that Sam*u*el Clemens is a *w*riter. ($\mathrm{B}_a(Wu)$.)

To explain this case we extend our substitutional language $\mathcal{L}$ with an operator for belief, B_a (informally: 'a believes that'), and employ doxastic associative models. Doxastic models represent the portion of the level of sense which is doxastically accessible to a subject a.

A doxastic model is a 6-tuple $\mathcal{D} = \langle S, R, \mathcal{C}, c, \mathcal{P}, v\rangle$ where: S is a non-empty set of indices $s, t, \ldots$ (intuitively, doxastic states); $R \subseteq S \times S$; $\mathcal{C}$ and $\mathcal{P}$ are as before; $c : S \to \wp(\mathcal{C})$ with $c(s)$ being the substitution class for some $s \in S$ and $\mathcal{C} = \bigcup_{s\in S} c(s)$; finally, v is defined as follows: $v : \mathcal{C} \times S \to \wp(Atm)$ such that $v(\alpha, s) \subseteq Atm(\alpha)$ and $v : \mathcal{P} \times S \to \wp(Atm)$ such that $v(\varphi^n, s) \subseteq Atm(\varphi^n)$. Assignments σ to the nominal variables are relativized to the elements of S in the natural way. The clause for the belief operator is: $\mathcal{D}_\sigma \models_s \mathrm{B}_a(A)$ iff for all $t \in S$, if sRt then $\mathcal{D}_\sigma \models_t A$. The other clauses remain, *mutatis mutandis*, as before.

To model the doxastic state of an ignorant doxastic subject we put $a = \mathsf{a}$ and relativize the valuation to the ignorant agent a letting v_a be defined as follows: $v_\mathsf{a} : \mathcal{C} \times S \to \wp(Atm)$ such that $v_\mathsf{a}(\alpha, s) \subseteq Atm(\alpha)$ and $v_\mathsf{a} : \mathcal{P} \times S \to \wp(Atm)$ such that $v_\mathsf{a}(\varphi^n, s) \subseteq Atm(\varphi^n)$. Restricted doxastic models represent the portion of the level of sense which is doxastically accessible to an ignorant doxastic subject a. The following restricted model is one of those which capture the situation a faces with respect to the pair (1) and (1*):

$\mathcal{D}_\mathsf{a} = \langle S, R, \mathcal{C}, c, \mathcal{P}, v_\mathsf{a}\rangle$, where: $S = \{s, t\}$; $R = \{\langle s, t\rangle, \langle s, s\rangle, \langle t, t\rangle\}$; $\mathcal{C} = c(s) = c(t) = \{m, u\}$; and $\mathcal{P} = \{W\}$. The assignments to the nominal constants and predicate letters is as follows.

For 'Mark' (m): $v_\mathsf{a}(m, s) = v_\mathsf{a}(m, t) = \{Wm\}$.

For 'Samuel' (u): $v_\mathsf{a}(u, s) = \{\ \}$; $v_\mathsf{a}(u, t) = \{Wu\}$.

For 'writer' (W): $v_\mathsf{a}(W, s) = \{Wm\}$; $v_\mathsf{a}(W, t) = \{Wm, Wu\}$.

We obtain: $\mathcal{D}_\mathsf{a} \models_s \mathrm{B}_\mathsf{a}(Wm)$; $\mathcal{D}_\mathsf{a} \not\models_s \mathrm{B}_\mathsf{a}(Wu)$; and $\mathcal{D}_\mathsf{a} \not\models_s m \ddot{=} u$. (In the above model a associates Wu with W but does not associate it with u.) In enlightened doxastic models $\mathrm{B}_a(Wm)$, $\mathrm{B}_a(Wu)$, and $m \ddot{=} u$ will be true at s.

8. Name Change. As a last example we shall consider the case of a simple sentence, in which we use synonymous names in a conventionally restricted way rather than being ignorant about their synonymy. Consider the following situation. There was a phase of time a certain city was called 'Leningrad' and a time before and after that phase it was called 'St Petersburg'. On the assumption that Albert visited this city in the strech of time it was called 'St Petersburg', it seems that (9) is true, whereas (9*) seems to be false.[14]

(9) *A*lbert never *v*isited *L*eningrad, but he *v*isited St *P*etersburg.
(In symbols: $\neg \mathrm{P}(Val) \wedge \mathrm{P}(Vap)$.)

(9*) *A*lbert never *v*isited St *P*etersburg, but he *v*isited *L*eningrad.
(In symbols: $\neg \mathrm{P}(Vap) \wedge \mathrm{P}(Val)$.)

To explain anti-substitution intuitions in this case, we extend our substitutional language with an operator P for 'It was the case that' and consider conventionally restricted temporal models. Such a model is a 7-tuple $\mathcal{T}_{\mathsf{C}} = \langle T, <, t_0, \mathcal{C}, c, \mathcal{P}, v_{\mathsf{C}} \rangle$ restricted by some convention c where: T is a non-empty set of indices $t, t', t'', ...$ (intuitively, instants of time); $<$ is a binary relation on T; t_0 is "now"; $\mathcal{C}$ and $\mathcal{P}$ are as usual; $c : T \rightarrow \wp(\mathcal{C})$ with $c(t)$ being the substitution class for some $t \in T$ and $\mathcal{C} = \bigcup_{t \in T} c(t)$; and $v_{\mathsf{C}} : \mathcal{C} \times T \rightarrow \wp(Atm)$ such that $v_{\mathsf{C}}(\alpha, t) \subseteq Atm(\alpha)$ and $v_{\mathsf{C}} : \mathcal{P} \times T \rightarrow \wp(Atm)$ such that $v_{\mathsf{C}}(\varphi^n, t) \subseteq Atm(\varphi^n)$. For simplicity we let the substitution classes be constant across instants of time (i.e., for any $t \in T$, $c(t) = \mathcal{C}$). The clause for the P-operator takes the familiar shape: $\mathcal{T}_{\mathsf{C}\sigma} \models_{t_0} \mathrm{P}(A)$ iff for some $t \in T$, $t < t_0$ and $\mathcal{T}_{\mathsf{C}\sigma} \models_t A$. The clauses for negation and conjunction are also as usual.

A conventionally restricted model which captures anti-substitution intuitions concerning the pair (9) and (9*) is the following one:

$\mathcal{T}_{\mathsf{C}} = \langle T, <, t_0, \mathcal{C}, c, \mathcal{P}, v_{\mathsf{C}} \rangle$, where: $T = \{t_1, t_2, t_3, t_4, t_5, t_6, t_7\}$; $t_0 = t_7$; $<$ (intuitively, the earlier-later relation) is a linear order on T (i.e., $<$ is irreflexive, transitive, and weakly connected; we shall assume that the structure of time is isomorphic to real numbers); $\mathcal{C} = \{a, p, l\}$, where $c(t) = \mathcal{C}$ for all $t \in T$; and $\mathcal{P} = \{V\}$.

The convention c for the restriction of the valuation of 'St Petersburg' (symbol: p), 'Leningrad' (symbol: l), and 'visits' (symbol: V) is as follows:

1. We let $\{t_1, t_2, t_5, t_6, t_7\}$ be the phase of time in which Leningrad/St Petersburg sentences are to be evaluated in terms of atomic sentences which contain the name 'St Petersburg' rather than 'Leningrad' (= *p-phase*).

2. We let $\{t_3, t_4\}$ be the phase of time in which Leningrad/ St Petersburg sentences are to be evaluated in terms of atomic sentences which contain the name 'Leningrad' rather than 'St Petersburg' (= *l-phase*).

In accordance with convention c we shall consider restricted valuations which satisfy the following conditions:

[14]This is a variant of an example originally proposed in Saul (1997). For an elaborate reference guided account of anti-substitution intuitions in the case of name change (which proceeds in terms of a Montagovian framework) see Zimmermann (2005).

(a) The conventionally restricted associates of the nominal constant a which represent the sense-extensions of 'Albert' contain in their conforming portions Vap at $t_{1,2,5,6,7} \in T$ (i.e., in the p-phase) but Val at $t_{3,4} \in T$ (i.e., in the l-phase).

(b) The conventionally restricted associates of the predicate letter V which represent the sense-extensions of the predicate 'visits' are arranged in such a way that (i) no pure atomic sentence which is composed from V and l is in the sense-extension of V in the p-phase, and (ii) no pure atomic sentence which is composed from V and p is in the sense-extension of V in the l-phase.

Moreover, the actual sense-extension of V will not contain Val in the l-phase but it will contain Vap in the p-phase. The following valuation satisfies convention c:

For 'Albert' (a):
$v_{\mathsf{C}}(a, t_{1,2,5,6,7}) = \{Vap\}$; $v_{\mathsf{C}}(a, t_{3,4}) = \{Val\}$.

For 'St Petersburg' (p):
$v_{\mathsf{C}}(p, t_{1,2,5,6}) = \{Vap\}$; $v_{\mathsf{C}}(p, t_{3,4}) = v_{\mathsf{C}}(p, t_7) = \{\ \}$.

For 'Leningrad' (l):
$v_{\mathsf{C}}(l, t_{1,2,5,6}) = v_{\mathsf{C}}(l, t_7) = \{\ \}$; $v_{\mathsf{C}}(l, t_{3,4}) = \{Val\}$.

For 'visits' (V):
$v_{\mathsf{C}}(V, t_{1,2,5,6}) = \{Vap\}$; $v_{\mathsf{C}}(V, t_{3,4}) = v_{\mathsf{C}}(V, t_7) = \{\ \}$.

We have: $\mathcal{T}_{\mathsf{C}} \models_{t_{1,2,5,6}} Vap$; $\mathcal{T}_{\mathsf{C}} \not\models_{t_{3,4,7}} Vap$; $\mathcal{T}_{\mathsf{C}} \not\models_{t_{1,2,3,4,5,6,7}} Val$; $\mathcal{T}_{\mathsf{C}} \models_{t_7} \neg\mathrm{P}(Val) \wedge \mathrm{P}(Vap)$; $\mathcal{T}_{\mathsf{C}} \not\models_{t_7} \neg\mathrm{P}(Vap) \wedge \mathrm{P}(Val)$; and $\mathcal{T}_{\mathsf{C}} \not\models_{t_{1,2,3,4,5,6,7}} p \ddot{=} l$.

In the unrestricted model $\mathcal{T} = \langle T, <, t_0, \mathcal{C}, c, \mathcal{P}, v\rangle$ convention c does not hold. The (conforming portions of the) associates of a, p, and l will contain $\{Vap, Val\}$ as a subset. And the actual associates of V will contain both Vap and Val. In the unrestricted model both $\neg\mathrm{P}(Val) \wedge \mathrm{P}(Vap)$ and $\neg\mathrm{P}(Vap) \wedge \mathrm{P}(Val)$ will be false, whereas $p \ddot{=} l$ will be true. Thus in conventionally restricted models the names 'St Petersburg' and 'Leningrad' are not synonymous with respect to the level of sense, whereas in the unrestricted model one cannot discern phases.

4.2 Enlightened anti-substitution intuitions

Many reference guided theorists have enlightened anti-substitution intuitions and think that the original sentences (♯) may seem to be true to us and the sentences (♯*) which are obtained from them by a substitution of the synonymous singular terms may seem to be false, despite the fact that we also know that these terms are synonymous (in virtue of being co-referential).

We shall reflect upon enlightened anti-substitution intuitions from our non-referential perspective. So the intuition is that, for instance, (2) may seem true to us and (2*) false, even though we know that ($2^\dagger$) is true.

(2) *C*lark Kent went *i*nto the phone booth and *S*uperman came *o*ut. ($Ic \wedge Os$.)

(2*) *S*uperman went *i*nto the phone booth and *C*lark Kent came *o*ut. ($Is \wedge Oc$.)

($2^\dagger$) *S*uperman $\doteqdot$ *C*lark Kent. ($s \doteqdot c$.)

Now, what can be said about this from the present perspective?[15] Perhaps, what might be relevant in this case is the difference of ($2^\dagger$) and ($2^\ddagger$) with respect to cognitive significance.

($2^\ddagger$) *S*uperman $\doteqdot$ *S*uperman. ($s \doteqdot s$.)

($2^\dagger$) seems to be informative, whereas ($2^\ddagger$) seems to express a triviality. In view of the fact that on the present account the senses—more exactly, the nominal definitions—of 'Superman' and 'Clark Kent' are exactly the same, we cannot explain the difference of ($2^\dagger$) and ($2^\ddagger$) in cognitive significance in terms the difference of the senses of these names. (We have illustrated this in Section 2 in terms of 'Hesperus' and 'Phosphorus'.) Looking to the level of sense, though, we can explain this difference in terms of sense-extensions. ($2^\dagger$) and ($2^\ddagger$) differ in *cognitive significance*, because 'Superman' and 'Clark Kent' have distinct sense-extensions. We say that the sense-extensions of two terms are identical just in case they contain exactly the same elements; and that they are distinct otherwise. (What matters here, of course, are unrestricted sense-extensions.)

To see that 'Superman' and 'Clark Kent' have distinct sense-extensions, it suffices to realize that the portions of these sense-extensions which contain elements which are composed from monadic pure predicates can never be the same. The sense-extension 'Superman', i.e. {..., Superman is gentle, ...}, will be distinct from that of 'Clark Kent', i.e. {..., Clark Kent is gentle, ...}.[16]

[15] I must admit that I am not sure whether I have enlightened anti-substitution intuitions.

[16] The presence of atomic sentences which are composed from monadic pure predicates in the sense-extension of a singular term is vital for the present account of cognitive significance. To realize this consider the highly artificial case of two constants a and b both having $\{Gab\}$ as their unconstrained common associate. According to our account, then, the sentences $a \doteqdot b$ and $a \doteqdot a$ will not differ in cognitive significance. This case does not only show a limitation of the theory. It also points out that the identity of the associates of a pair of nominal constants does not reduce to their typographical identity. With regard to the limitation, though, we might find consolation in the fact that this example is indeed highly artificial, as it is unlikely that these associates will represent the level of sense of an intuitively acceptable language. For instance, let a symbolize the numeral '2', let b symbolize the numeral '1', and let G symbolize the predicate $>$. On the assumption that the sense-extension of these numerals, i.e. $\{2 > 1\}$, matches with the sense-extension of $>$ with respect to the sentence '$2 > 1$', we won't be able—as the nominal definitions for the numerals and the meaning postulates for $>$ dictate—to claim also the

The intuitions of the enlightened which say that (2) is true and (2*) is false even though they know the truth of (2†) might be, I suggest, due to two central factors: (i) the fact that 'Superman' and 'Clark Kent' reflect distinct sense-extensions; and (ii) the fact that different portions of these sense-extensions are relevant for the truth of each of the conjuncts in (2) and (2*). This kind of explanation is also available in the other puzzle cases discussed above.

References

Fine, Kit (2002) *The Limits of Abstraction*, Oxford: Clarendon Press.

Fodor, Jerry A. (1975) *The Language of Thought*, Cambridge (Mass.): Harvard University Press.

Kripke, Saul (1976) "Is there a problem about substitutional quantification?", in G. Evans and J. McDowell (eds.) *Truth and Meaning: Essays in Semantics*, Oxford: Oxford University Press, 325–419.

Leblanc, Hugues (1976). *Truth-Value Semantics*, Amsterdam: North-Holland.

Peacocke, Christopher (1999) *Being Known*, Oxford: Clarendon Press.

Saul, Jennifer (1997) "Substitution and simple sentences", *Analysis* **57**, 102–108.

Saul, Jennifer (2007) *Simple Sentences, Substitution, and Intuitions*, Oxford: Oxford University Press.

Tarski, Alfred (1983) "The establishment of scientific semantics", in A. Tarski *Logic, Semantics, Metamathematics*, translated by J. H. Woodger, second edition by J. Corcoran, Indianapolis (Ind.): Hackett, 401–408. (English translation of the German translation of Tarski, Alfred (1936) "O ugruntowaniu naukowej semantyki", *Przeglad Filozoficzny* **39**, 50–57.)

Wieckowski, Bartosz (2008) "Predication in fiction", forthcoming in M. Peliš (ed.) *The Logica Yearbook 2007*, Prague: Filosofia.

Wieckowski, Bartosz (submitted ms.) "Associative substitutional semantics and quantified modal logic".

Zimmermann, Thomas Ede (2005) "What's in two names?", *Journal of Semantics* **22**, 53–96.

truth of, e.g., '2 is a number' or '1 has a size' with respect to the level of sense. An intuitively appealing language, though, should allow monadic predication.