Are Intensions Necessary?
*Sense as the Construction of Reference**

Almerindo E. Ojeda
University of California at Davis
aeojeda@ucdavis.edu

§1. Introduction

More than half a century ago, Carnap (1947) defined the *extension* of an expression in a particular state of affairs as its reference in such a state, while construing the *intension* of an expression as a function that assigned, to each state of affairs, the extension of that expression in that state. He then considered taking the meaning of an expression to be its intension. The view that the meaning of an expression is its intension became the leading idea of *possible-worlds semantics*, an interpretive framework which soon became the foundation of the classic interpretation of modal logic (Kripke, 1959, 1963a, 1963b). Moreover, as shown in the large body of work emanating from the pieces collected in Montague (1974), the view that meaning is intension has shed light on many critical issues for the semantics of natural languages, and has become a cornerstone of that enterprise. Chief among these issues are the variable informativeness of identity statements, the failure of substitution in opaque contexts, the compositional interpretation of modal verbs and adverbs, the non-trivial nature of counterfactuals, and the nonsynonymy of vacuous predicates.

Yet, as Carnap himself realized, the identification of meaning and intension runs into serious difficulties when it attempts to interpret expressions that have the same intension yet differ in meaning. Consider for example the two terms in (1), the two predicates in (2), and the two sentences in (3).

(1)   a. *two plus three*
        b. *six minus one*

(2)   a. *equilateral triangle*

---

    b. *equiangular triangle*

(3)   a. *The angular sum of any triangle is 180°.*
       b. *The angular sum of any square is 360°.*

It is generally agreed that the two terms in (1) have the same intension. It is the constant function that assigns the number five to each state of affairs. Yet, as shown in (4), these terms cannot have the same meaning, as substituting the one for the other in the context of a larger expression may affect meaning.

(4)   a. *Johnny knows what two plus three is.*
       b. *Johnny knows what six minus one is.*

As a matter of fact, this substitution would even change the truth value of the resulting expression if Johnny's limited knowledge of arithmetic extends to addition but not to subtraction.

The same points can be made about the two predicates in (2) and the two sentences in (3). For, the common intension of the two predicates in (2) is the function that assigns the set of triple-sixty triangles to each state of affairs while the common intension of the two sentences (3) is the function that assigns truth to each state of affairs. Yet, the two predicates in (2) do not have the same meanings—and neither do the two sentences in (3):

(5)   a. *Johnny remembers what an equilateral triangle is.*
       b. *Johnny remembers what an equiangular triangle is.*

(6)   a. *Johnny believes the angular sum of any triangle is 180°.*
       b. *Johnny believes the angular sum of any square is 360°.*

But the problems faced by intensional semantics do not come only from the language of mathematics. Consider, for example, the case of *bought* and *sold*. Notice that these two predicates will refer to the same set of entities, as something is bought if and only if it is sold. In fact, they will refer to the same set of entities in every state of affairs. This is something that is required by the very logic of buying and selling. Yet, *bought* and *sold* do not have the same meaning. Thus, a car bought reluctantly does not mean the same thing as a car sold reluctantly. In the first case, we have the reluctance of the buyer; in the second that of the seller. And the case of *bought* and *sold* holds generally of *converse predicates* like *given/taken*, *borrowed/lent*, *leased/rented*, *imported/exported*, *immigrated/emigrated*, and so on.

To solve these kinds of problems, Carnap (1947, §§14, 15) invoked a *structural theory of meaning*. As this theory would have it, the meaning of an expression φ is not the intension of φ, but rather the way in which the intension of φ is built out of the intensions of the constituents of φ. Since the expressions in (1)–(3) all build their intensions out of different intensions, they would have to build their intensions in a different way. Consequently, they would have different meanings, and the distinctions in (1)-(3) could be drawn as desired.

To illustrate, take the contrast between *bought* and *sold*. Although these two predicates would refer to the same set of entities, one of them would build this set through the buying relation while the other would do so through the selling relation. But these are different binary relations. Although the pairs they are true of can be placed in a one-to-one correspondance in which the objects bought and sold in these pairs would be the same, the buyers and sellers in these pairs would be different—at least in all but the fictitious sales in which buyers and sellers coincide.

It follows that the structural theory of meaning is strong enough to carry possible-worlds semantics over the wall of expressions which are cointensional yet nonsynonymous. The question thus arises as to whether the structural theory of meaning is strong enough to carry *all* of the weight possible-worlds semantics can carry—or whether a structural theory of meaning can dispense with intensions altogether. The purpose of this paper is to initiate an argument that it can.

It should be acknowledged from the outset that possible-worlds semantics is a research program that has been intensively pursued for more than half a century. Consequently, it will not be possible for us, in the span of one paper, to do justice to all of the results this research program has produced. But this does not mean that we should not try to *begin* to produce these results within an extensional framework. For, if our argument is successful, semantics would only stand to gain, as we could then trade possible worlds for purely extensional semantic structures. While possible worlds are nondenumerably infinite, in both number and complexity, the semantic structures we envisage are at most denumerably infinite in both of these respects. This would represent a net gain in the computational (and hence psychological) adequacy of semantics. Moreover, these semantic structures would come at no additional cost—at least if they were already purchased in order to account for (1)–(3).

Our paper is organized as follows. We begin in §2 by sketching *PEST*, a

structural theory of meaning that is purely extensional in nature. We continue in §3 by showing how *PEST* can account for the facts that motivated possible-worlds semantics in the first place. Crucial to our accounts will be proof-theoretic views of modality and counterfactuality. We will therefore elaborate on them in §4 and §5.

At this point we shift gears and turn to objections that might be raised against *PEST*. For, it might be objected that semantic structure overdetermines meaning, that semantic structure underdetermines meaning or, at a more formal level, that no *PEST* account of self-reference is possible within set theory. §6 responds to the first of these objections by enriching *PEST* with notions of indirect interpretation and semantic equivalence; §7 addresses the second by adopting lexical decomposition; §8 meets the third by espousing *hyperset theory*— the consistent generalization of set theory that is being currently developed in response to self-referential phenomena arising in philosophy, artificial intelligence, and computer science (see Barwise and Moss, 1996). We sum up in §9 by listing the advantages that a purely extensional theory of meaning has over an intensional one. The paper also includes two appendices. *Appendix 1* develops an *PEST* grammar for a fragment of English. We hope that this grammar will make *PEST* concrete (and suggest a natural way to incorporate this theory of meaning in a grammar); *Appendix 2* shows how *PEST* can begin to account for first-order entailments.

We close this introduction by pointing out that Carnap was hardly the only one to develop a structural theory of meaning. In fact, he may not even have been the first, as a case can be made that Lewis (1944, 245f) had such a theory in mind a few years before the publication of Carnap's *Meaning and Necessity*. And, after Carnap, the structural theory of meaning has been developed in Lewis (1972, §V) and, most notably, in Cresswell (1985). It should be pointed out, however, that all of these structural theories of meaning were intended as *supplements* to possible-worlds semantics, not as *alternatives* to it. And a similar point can be made about Forbes (1989, Chapter 5), who interprets complements of verbs of attitude in terms of both structured extensions (states of affairs) and intensions (propositions).

Outside the intensional community, structural theories of meaning have been developed by proponents of Interpreted Logical Forms (Larson and Ludlow, 1993), and Russellian Annotated Matrices (Richard, 1990). Interestingly, the structural theory of meaning is supplemental here as well. In these cases, it supplements the *form* of a linguistic expression, which is taken to be as much a part of the expression's meaning as is its content (reference). Structural theories

of meaning that are purely extensional have been proposed in connection with Russellian propositions (Soames 1987, King 1995). In addition to this, computational analogs to the structural theory of meaning have been presented. Some of these are even purely extensional in nature. See for example Moschovakis (1993), where senses and references are construed, respectively, as interpretive algorithms and their values. It is with these purely extensional theories that the proposals in this paper are aligned.

§2. *A purely extensional structural theory of meaning*

There are several ways of formulating a structural theory of meaning. Here we will sketch one that is based on an *arboreal* conception of structure. We do this to underscore—and eventually to exploit—the close parallelism that exists between syntactic and semantic structures. Our arboreal formulation of a structural theory of meaning is due, essentially, to Lewis (1972, §V).

Intuitively, a purely extensional structural theory of meaning is a theory issuing from the idea that the sense of an expression is a *referential structure*—a tree which represents the way in which the reference of an expression is built from the references of its constituents (and the way in which they combine). Take for example the term in (1a). It refers to the number five by building it from the references of *two*, *plus*, *three*. Thus, given the theory we are about to sketch, the sense of this term could be the tree in (7).[1]

(7)

$$|two\ plus\ three|$$

$$|two| \qquad |plus\ three|$$
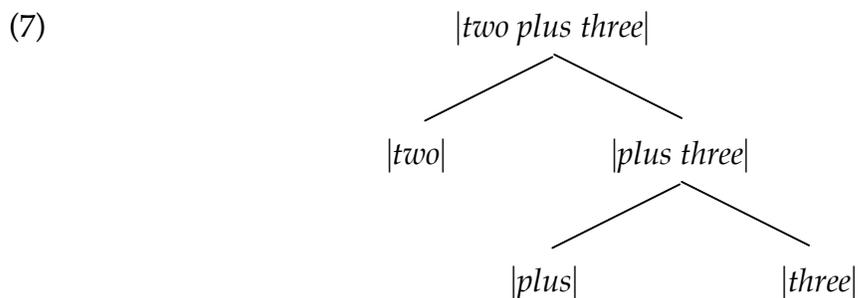
$$|plus| \qquad |three|$$

Now take the term in (1b). It too refers to the number five. Yet, it does not do so by building it from the references of *two*, *plus*, *three*. It instead arrives at it by pulling together the references of *six*, *minus*, *one*. A purely structural theory of meaning could therefore regard the tree in (8) as the sense of (1b).

---

[1] Here and henceforth we will use $||x||$ and $|x|$ to represent the sense and the reference, respectively, of any expression x.

(8)

$$|\textit{six minus one}|$$

$$|\textit{six}| \qquad |\textit{minus one}|$$

$$|\textit{minus}| \qquad |\textit{one}|$$
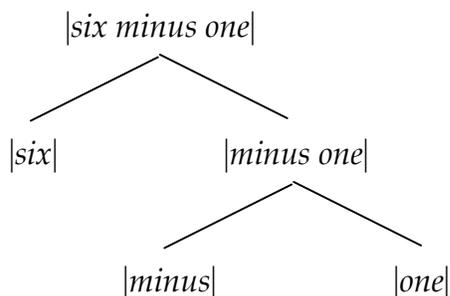
If referential trees represent senses, and if their roots represent references, then the structural theory of meaning can describe (1a) and (1b) as two expressions with different sense but identical reference. This, of course, is as desired. Yet, as we saw above, it is something that possible-worlds semantics cannot do without a structural theory of meaning. But possible worlds played no role in (7) and (8). This means that possible worlds are neither necesary nor sufficient to account for the contrast in (1).

The structural theory of meaning we are developing construes senses in terms of structures that represent the way in which an expression builds its reference. In so doing, the theory captures, in a rather natural way, the blinding Fregean intuition that the sense of an expression is the way in which that expression presents its reference (Frege, 1892). As far as I can see, this is not something that the intensions of possible-worlds semantics can do.

Related to this point, notice that our structural theory of meaning entails that the reference of an expression is a well-defined part of its sense (a root is a well-defined part of a tree). Consequently, sense would determine reference—which is again something Frege (1892) would desire. It goes without saying that the converse is not true, as different trees may have the same root. This, of course, is a welcome result; it has been the very basis for distinguishing between sense and reference in the first place.

The nodes of referential trees should not be confused with their labels. Take for example the roots of (7) and (8). Although their labels may suggest oth-erwise, these nodes are two instances of one and the same reference, namely the number five. We could therefore substitute either one of these labels for the other without altering the trees they belong to in any way. Or we could substitute either one of these labels for |*five*|. All of these substitutions will preserve the original trees *in every respect*. These identities among trees are, of course, essential to our purposes. For if it were not for them, intensionality would creep back into our trees and snuff the purely extensional nature of the theory.

Notice that all of the complex referents of (7) can be built from the *referents* of their immediate constituents. For, suppose that $|plus| = \lambda x \lambda y[x+y]$, that $|two| = 2$, and that $|three| = 3$. The complex referent $|plus\ three|$, which is $\lambda y[3+y]$, can be built from its immediate constituent referents $|plus|$ and $|three|$ by applying the former, as a function, to the latter. And the complex referent $|two\ plus\ three|$, which is 5, can be built from its own immediate constituent referents, which are $|two|$ and $|plus\ three|$. The same point can be made, of course, for the complex referents of (8). As a matter of fact, the same points can be made of most kinds of complex referents of natural language expressions.

But not all of the complex referents of a natural language expressions can be built from their immediate constituent referents. Some of them must be built from their immediate constituent *senses* instead. Take for example the sentences in (4). These sentences attribute to Johnny two *different* properties. In (4a) it is the property of knowing what two plus three is; in (4b) it is the property of knowing what six minus one is. This entails that these properties cannot be built from their immediate constituent referents, for if they could, then they would not be different properties.[2] Yet, the properties predicated of Johnny in (4) can be built from the *senses* of their immediate constituents—at least if (7) and (8) are subtrees of these senses, which is what our structural theory of meaning would require.

The properties predicated of Johnny in (4) are not the only referents that take us beyond the referents of their immediate constituents. All properties which ascribe propositional attitudes, purposes, and reports do so as well (see (14)–(16) below). As do modal assertions (17). Following standard usage, we will say that all these referents are *opaque* (as opposed to referents which can be built from the referents of their immediate constituents, which will be called *transparent*).

It is true, of course, that one could stipulate that the referents of the immediate constituents of opaque referents are their senses (or that the *indirect referents* of these constituents are their *customary senses*, as Frege, 1892, put it). But this terminological move would only save the letter of the generalization by killing its spirit. It seems preferable to forgo of the claim that the reference of an expression can be built from the referents of its immediate constituents. As we shall see below, our structural theory of meaning can do this and still abide by a principle of compositionality that requires that the sense of an expression be built

---

[2] I assume, of course, three things. First, that *knows* has the same referent in both sentences. Second, that *what two plus three is* and *what six minus one is* have the same referent in both sentences. And third, that the compositional process yielding the different properties from these constituents is the same in both sentences.

from the senses of its (immediate) constituents. Needless to say, this does not open the doors to intensionality, as senses are nothing more than referential trees (or ways in which referents combine) for us. They are not intensional functions.

Perhaps we should be more precise at this point. Let L be any natural language and let G be an adequate grammar of L. Let x be any expression of L. We shall assume (i) that x must be either semantically simple or semantically complex relative to G, (ii) that G assigns one and only one reference $|x|$ to x, (iii) that $|x|$ is assigned in accordance with a principle of compositionality that requires $|x|$ to be a function(al image) of its immediate consituent senses (meanings), of which, (iv) there are but finitely many.[3] Given these assumptions, the purely extensional structural theory of meaning we are proposing is the theory based on the premise that the meaning $\|x\|$ that G assigns to x is as follows.

(9) a. If x is semantically simple relative to G, then $\|x\|$ is the one-node tree consisting solely of $|x|$. Naturally, $|x|$ will be the root of this tree as well as its only leaf.

  b. If x is semantically complex relative to G then, by the assumptions in (ii)-(iv) above, there is one and only one n-place function φ such that $\varphi(m_1, m_2, \ldots, m_n) = |x|$, where $m_1, m_2, \ldots, m_n$ are the immediate constituent senses of $|x|$. $\|x\|$ is the smallest of the trees which satisfy the following conditions, where $1 \leq i \leq n$.

   1. All of the m_i are subtrees of $\|x\|$.
   2. $|x|$ immediately dominates the roots of all of the m_i.
   3. $|x|$ is the root of $\|x\|$.

Since all the expressions of a language will have trees for meanings, we may use the terms *sense, meaning, reference tree*, and *semantic tree* interchangeably.

We will henceforth refer to this purely extensional structural theory of meaning as *PEST*. It should be clear that, unlike the structural theories of meaning that were developed by Lewis (1943-4), Carnap (1947), Lewis (1972), Cresswell (1985), and others, *PEST* is not intended as a *supplement* to possible-worlds semantics, but rather as an *alternative* to it. *PEST* is a theory of meaning that dispenses with intensions, the key constructs of possible-worlds semantics.

If need be, we can define a *tree* as an ordered pair <N, D> in which N is a set, D is a binary relation in N, and x, y, z are variables over N such that

---

[3] As we are about to see, all the senses involved, by (iii), in the construction of a referent, will be purely extensional entities, as they will be nothing more than trees of referents. Consequently, senses will be just as extensional as referents.

(T1) $\neg\exists x[xDx]$	(*Irreflexivity*)
(T2) $\forall x\forall y[xDy \rightarrow \neg\, yDx]$	(*Asymmetry*)
(T3) $\forall x\forall y\forall z[[xDy \wedge yDz] \rightarrow xDz]$	(*Transitivity*)
(T4) $\exists x\forall y[x \neq y \rightarrow yDx]$	(*Rootedness*)
(T5) $\forall x\forall y\forall z[[xDy \wedge xD\,z] \rightarrow [yDz \vee zDy]]$	(*Noncyclicity*)

If *t* is a tree *<N,D>*, then *N* is called the set of *nodes* of *t* and *D* is called the relation of *dominance* of *t* (and is read *is dominated by* when used as an infix, as in T1-T5). The *root* of *t* is the greatest element of *N* in terms of *D*. A *leaf* of *t* is an element of *N* that is minimal in terms of *D*. A *sense* (or meaning), then, is a tree *<N,D>* in which *N* is a set of referents[4] and *D* is a relation which referents bear to those referents in whose construction they partake.

A referential tree *t* can be visualized in terms of a *Hasse diagram d* in the usual way (the vertices of *d* correspond to the nodes of *t*, and the paths of *d* correspond to the relation of dominance of *t*). We have already encountered such diagrams in (7) and (8) above. Referential trees can also be represented by labeled bracketings. Take, for example, the trees diagramed in (7) and (8). They can be represented as the bracketings in (7′) and (8′).

(7′) $[_{|two\ plus\ three|}\ |two|\ [_{|plus\ three|}\ |plus|\ |three|\ ]]$

(8′) $[_{|six\ minus\ one|}\ |six|\ [_{|minus\ one|}\ |minus|\ |one|\ ]]$

Although compact, labeled bracketings can be rather terse. We will therefore choose to represent semantic trees in the sequel in terms of Hasse diagrams rather than labeled bracketings. It should be pointed out, however, that for reasons to be given below, the left-to-right ordering of the nodes of these diagrams is—like the linear order of the bracketed items in the corresponding labeled bracketings—of no theoretical significance. These orderings are given only to aid the reader in identifying the expressions whose meanings are being represented.

Semantic trees are useful construals of the meanings of semantically complex expressions. But they play an important role for semantically simple expressions as well, as they allow us to distinguish between their senses and their references. For, let *x* be a semantically simple expression. Its sense $\|x\|$ in *PEST* is a tree *<N, D>* in which *N* is the singleton of $|x|$ and *D* is the empty set.

---

[4] Or rather *reference tokens*, as a reference may serve as two or more nodes.

This is, to be sure, a degenerate tree. Yet, it is still different—and richer—than the sole element |*x*| of *N*, which would be the reference of *x*. The semantic trees of *PEST* can therefore allow us to uphold the difference between sense and reference regardless of the semantic complexity of an expression. And this can in turn allow us to capture generalizations. Consider, for example the clause in (9b) above. If it were not for the fact that semantically simple expressions have a distinct sense (= a tree), we would have to split (9b) in two clauses: one for preterminal nodes (which would lack constituent subtrees) and another one for higher nodes (which would have them).

Notice that *PEST* abides by a principle of compositionaity that requires that the sense of a complex expression be a function(al image) of its immediate constituent senses. For, recall that we have assumed in (iii) above that the referent of a complex expression is a function of its immediate constituent senses. And, according to (9b), the sense of an expression is determined by its referent and its immediate constituent senses. Putting these two facts together, the sense of a complex expression is indeed a function of its immediate constituent senses.

Finally, notice that, unlike syntactic trees, semantic trees do not linearly order their nodes. This should be as desired. For, while linear order is morphosyntactically real (it corresponds to the order in which the constituents are uttered over time), linear order does not correspond to anything real in the composition of meanings. It should be emphasized that this does not mean that morphosyntactic linear order is devoid of meaning. This order may well have a semantic effect. And if it does, *PEST* will be able to account for it. For, notice that the arguments m_1, m_2, …, m_n of φ in (9b) are *ordered*. And the order of these arguments could *both* reflect the linear order of the morphosyntactic expressions that contributed these arguments *and* affect the value that φ assigns to them. If it did, then *PEST* would have acknowledged—and accounted for—a semantic effect of linear order.

Readers in need of a concrete example of the theory delineated in (9) may turn now to the *Appendix*, where a grammar that incorporates that theory is presented. These readers should be warned, however, that a full understanding of the workings of that grammar presuppose material presented in the other sections of this paper.

§3. Beginning to cover the data

Intensional semantics was developed in order to solve a number of prob-

lems that drove extensional semantics into the ground. Chief among them were (i) the variable informativeness of identity statements, (ii) the failure of substitution in opaque contexts, (iii) the compositional interpretation of modal verbs and adverbs, (iv) the nontrivial nature of counterfactuals, and (v) the nonsynonymy of vacuous predicates. The purpose of this section is to show that *PEST* can solve all of these problems as well as intensional semantics can (if not better). To do so *PEST* will need to be coupled with proof-theoretic interpretations of modality (Boolos, 1993) and counterfactuality (Rescher, 1964).

Consider first (i). While the identity statement in (10a) is informative, the identity statement in (10b) is not.

(10)   a. *The morning star is the evening star.*
       b. *The morning star is the morning star.*

This is a problem for traditional, nonstructural, extensional semantics. For, since *the morning star* and *the evening star* have the same extension, extensional semantics should predict that both (10a) and (10b) make the perfectly uninformative statement that something—in this case a planet—is itself. But this prediction is not borne by the facts, as only (10b) is making such an uninformative statement.

To solve this problem, intensional semantics notes that *the morning star* and *the evening star* could have referred to different entities under different circumstances, and therefore assigns different intensions to these noun phrases. (10a) could now be informative—at least if we assume that the purpose of identity statements is to assert that two different intensions have the same extension in the real world. (10b), on the other hand, would be completely uninformative, as it would be asserting that the value some function assigns to some argument is the value that function assigns to that argument (where the function is the intension of *the morning star*, the argument is the real world, and the value is the extension of *the morning star* in the real world).

But *PEST* can account for the difference in informativeness between (10a) and (10b) as well. In fact, it may even be slightly better at it than the intensional approach. Notice first that *PEST* will assign different meanings to *the morning star* and *the evening star*:

(11) a. |*the morning star*| b. |*the evening star*|

```
        |the|         |morning star|              |the|         |evening star|

                    |morning|    |star|                    |evening|    |star|
```
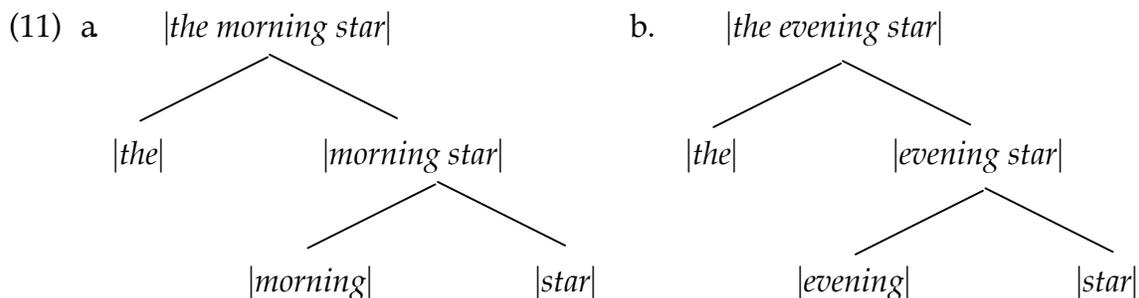
Notice next that these differences in meaning can explain why (10a) is an informative statement while (10b) is not—at least if we assume that the purpose of identity statements is to assert that some extension can be built in two different ways. For, (10a) would be stating that the extension built as indicated in (11a) is the same as the extension built as indicated in (11b). (10b), on the other hand, would be saying that the extension built as indicated in (11a) is the extension built as indicated in (11a).

In fact, notice that *PEST* is able to attribute the semantic difference between *the morning star* and *the evening star* in terms of the *actual* reference of *morning* and *evening*. The intensional approach would instead attribute this difference to the *potential* reference of these nouns. The *PEST* account of this difference is therefore simpler, less abstract, and more intuitive than that of the intensional account.

It should go without saying that the structural nature of meaning played a crucial role in the *PEST* account of the difference between (10a) and (10b). For, if it weren't for the difference in the way their extensions are built, *the morning star* and *the evening star* would be semantically indistinguishable, and the difference in the informativeness of (10a) and (10b) would vanish.

Turning next to issue (ii), notice that the sentence in (12a) may be true while the sentence in (12b) may be false.

(12)   a. *Chris believes that Paris is in France.*
        b. *Chris believes that Lima is in Peru.*

This is a problem for nonstructural versions of extensional semantics. For, given that the two sentential complements in (12) have the same truth value, extensional semantics would assign them the same meaning. Consequently, we should be able to substitute one of these complements for the other without affecting truth value—at least if we assume a natural principle of substitution that ensures that meaning is preserved under the substitution of synonyms for

synonyms. Yet, as we have seen, this is not the case. (12a) could well be true while (12b) could be false (and *vice versa*).

To save some natural principle of substitution, intensional semantics observes that the two sentential complements in (12) could have different truth values in different states of affairs, and therefore assigns them different intensions. The two complements would thereby have different meanings, and need not be interchangeable *salva veritate*.

But *PEST* can account for the failure of substitution in (12) as well. The sentential complements of (12) fail to substitute for each other because they differ in meaning, and they differ in meaning because they will have reached their truth values in different ways. While one will have reached its truth value through the extensions of *Paris* and *France,* the other will have done so through the extensions of *Lima* and *Peru*.
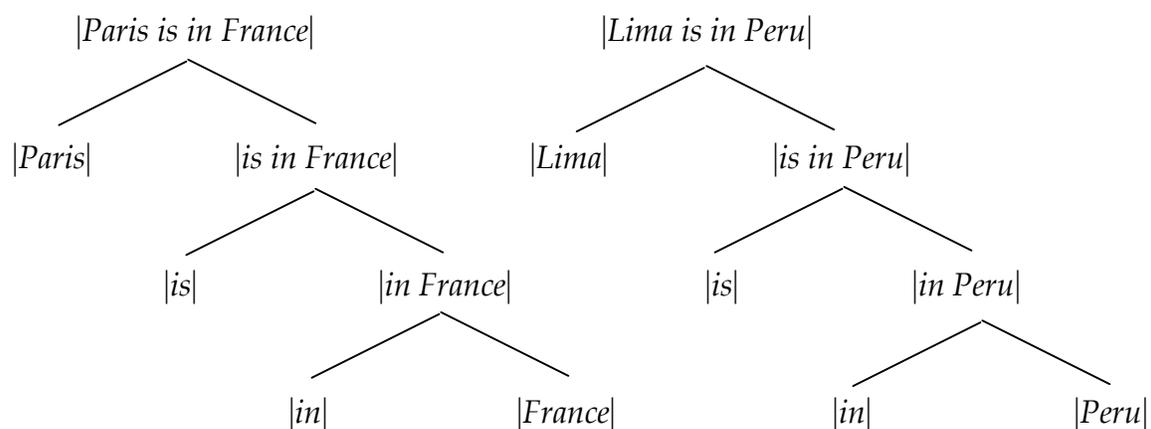
We will say that the trees above are examples of *propositions*—referential trees that describe the way in which sentences acquire their truth values (or, more simply, referential trees whose root is a truth value).

It should be pointed out that failure of substitution of coextensional expressions affects not just entire belief complements, but *parts* of belief complements as well. Thus, the two sentences in (13) may well differ in truth value—let alone meaning—even though *Paris* and *my favorite city* have exactly the same reference.

(13)   a. *Sandy believes that Paris is in France.*
       b. *Sandy believes that my favorite city is in France.*

And *PEST* fares just as well as the intensional approach to these partial

substitutions.

The *PEST* account of the failure of substitution of coextensionals in beliefs will account also for the failure of substitution of coextensionals in the complements of all verbs of propositional attitude (*suspect, discover, know, doubt, forget, remember*), purpose (*seek, want, hope, wish, try*), report (*say, suggest, mutter, quip, chime*), and modality (*must, can, may, shall, will*):

(14)   a. *Pat suspects that Dr. Seuss wrote this story .*
          b. *Pat suspects that Theodore Geisel wrote this story.*

(15)   a. *Robbie is trying to figure out what the positive square root of 2500 is.*
          b. *Robbie is trying to figure out what the number of states in the USA is.*

(16)   a. *Lois said she saw Superman fly by.*
          b. *Lois said she saw Clark Kent fly by.*

(17)   a. *Nine must be smaller than ten.*
          b. *The number of planets must be smaller than ten.*

In short, *PEST* seems to be able to account for the failure of substitution of coextensionals (and their parts) in all opaque contexts.[5]

It should not escape the reader that the structural nature of meaning played a critical role in the preceding account of the failure of substitution. For, if it weren't for their structural difference, the coextensionals in question would be semantically indistinguishable, and hence intersubstitutable. And, as for (i), the *PEST* account of (ii) is based on an extensional semantic difference between the expressions in question, not on an intensional one—as the intensional approach would require. Gains in simplicity, concreteness, and intuitiveness would thus continue to accrue.

Let's turn next to (iii). It follows from the laws of logic that George Bush, the current president of the United States, is identical to himself, not that he is married to Laura Bush, as he currently is. Thus, if the sentences in (18) involve *logical necessity*, then (18a) is true and (18b) is false.

---

5 At this point, readers familiar with the thorny issues surrounding the interpretation of proper names will appreciate the fact that *PEST* can incorporate both the Millian and the Russellian interpretations of proper names (although not at the same time, of course). See §7 below, where we will also consider the possibility that the sense of a simple proper name *N* is a two-node tree with root |*N*| and leaf *N*.

(18)   a. *George Bush has to be George Bush.*
        b. *George Bush has to be married to Laura Bush.*

Similarly, while the laws of logic allow George Bush to be married to someone other than Laura Bush, they do not allow George Bush from being something other than himself. Thus, if the sentences in (19) involve *logical possibility*, then (19a) is true while (19b) is false.

(19)   a. *George Bush might not be married to Laura Bush.*
        b. *George Bush might not be George Bush.*

These facts create a serious problem for traditional extensional semantics—at least if this semantics is to operate compositionally (Dowty, Wall, Peters, 1981, 141f). For, here the necessity operator must combine with truth and yield truth in (18a) but falsity in (18b). Similarly, the possibility operator must combine with falsity and yield truth in (19a) but falsity in (19b). Needless to say, this indeterminacy of interpretation is inadmissible under the Principle of Compositionality, which requires that the meaning of a complex expression be a *function* of the meanings of the parts (and the way they combine).

To overcome this problem, traditional extensional semantics could extend its set of truth values by subdividing plain truth into *necessary truth* and *contingent truth*, and by subdividing plain falsity into *necessary falsity* and *contingent falsity* (Kearns, 1981). Modal operators could be made to be truth functional under this setup. Yet, two new truth values would be called for. Moreover, this solution would not shed much light on the meanings of the modal operators. For, if taken as well-defined truth-functional operators, *necessarily p* comes out as necessarily true if and only if *p* is necessarily true, and *possibly p* is rendered necessarily true if and only if *p* is not necessarily false.

The facts in (18) and (19) are not a problem for intensional semantics—at least if we adopt an interpretation of modality in which the necessity operator is interpreted as a function that assigns (i) truth to the constant function that assigns truth to every possible world and (ii) falsity to every other function from worlds to truth values, while the possibility operator is interpreted as a function that assigns (i) falsity to the constant function that assigns falsity to every possible world and (ii) truth to every other function from worlds to truth values.

But the facts in (18) and (19) need not be a problem for *PEST* either—at least if we adopt the proof-theoretic interpretation of modality developed in

Provability Logic (Boolos, 1993). For if we do, then we could define the key modal operators (on any particular occasion of their use) as shown in (20), where *x* is a variable over propositions and Γ is a constant that stands for the set of propositions which are given, postulated, or simply taken for granted (on the particular occasion in which the interpreted operators are used).[6]

(20)   a. $\Box = \lambda x[\Gamma \vdash x]$
        b. $\Diamond = \lambda x[\neg\Gamma \vdash \neg x]$

(20a) says that the necessity operator (on any occasion of its use) is the function that characterizes the (set of) propositions that are provable from the (set of) propositions that are taken for granted (on that occasion of use). (20b) states that the possibility operator (on any occasion of its use) is the function that characterizes the (set of) propositions that are not disprovable from the (set of) propositions that are taken for granted (on that occasion of use).

It follows that, for any proposition *p*, the propositions $\Box p$ and $\Diamond p$ would be interpreted, compositionally, as shown in (21).

(21)   a. $\Box p = \Box(p) = \Gamma \vdash p$
        b. $\Diamond p = \Diamond(p) = \neg\Gamma \vdash \neg p$

(21a) says that a proposition is necessary (on a particular occasion of its use) if and only if it is provable from the set of propositions that are taken for granted (on that occasion of use). (21b) says that a proposition is possible (on a particular occasion of its use) if and only if it is not disprovable from the set of propositions that are taken for granted (on that occasion of use).

Now, let us say that a proposition *p* is *consistent* with a set of propositions if and only if *p* is not disprovable from them. If we construe consistency in this natural way, then (20b) states that the possibility operator (on any occasion of its use) is the function that characterizes the (set of) propositions that are consistent with the (set of) propositions that are taken for granted (on that occasion of use). Similarly, (21b) asserts that a proposition is possible (on a particular occasion of its use) if and only if it is consistent with the (set of) propositions which are taken

---

[6] To emphasize similarities with traditional theories of modality, we have adopted a contextual interpretation of the negation symbol '¬'. We may therefore use this symbol coherently both with object-language expressions '*p*' and with metalanguage expressions 'Γ $\vdash$ *p*'. If contextual interpretations are to be avoided—or if differences in logical type are to be emphasized—we could reserve the negation symbol '¬' for object-language expressions '*p*', and use a different symbol, say the slashed turnstile '$\nvdash$', for the negation of metalinguistic provability statements 'Γ $\vdash$ *p*'.

for granted (on that occasion of use).

The set Γ of givens mentioned in (20) and (21) is quite variegated. It contains a set Γ_L of logical givens, a set Γ_P of physical givens, a set Γ_E of epistemic givens, a set Γ_D of deontic givens—and so on. Each one of these sets will give rise to a corresponding notion of modality. Thus, since the sentences in (18) and (19) were taken to convey logical modalities, they would be interpreted, compositionally, as follows,[7]

(18′) a. $\Gamma_L \vdash$ *George Bush is George Bush*
      b. $\Gamma_L \vdash$ *George Bush is married to Laura Bush*
(19′) a. $\neg\Gamma_L \vdash \neg$*George Bush is not married to Laura Bush*
      b. $\neg\Gamma_L \vdash \neg$*George Bush is not George Bush*

where (19′) reduces to (19′′) by the Law of Double Negation.

(19′′) a. $\neg\Gamma_L \vdash$ *George Bush is married to Laura Bush*
       b. $\neg\Gamma_L \vdash$ *George Bush is George Bush*

Notice that (18a) will come out true and (18b) false under these analyses. For, while there exist logical postulates that imply that George Bush is George Bush (18′a), there are no logical postulates that require that George Bush be married to Laura Bush (18′b). Similarly, notice that (19a) will be true and (19b) false under these analyses. For, while there are no logical givens that imply that George Bush is married to Laura Bush (19′′a), there *are* logical givens that imply that George Bush is George Bush (19′′b).

The correct compositional interpretation of the sentences in (18) and (19) can therefore be attained within *PEST* as well. In fact, it may be argued that *PEST* can handle these facts to advantage, as it can also account for the intuitive rationale behind the truth values assigned to these sentences. Indeed, if someone were to ask why the sentences in (18) and (19) were true or false, an appeal to logical givens would be eventually made—even if truth values in alternative worlds, states, or circumstances were learnedly invoked. As Hamlyn (1967) put it,

> In ordinary circumstances the assertion "That statement must be true" is likely to be made at the conclusion of an argument, and the use of the words "That must be true" invites completion with "because…". The argument in question may be one of a number of different kinds […] In all such cases we invoke reasons […]

---

[7] To enhance readability, we let sentences stand here for the propositions they convey.

this is so often the case that to say in any context "It must be true" is to invite the question "Why?," a question that can be answered only by providing the relevant reasons. In other words, a statement that must be true may be taken as one whose truth is in an important sense necessitated by the reasons adduced in its support (Hamlyn, 1967, 198).

In addition to being more intuitive, the proof-theoretic view of modality is simpler. The set of provables of predicate logic is, for example, recursively enumerable, and the set of provables of monadic predicate logic (which includes all of propositional logic) is recursive (Partee et al., 1990, §8.6.5). The set of possible worlds, on the other hand, is not even denumerable—let alone recursively so.

Once again, the structural theory of meaning played a crucial role—this time in our analyses of (18) and (19). For, if it weren't for this theory, the modal operators would have to apply to truth values rather than to propositions. But this would lead to incoherence. For, while one can ask whether a proposition is provable or consistent, it makes no sense to ask whether a truth value is.

Although the theory of modality sketched in (20) and (21) is not new, it is little known outside of Provability Logic. We will therefore survey its noteworthy tradition in §4 below. We will also offer some formal comparisons between the proposed and the traditional views of modality. In the meantime, let us turn to (iv) and consider the sentences in (22), which are adapted from Chierchia and McConnell-Ginet (2000, 258).

(22)   a. *If Proust had been on the Titanic, he would probably have been among the fatalities of the shipwreck.*
       b. *If Proust had been on the Titanic, he would probably have been among the survivors of the shipwreck.*

Intuitively, (22a) is true and (22b) is false. Yet, traditional extensional semantics would predict that both (22a) and (22b) are true—at least if these sentences involved material implication, since their common antecedent is false, and any sentence expressing a material implication from a false antecedent is, by definition, true. More generally, traditional extensional semantics would predict, incorrectly, that all counterfactuals are trivially true (and hence trivial).

Notice that the problem (22) poses for traditional extensional semantics is quite acute, as the only difference between the sentences in (22) is the propositional content of their consequents. Yet, this content is completely invisible to that framework: since the two consequents are false, they are extensionally indis-

tinguishable from each other.

Based on the observation that the common antecedent of the sentences in (22) *could* be true, possible-worlds semantics can reasonably claim that counterfactuals should be interpreted in a world w which, though unreal, is still possible. If we furthermore assumed that w was *as close as it could be to the real world*, then we could account for why (22a) was felt to be true and (22b) false. For, suppose w was a possible world in which the passenger manifest of this world's *Titanic* included Marcel Proust. Suppose also that w was as close to the real world as possible. It seems clear that w would include a massive loss of life brought about by the liner's fateful shipwreck. Thus, if interpreted in w, (22a) would indeed be true and (22b) false as desired (see Lewis, 1973).

But *PEST* can also account for counterfactual reasoning—at least if we can rework the preceding notion of closeness into the proof-theoretic view of counterfactuals developed in Goodman (1947), Rescher (1964), and McCawley (1993, 538ff). In the view we envisage, the interpretation of a counterfactual $p \rightarrow q$, on any particular occasion of linguistic use, would be given by (23),

(23) $p \rightarrow q \equiv \exists \Gamma'[\Gamma' \vdash q]$

where Γ' is a set of propositions that includes p, is consistent with p, and is otherwise as close as possible to the set Γ of propositions that are *actually* given, postulated, or taken for granted on that occasion of use. If it exists, Γ' is intended to be a set of propositions which are only *hypothetically* given, postulated, or taken for granted; it is a set of propositions which are *assumed solely for the sake of argument.* So (23) states that a counterfactual is true iff its consequent is provable from a suitable set of hypothetical givens.

The proof-theoretic analysis of counterfactuals we have just sketched yields the contrast in (22)—if a suitable notion of closeness can be defined and justified. For, note that there exists a consistent set Γ' of propositions that includes the claim that Proust was a passenger of the *Titanic* as well as the fact, already present in the set of actual givens, that the *Titanic* shipwrecked in its maiden voyage in an accident that led to the death of most of its passengers. Note also that it follows from this Γ' that Proust was probably among the fatalities of the shipwreck—which in turn means, given (23), that (22a) is true. On the other hand, there is no consistent set Γ' of propositions that includes the claim that Proust was a passenger of the *Titanic* (and is as close as possible to the actual set of givens) from which it can be proved that Proust was probably among its survivors. But this means, according to (23), that (22b) is false.

Note that the possible-worlds account of counterfactuals succeeds by in-terpreting counterfactuals in worlds that satisfy their antecedents. In such worlds, the truth values of material implications with true antecedents depend solely on the truth values of the consequents—which is as desired. If successful, the proof-theoretic account of counterfactuals we have sketched would owe its success to the fact that proof is weaker than material implication. For, while any-thing is materially implied by a false antecedent, not everything is provable from false premises.

It would be impossible to present, in a paper of this scope, a full defense of the proof-theoretic view of counterfactuals. It should be pointed out, however, that this view seems to be able to account for the well-known paradoxes of counterfactual reasoning—namely the failure of counterfactual transitivity, the failure of reinforcement of the counterfactual antecedent, and the failure of counterfactual contraposition. See §5 below, which is where we have placed the discussion of these paradoxes in order to maintain the focus of this section.

But before we leave the topic of counterfactuals, we should note that the points about simplicity we made for the *PEST* account of the modal facts in (18)-(19) apply also to the *PEST* account of the counterfactual facts in (22), as both accounts are proof-theoretic rather than intensional.

We should also emphasize that the structural nature of *PEST* is crucial to any success we can claim in (22). For, if we had an extensional theory of meaning which was not structural in nature, then the propositional content of the antecedents and the consequents of the sentences in (22) would remain invisible, and the provability or unprovability of their consequents would be unattainable. For provability and unprovability are relations among propositions, not among truth values.

Turning finally to (v), note that the two sentences in (24) make completely different assertions.

(24)   a. *A unicorn set its head on the maiden's lap.*
       b. *A cyclops set its head on the maiden's lap.*

Yet, accounting for the difference between these assertions would be impossible for traditional extensional semantics—at least if it were to regard the empty set as the common reference of both *unicorn* and *cyclops*.

Intensional semantics has no problem accounting for the contrast in (24). Since there is no necessary connection between the existence of unicorns and cyclops, possible worlds may be found in which only one of these two kinds of mythical creatures exists. Consequently, *unicorn* and *cyclops* could be assigned different intensions, and the two sentences in (24) would make predictably different assertions as desired.

Intensional analyses work well with predicates which, like *unicorn* and *cyclops*, are *contingently* empty. Yet, they reach an impasse with predicates that are *necessarily* empty. These predicates are legion. Take, for example, the ones in (25).

(25)   a. *square circle*
       b. *married bachelor*
       c. *four-sided triangle*
       d. *number divisible by zero*
       e. *real square root of a negative number*

As intensional semantics would have it, all of the predicates in (25) would be referentially empty in every possible world. Consequently they would have to be assigned the same intension, and be regarded synonymous. This, of course, would be incorrect. For proof that one of these predicates is absurd is not proof that all of them are.

PEST has no problem distinguishing among the predicates in (25)—even if all of them were equally absurd. For, according to *PEST*, all of these predicates would refer to the empty set. Hence, they would all be equally absurd. Yet, this referential emptiness would have been arrived at in a different way for each one. And each of these ways would have been preserved in a different compositional structure. This means that *PEST* can predict that the predicates in (25) would all differ in meaning (and insofar as each of these meanings corresponds to a distinct proof of referential emptiness, proof that one of them was absurd would not be proof that all of them are).

Note that this analysis extends naturally to contingently empty predicates like *unicorn* and *cyclops*. For these nouns can also be said to use different means (say, single-horned *vs.* single-eyed foreheads) to attain their common, empty, reference. Since each of these means would be preserved in the predicates' compositional structures, the appropriate semantic difference would emerge between them as well.

It must be acknowledged that this analysis of (24) requires that predicates like *unicorn* and *cyclops* involve more semantic structure than meets the eye, and some readers may have reservations about admitting such structure into their analyses. For such readers, *PEST* allows another account of the contrast in (24). It consists in admitting fictional entities like unicorns and cyclops into the universe of discourse (McCawley 1993, 428). A universe of discourse is, after all, the set of things that can be talked about, not the set of entities that exist. Now, since unicorns and cyclops are different fictional creatures, the contrast in (24) can be straightforwardly accounted for. We should point out, however, that this analysis does not seem available for predicates like the ones in (25), as they would require entities which are not just fictional, but downright absurd (and thus nonexistent).

Moreover, we will come across further evidence for "invisible" semantic structure, and that this evidence will not come from entities from fictional domains (see §6 below). So hypothetical semantic structure may well be unavoidable. But that is a different issue; all we needed to establish at this point is that *PEST* can solve the problem posed by vacuous predicates. And, as we have seen, it can. In fact, it can do so in more than one way.

§4. *The proof-theoretic view of modality*

The proof-theoretic view of modality we have sketched in (20) is hardly new. It can be traced back to Frege's *Begriffsschrift*.[8] Here statements of necessity were taken to hint at the existence of premises from which these statements may be inferred, while statements of possibility were said to suggest the ignorance of laws from which the negatives of these statements would follow (Kneale and Kneale, 1962, 548):

> The apodictic judgement differs from the assertory in that it suggests the existence of universal judgements from which the proposition can be inferred, while in the case of the assertory one such a suggestion is lacking. By saying that a proposition is necessary, I give a hint about the grounds for my judgement […]
>
> If a proposition is advanced as possible, either the speaker is suspending judgement by suggesting that he knows no laws from which the negative of the proposition would follow or he says [equivalently] that the generalization of the

---

[8] And perhaps even to Leibniz. For, according to Hacking (1994, 31), Leibniz understood well that "the concept *logically necessary proposition* is to be explained in terms of proof, and not in terms of truth in all possible worlds (as if necessity represented some constraint on what worlds can be created)."

negation is false (Frege, 1879, §4).

A construal of necessity in terms of provabilty from certain axioms was explored also by Kurt Gödel, who wedded this construal to an unusual, syntactic, interpretation of modality. As shown by Richard Montague and David Kaplan, however, these decisions led to ultimately unsatisfactory results. Fortunately, none of these results obtains if we drop the attempt to make necessity a syntactic predicate attaching to *names* of propositions, and use instead the necessity operator as a propositional operator in the usual way (Prior 1967, 8f).

In §65 of his classic *Elements of Symbolic Logic*, Reichenbach drew a distinction between absolute and relative modality. While absolute modality deals with statements that are made "without reference to special conditions", relative modality pertains to statements that are made "with reference to other facts". He then proposed the following proof-theoretic definitions of these modalities:

(26) **Absolute Modalities**
   a. p is logically necessary $=_{df}$ 'p' is a tautology
   b. p is logically possible $=_{df}$ neither 'p' nor '$\neg p$' is a tautology

(27) **Relative Modalities**
   a. q is logically necessary relative to p $=_{df}$ 'q' is derivable from 'p'
   b. q is logically possible relative to p $=_{df}$ neither 'q' nor '$\neg q$' is derivable from 'p'

The opening paragraph of this paper incurred in a simplification which we must now confess to. Strictly speaking, Carnap (1947) defined intensions, not as we said, in terms of states of affairs, but rather in terms of *descriptions* of states of affairs:

> A class of sentences in [a particular language system] S1, which contains for every atomic sentence either this sentence or its negation, but not both, and no other sentences is called a **state-description** in S1, because it obviously gives a complete description of a possible state of the universe of individuals with respect to all properties and relations expressed by predicates of the system. Thus the state-descriptions represent Leibniz' possible worlds or Wittgenstein's possible states of affairs […] The class of all those state-descriptions in which a given sentence C_i holds is called the range of C_i […] [The rules that determine the ranges of sentences give] an interpretation for all sentences in S1, since to know the meaning of a sentence is to know in which of the possible cases it would be true and in which not, as Wittgenstein has pointed out (Carnap, 1947, §2).

True to this *linguistic* view of possible worlds, Hintikka (1961) developed a semantics for modal logic in which necessity and possibility were defined in terms of sets of propositions rather than worlds. Consider (28) in this regard, where p is a proposition and w is a maximally consistent set of propositions (henceforth *modset*).

(28) a. $\Box p \equiv \forall w[p \in w]$
      b. $\Diamond p \equiv \exists w[p \in w]$

As might be expected, Hintikka's semantics of modal logic is equivalent to Kripke's. Yet, Hintikka's modsets may be more tractable than Kripke's worlds. Moreover, by distinguishing a set L of laws from each modset, Dunn (1973) was able to define Kripke's accessibility relation between worlds in terms of what propositions are chosen as laws from the modsets that correspond to those worlds. He also showed that these choices correspond to the celebrated axiomatic systems of modal logic (*M*, *S4*, *B*, *S5*) discussed by Lewis and Langford (1946). Using these laws, Sowa (2002) defined the modal operators as follows.

(29) a. $\Box p \equiv L \models p$
      b. $\Diamond p \equiv \neg L \models \neg p$

These analyses are, of course, very close to the ones we gave in (21). In fact, they differ only insofar as they involve a set of *laws* that relate to the modalized propositions in terms of *entailment*. Our proposals instead involve a set of *givens* that relate to the modalized propositions in terms of *provability*.

Finally, although provability logicians do not claim to provide the *intended* interpretation of the modal operators of modal logic, Boolos does suggest that Quine (or someone of a Quinean persuasion) might just make that claim:

> Far from undermining Quine's critique of modality, provability logic provides an example of the interpretation of the box whose intelligibility is beyond question. Quine has never published an opinion on the matter, but it would be entirely consonant with the views he has expressed for him to hold that provability logic is what modal logicians should have been doing all along (Boolos, 1993, xxxiii-xxxiv).

Philosophical persuasions aside, we *have* advanced the interpretation of the modal operators developed by provability logic as the intended interpretation of the modal operators employed by natural language.

Once modal operators are interpreted proof-theoretically, modal logic becomes a tool for investigating provability and consistency in a rigorous way. But modal logic is not a particular logical system; it is a large family of logical systems. We have already mentioned the four systems *M*, *S4*, *B*, *S5* discussed by Lewis and Langford (1946). But many others have been proposed (Garson, 1993, §8). Of all these systems, provability logicians have selected one for special attention. It is *GL*, a system of propositional modal logic named in honor of Gödel and Löb. *GL* has only two axioms (in addition to all of the tautologies of propositional logic). They are (K) and (W), where *p* and *q* are propositions which may, of course, contain modal operators.

(K) (Kripke) $\Box(p \rightarrow q) \rightarrow \Box p \rightarrow \Box q)$ (*Modal Distribution*)
(W) (Löb) $\Box(\Box p \rightarrow p) \rightarrow \Box p$ (*Minimal Soundness*)[9]

In addition, *GL* has only one rule of inference (other than *Modus Ponens*). It is the binary relation consisting of all the ordered pairs of the form in (N), where *p* is, again a proposition which may contain modal operators.

(N) (Kripke) $(p, \Box p)$ (*Necessitation*)

GL is *normal* in the sense—now standard—that its set of theorems contains all of the instances of the distribution axiom, all of the tautologies of propositional logic, and all of the consequences of *Modus Ponens*, *Necessitation*, and *Substitution* (of which the third is a rule of inference that can be derived in *GL*). In fact, *GL* is little more than the very definition of normality, as it exceeds it only by requiring *Minimal Soundness*.[10]

So *GL* is the tool of choice for investigating provability and consistency in a logical system. Here we will use it to study the logical systems required by our proof-theoretic view of modality. They are the systems Γ whose only (nonlogical) axioms are the propositions which are given, postulated, or simply taken for granted on a particular occasion of linguistic use, and have no (nonlogical) rules of inference. To do this, all we need to do is let the boxes of (K), (W), (N) stand for *provability in* Γ, and allow the propositional variables in (K), (W), (N) to range

---

[9] (W) is better known in the literature as *Löb's Theorem*. I have taken the liberty to dub it here *Minimal Soundness* to underscore the fact that this axiom requires a proposition to be (provably) *sound* only when it absolutely has to, namely when it is *provable* (Boolos, 1993, 55).

[10] To see how small an excess this is, notice that the original definition of normality (Kripke, 1963b) required full soundness (see text below). As to the rationale behind normality, this is the only property all the systems of modal logic surveyed in Garson (1993) have in common.

over propositions (well-formed trees whose roots are truth values).

Using *GL* to investigate our logical systems we can show that the proof-theoretic view of modality we have proposed (henceforth *PTM*) shares a number of formal properties with the traditional ones. Notice, for example, that the two modality operators are interdefinable under the proposed interpretation, since

(30) $\Box p \equiv \Gamma \vdash p \equiv \neg \neg \, \Gamma \vdash \neg \neg p \ \equiv \neg \Diamond \neg p$

(31) $\Diamond p \equiv \neg \Gamma \vdash \neg p \equiv \neg \Box \, \neg p$

In fact, they are interdefinable under the proposed interpretations *in exactly the same way* as they are under their traditional interpretations. Furthermore, both *iterated modalities*, (32a) and (32b), are provable in *GL*.[11]

(32) a. $\Box p \rightarrow \Box \Box p$
      b. $\Diamond p \rightarrow \Diamond \Diamond p$

(32a) says that if some proposition is necessary, then it is necessary that it is necessary; (32b) says that if some proposition is possible, then it is possible that it is possible.

Whether (32a) and (32b) should be valid in modal logic is "both a disputed question and one of some obscurity," write Hughes and Cresswell (1996, 51f), who add that

> It is, however, at least a reputable and plausible view that in certain well-established senses of 'necessary', it should be answered in the affirmative; it is, for example, plausible to maintain that whenever a proposition is *logically* necessary, this is never a matter of accident but is always something which is logically bound to be the case (Hughes and Cresswell, *loc. cit.*).

(32a) and (32b) are valid, for example in *S4* and *S5*, which are the most widely known among the systems of modal logic in Lewis and Langford (1946). And in the context of provability logic—a context, if there ever was one, in which necessity is interpreted logically—they are perhaps more than justified. Had (32a) not been a theorem of *GL*, writes Boolos (1993, 11), "we should have been interested in the smallest normal extension of *GL* in which it was one!"

But *PTM* also differs from the traditional view of modality in certain

---

[11] See Boolos (1993, 11f) for a proof of (32a). (32b) follows from (32a) by the interdefinability of necessity and possibility: $\Diamond p \equiv \neg \ \neg p \rightarrow \neg \ \ \neg p \equiv \neg \neg \Diamond \neg \neg \Diamond \neg \neg p \equiv \Diamond \Diamond p$.

formal respects. Notice that, according to (21), every possibility sentence $\neg\Gamma \Vdash \neg p$ involves two negations of the necessity sentence $\Gamma \Vdash p$: the *internal* negation of p and the *external* negation of $\Gamma \Vdash p$ as a whole. Since either one of these negations may occur without the other, we must acknowledge the existence of two additional propositions: $\Gamma \Vdash \neg p$ and $\neg\Gamma \Vdash p$. The four propositions resulting from this discussion are listed in (33). They may be said to form the *closure of* $\Gamma \Vdash p$ *under internal and external negation.*

(33)   a. $\Gamma \Vdash p$
       b. $\Gamma \Vdash \neg p$
       c. $\neg\Gamma \Vdash p$
       d. $\neg\Gamma \Vdash \neg p$

Relative to a set Γ of givens, the propositions in (33) assert that p is provable (33a), disprovable (33b), unprovable (33c), and undisprovable (33d).

It is at this point that *PTM* and its traditional counterpart begin to diverge. Notice that the propositions in (33) may form the square of opposition in (34), *but only if* Γ *is consistent*. For if it isn't, then we would not be able to ensure (a) that the two contraries are not both true, (b) that the two subcontraries are not both false, and (c) that the superalternates imply the subalternates.

(34)

$$\Gamma \Vdash \boldsymbol{p} \text{———} \Gamma \Vdash \neg\boldsymbol{p}$$
$$\neg\Gamma \Vdash \neg\boldsymbol{p} \text{———} \neg\Gamma \Vdash \boldsymbol{p}$$
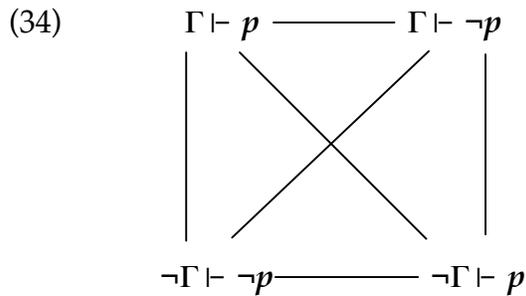
This square of opposition can be represented in terms of necessity and negation:

(35)

$$\Box\boldsymbol{p} \text{———} \Box\neg\boldsymbol{p}$$
$$\neg\Box\neg\boldsymbol{p} \text{———} \neg\Box\boldsymbol{p}$$
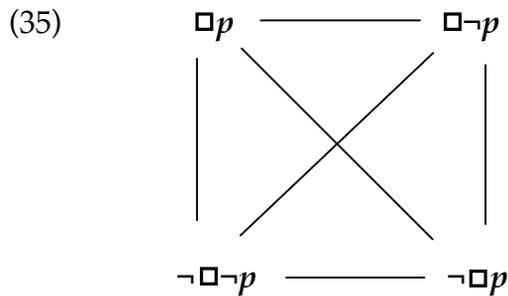
And we may leave it to the reader to show that this square of opposition can also

be represented solely in terms of possibility and negation.

The traditional view of modality differs at this point from *PTM* in that the traditional view needs no assumption of consistency to establish (35) and its variants as a *bona fide* square of opposition. As a matter of fact, it needs no assumptions whatsoever to do this. (35) and its possibility variant are direct consequences of the traditional interpretation of modality operators in terms of possible worlds.

Another interesting point of divergence between *PTM* and the traditional view of modality deals with the following statement.

(36)　　$\Box p \rightarrow p$

Traditionally interpreted, (36) says that any proposition is true (in the real world) if it is true in every possible world. Taken this way, (36) is obviously true. Yet, interpreted proof-theoretically, (36) says that any proposition is true if it is provable. This statement may be true, but only if the set of postulates from which the proposition is provable, is also *sound* (which is hardly something that can be taken for granted). In fact, (36) is the very formulation of soundness under the proof-theoretic interpretation.

Also interesting is (37), which is the converse of (36).

(37)　　$p \rightarrow \Box p$

Taken in its traditional interpretation, (37) says that any proposition is true in every possible world if it is true in the real world. This, of course, is just false. Taken proof-theoretically, however, it says that any true proposition is provable. This would be true if the set of postulates from which the proposition is provable is *complete*. Yet, as Gödel has taught us, no complete set of postulates can be found which is both consistent and strong enough to axiomatize even the truths of arithmetic. So (37) can be true, but only in the simplest of logical systems.

§5. *The proof-theoretic view of counterfactuals*

We claimed in §3 above that the proof-theoretic view of counterfactuals (henceforth *PTC*) sketched in (23) can account for the paradoxes of counterfactual reasoning—namely the failure of counterfactual transitivity, the failure of rein-forcement of the counterfactual antecedent, and the failure of counterfactual

contraposition. Indeed, while factual arguments are transitive, counterfactual arguments need not be. This is a fact that *PTC* can predict. Consider for example the following instance of counterfactual reasoning.[12]

(38)   a. *If Bush had been born in Afghanistan, then he would be a member of the Taliban.*
       b. *If Bush were a member of the Taliban, then he would have revealed American military secrets to Osama bin Laden.*
       c. *If Bush had been born in Afghanistan, then he would have revealed American military secrets to Osama bin Laden.*

Let Γ' be a set of propositions that includes the claim that Bush was born in Afghanistan. If Γ' is as close as possible to the set of actual givens, then Γ' would also have to contain the proposition that Bush did not become president of the United States (given that presidents of the United States must be born in the United States)—and, therefore, that he did not have access to any American military secrets (given that this access would come to him only through presidential privilege). If Γ' were furthermore consistent, then it would *not* be provable from it that Bush would have revealed any military secrets to Osama bin Laden. In fact, it would be provable from Γ' that he *wouldn't* have done it. In any case, (38c) would be false according to (23)—and this even if, according to (23), (38a) and (38b) were each true in their own right. Transitivity therefore fails among counterfactuals.

Second, while a factual conditional holds after its antecedent has been strengthened, a counterfactual conditional does not have to. This too is a fact that can be predicted by *PTC*. Consider for example the sentences in (39).

(39)   a. *If Bill and Mary had been at the party, they would have danced together all night long.*
       b. *If Bill and Mary had been at the party, and if Bill had broken his leg, they would have danced together all night long.*

(39a) may be true according to (23) because Bill and Mary's dancing together all night long may follow from a consistent set Γ' of propositions that *includes* the assumption that Bill and Mary were at the party (as it must by the requirement that Γ' include the antecedent of a counterfactual) but *excludes* the assumption that Bill broke his leg (as it must by the requirement that Γ' be as close as possible

---

[12] All the examples of this section are from McCawley (1993, 528-540).

to the set of actual givens). Yet, (39b) would be false according to (23), as Bill and Mary's dancing all night together does not follow from any consistent set Γ′ of propositions that *includes* the assumption that Bill broke his leg (as it must by the requirement that Γ′ include the antecedent of a counterfactual) and is also, as required, as close as possible to the actual set of givens. Thus, counterfactuals need not hold under the strengthening their antecedents.

Third, while contraposition is a valid form in standard, factual, reasoning, it is not a valid form of counterfactual argument. This too is as expected according to *PTC*. Consider for example the sentences in (40).

(40)   a. *Even if Michael Jordan had played for the Lakers, he would still have been declared the best player in the history of basketball.*
       b. *Even if Michael Jordan had not been declared the best player in the history of basketball, he would still not have played for the Lakers.*

Suppose (40a) is true. Under (23), this means that Michael Jordan's being declared the best player in the history of basketball follows from a consistent set Γ′ of propositions that includes the assumption that Michael Jordan played for the Lakers and is as close as possible to the actual set of givens. This is entirely possible, as Michael Jordan could well be declared the best player in the history of the sport regardless of the team he played in. But let Γ″ be a consistent set of propositions that includes the assumption that Michael Jordan had not been declared the best player in the history of basketball (and is as close as possible to the set of actual givens). It does not follow from Γ″ that Michael Jordan did not play for the Lakers (none of the actual players of the Lakers have in fact been so declared). By (23), (40b) is therefore false. Contraposition therefore fails in (40) as expected.

It should be pointed out that Lewis' classic account of counterfactuals in terms of possible worlds is unable to account for this failure of counterfactual contraposition (see McCawley, 1993, 534). For, according to Lewis' account, (40b) is to be interpreted in a world which (i) satisfies its antecedent and (ii) is as close as possible to the real world. But the worlds closest to the real world in which Michael Jordan was not declared the best player in the history of the sport *are* worlds in which he did not play for the Lakers. So (40b) would be true, and contraposition would hold, incorrectly, in (40).

But there are other cases in which *PTC* succeeds where possible-worlds accounts fail. Recall that possible-worlds accounts of counterfactuals succeed by interpreting counterfactuals in worlds in which their antecedents are true. But

what would happen if there *were* no such worlds? Consider for example the *necessarily* counterfactual conditionals in (41a) and (41b).

(41)   a. *If 3 were even, 4 would be odd.*
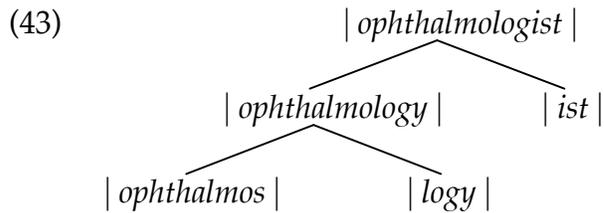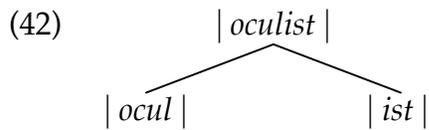   b. *If 3 were even, 4 would be prime.*

Notice that the common premise of (41a) and (41b) is indeed false in every possible world. This forces the possible-worlds approach to face the following dilemma. Either (i) there is no world in which to evaluate these counterfactuals or, (ii) both counterfactuals are true, as both are material implications from false antecedents in every world they are interpreted in. If we go for (i), then (41a) and (41b) are both uninterpretable; if we go for (ii), they are both true. In either case, the possible-worlds account has no way to distinguish between (41a) and (41b). But this is not the desired outcome. Intuitively, (41a) is true and (41b) is false.

Unlike possible-worlds accounts, *PTC* can account for the contrast in (41), as it relies on proof rather than on material implication. For, suppose Γ′ is a set of propositions that includes the claim that 3 is an even number. Suppose also that Γ′ is as close as possible to the set of actual givens. This means that Γ′ would have to include the statement that 4 is the successor of 3 as well as the statement that the successor of any even number is odd. It now follows from Γ′ that 4 would be odd. (41a) would therefore be true under (23). Yet, if Γ′ is consistent and as close as possible to set of actual givens, then (41b) would not follow from it (the successor of an even number need not be prime). (41b) would therefore be false under (23). *PTC* can therefore account, as desired, for the contrast in (41).

So *PTC* seems to be able to account for the paradoxes of counterfactual reasoning. In fact, it improves the possible-worlds accounts of counterfactuals in certain respects. But this assumes that we can define a relation of *closeness among givens* that can be justified independently of counterfactuals. This is a matter that we must leave for further research.

§6. Addressing possible objections: Overdetermination

PEST succeeded where other extensional theories failed because it provided a finer notion of meaning. It might be thought, however, that *PEST* goes overboard, and provides *too fine* a notion of meaning. Consider for example *oculist* and *ophthalmologist*. It would seem that these nouns are perfectly synonymous. Yet, if we attend to their compositional structures, these predicates should differ radically in meaning:

(42) | oculist |
       ╱   ╲
   | ocul | | ist |

(43) | ophthalmologist |
               ╱   ╲
        | ophthalmology | | ist |
          ╱   ╲
   | ophthalmos | | logy |

Thus, while both | *logy* | and | *ophthalmology* | should be part of the meaning of *ophthalmologist*, neither one of these two referents should be part of the meaning of *oculist*.

As usual, this problem is not unique to *PEST*; it affects all structural theories of meaning, including the intensional ones. One way to solve it is to adopt a *mixed* theory of interpretation in which some expressions are interpreted *directly* (or compositionally) in terms of constituents while others are interpreted *indirectly* in terms of nonconstituents. Thus *oculist* could be interpreted directly, as indicated in (42) while *ophthalmologist* could be interpreted indirectly as indicated in (44).

(44)   ‖ *ophthalmologist* ‖ = ‖ *oculist* ‖

Needless to say, it would be just as easy to interpret *ophthalmologist* directly as indicated in (43) and interpret *oculist* indirectly as indicated in (45). Which of these interpretive strategies is followed need not be left to chance. The strategy chosen could vary from speaker to speaker according to the order in which (s)he acquired these nouns. Thus, while the first of these nouns to be acquired could be interpreted directly, the second one would be interpreted indirectly in terms of the first.

(45)   ‖ *oculist* ‖ = ‖ *ophthalmologist* ‖

It should be borne in mind that expressions interpreted indirectly admit of homophones that are interpreted compositionally. Thus, even if *ophthalmologist* were interpreted indirectly in terms of *oculist*, it would still be possible for it to develop a homophone which has ophthalmology as part of its meaning. In fact, this sense may even displace the original one as the term's sense. In such a case,

oculist and *ophthalmologist* would obviously cease to be synonymous. If so, their difference in meaning would be accurately described by *PEST*—even if the two nouns were to continue to be coreferential.

Appeals to indirect interpretation are hardly unprecedented. Indirect interpretation is just *interpretation by definition*—or just *definition* pure and simple. It is therefore called for by a wide range of expressions, from *parallelogram* (which is interpreted as *quadrilateral with two pairs of parallel sides* is) to *felony* (which is defined by common laws or statutes in terms of other, presumably simpler, expressions). It is also called for by idioms, which should be interpreted indirectly as illustrated in (46).

(46) $\|$ *kick the bucket* $\|$ = $\|$ *die* $\|$
      $\|$ *throw the towel* $\|$ = $\|$ *give up* $\|$
      $\|$ *spill the beans* $\|$ = $\|$ *reveal confidential information* $\|$

But *PEST* is not out of the woods yet. Consider the sentence in (47). If *PEST* were to adopt classical theories of quantification, then it would have to assign two distinct compositional structures to this sentence. One of these structures would correspond to (48a); the other to (48b).

(47) *A man loves a woman.*
(48) a. $\exists x \exists y [man(x) \wedge woman(y) \wedge love(x,y)]$
      b. $\exists y \exists x [man(x) \wedge woman(y) \wedge love(x,y)]$

Now, to assign distinct compositional structures is to assign distinct meanings—at least in the context of a structural theory of meaning. And here is where *PEST* could run into trouble. For it is not clear what these distinct meanings would be. (48a) and (48b) are, of course, logically equivalent (see McCawley 1993, 2.5.22). And they are semantically equivalent as well, as they have exactly the same truth conditions (Barwise and Etchemendy 1990, 116).

But logical and semantic equivalence do not add up to semantic identity—at least in the context of a structural theory of meaning. Indeed, while (48a) can be taken to assert the existence of a man of certain characteristics, (48b) can be taken to assert the existence of a woman with certain characteristics. Intuitively, these are different assertions. So much so that they correspond to natural answers to different questions:

(49) A: *What can you tell me about men?*
      B: *That there is one that loves a woman* (= 48a).

(50)   A: *What can you tell me about women?*
        B: *That there is one that a man loves* (= 48b).

Having said this, notice that *PEST* still will support proofs of (51) and (52).

(51) $\exists x\exists y[man(x) \wedge woman(y) \wedge love(x,y)] \dashv\vdash \exists y\exists x[man(x) \wedge woman(y) \wedge love(x,y)]$
(52) $\exists x\exists y[man(x) \wedge woman(y) \wedge love(x,y)] \dashv\vDash \exists y\exists x[man(x) \wedge woman(y) \wedge love(x,y)]$

In so doing, it will regard (48a) and (48b) as both logically and semantically equivalent. And will do so in exactly the same ways other theories do—namely by deriving the metalogical statements in (51) and (52).

In short, *PEST* will acknowledge a semantic difference between (48a) and (48b). Yet, it will deem it subtle. Whether the "topical" difference between (48a) and (48b) we have proposed is real (and if it is, whether it is indeed the semantic difference predicted by *PEST*) is something that we will have to leave for future research. All we can hope to do here is argue that the facts in (47) and (48) need not count against *PEST*. In fact, they may actually count for it instead.

Similar replies can be given to objections from cross-linguistic claims of semantic overdetermination. To take a simple example, consider the German sentence in (53) and the Spanish counterpart thereof given in (54).

(53)   *Ich bin hungrig.*   (54)   *Yo tengo hambre.*
        I   am hungry   I have hunger
        'I am hungry.'   'I am hungry.'

(53) and (54) are, for all practical purposes, synonymous. They may serve, for example, as glosses of each other. Yet, although (53) and (54) will always convey one and the same truth value, they will do so in different ways. *PEST* thus predicts corresponding semantic differences between them. See (55) and (56) in this regard.

(55) | I am hungry | (56) | I have hunger |

        | I | | am hungry | | I | | have hunger |

             | am | | hungry | | have | | hunger |

But are (53) and (54) *perfectly* synonymous? Arguably not: while (53) asserts that

a selfsame entity has a particular property, (54) asserts that two distinct entities stand in a particular relation. But this, of course, is just what *PEST* predicts by assigning these sentences the meanings in (55) and (56).

On the other hand, it should be clear that *PEST* can also account for the "practical synonymy" between (53) and (54). And can do so in the same ways its competitors do. For note that the semantic similarity of (53) and (54) may arise either as a matter of fact (the lexical semantics of *hungrig* 'hungry' and *hambre* 'hunger' just happen to be very similar), or else as a matter of principle (via the logic postulate in (57a) or the meaning postulate in (57b)).

(57)   a. *hungry*(*i*) ⊣⊢ *have*(*i, hunger*)
       b. *hungry*(*i*) ⊣⊨ *have*(*i, hunger*)

Both the factual and the principled avenues are open to *PEST*, which can again claim that (53) and (54) are semantically distinct—even if only slightly.

Needless to say, cases like the ones discussed in this section are legion. Too many, in fact, to deal with in turn. Yet, it is our expectation that the following strategy will resolve all of these cases. First we determine whether two expressions that are alleged to be synonymous (but differ compositionally) are indeed perfectly synonymous. If they are, then one of them should be interpreted in terms of the other. If they fail to be synonymous, then the difference in compositional structure should be able to account for their semantic difference (as well as for the appearance of synonymy between them). Whether this expectation will be realized is something that future research will determine.

§7. *Addressing possible objections: Underdetermination*

Although finer than traditional versions of extensional semantics, *PEST* may still be thought to be too coarse a theory of meaning. Consider for example the nouns *human* and *person*. These two nouns have the same extension. Yet, they do not have the same meaning. While *human* refers to a set of entities that satisfy *biological* properties like being featherless, bipedal, and having opposable thumbs, *person* refers to a set of entities that satisfy certain *ethical* properties like having the right to be free from torture (and bearing the responsibility not to inflict it). Yet, if we were to discover nonhuman entities in the wild, in outer space, or in electronic circuitry, that satisfied the ethical properties of a person, then the referents of *human* and *person* would differ by virtue of the fact that they always had different meanings.

The problem that the *human*/*person* contrast poses for *PEST* is that both of these nouns are semantically simple (henceforth *monomorphemic*). Hence, it might seem that there is no semantic structure to them (or that both have the same degenerate structure, namely the tree whose only node is their common reference). Consequently, *PEST* would predict, incorrectly, that *human* and *person* are synonymous.[13]

But lacking in morphosyntactic structure is not the same as lacking in semantic structure. One could resort to the *lexical decomposition* of these nouns and claim that their meanings arise from the combination of sublexical meanings —biological properties in the case of *human* and ethical properties in the case of *person*:

(58) $|human| = \lambda x[|bipedal|(x) \wedge |featherless|(x) \wedge \ldots]$

$\lambda x[|bipedal|(x)] \quad \lambda x[|featherless|(x)] \quad \ldots$

(59) $|person| = \lambda x[|not\text{-}to\text{-}torture|(x) \wedge |not\text{-}to\text{-}be\text{-}tortured|(x) \wedge \ldots]$

$\lambda x[|not\text{-}to\text{-}torture|(x)] \quad \lambda x[|not\text{-}to\text{-}be\text{-}tortured|(x)] \quad \ldots$

Note that what we actually want to do is distinguish between the meanings of the predicates *nonfictional human* and *nonfictional person*, as there are certain fictional animals, extraterrestrials, and robots that are persons but not humans. So liberalizing our universe of discourse by incorporating fictional entities will not address the present difficulty (see §3 above).

A substantially more involved problem for *PEST* arises in connection with the interpretation of proper names. In a letter sent to P.E.B. Jourdain in 1914, Frege told the following tale:

> Let us suppose an explorer travelling in an unexplored country sees a high snow-capped mountain on the horizon. By making inquiries among the natives he learns that its name is 'Aphla'. By sighting it from different points he determines its position as exactly as possible, enters it in a map, and writes in his diary: 'Aphla is at least 5000 metres high'. Another explorer sees a snowcap-ped

---

[13] I am indebted to Jeremy Goard for pointing out the relevance of the *person*/*human* contrast to this discussion.

mountain on the southern horizon and learns that it is called 'Ateb'. He enters it in his map under this name. Later comparison shows that both explorers saw the same mountain. Now the content of the proposition 'Ateb is Aphla' is certainly not the same thing as the content of the proposition 'Ateb is Ateb' (Frege, 1980, 80).

Since *Aphla* and *Ateb* could be taken to be monomorphemic,[14] the moral we can draw from this tale is that coreferential names may be monomorphemic and still differ in meaning. If so, they would pose another problem for *PEST*. For, being monomorphemic, the meanings of these expressions would seem to be again structurally indistinguishable—and hence indistinguishable *tout court*.

Before addressing this problem, note that this difficulty burdens also the intensional theory of meaning—at least if we adopt the widespread view that proper names are *rigid designators*. For this view will force *Aphla* and *Ateb* to have the same reference in every possible world. Consequently, they would have the same sense. And, being semantically simple, they would have the same (structured) meaning as well. It should be pointed out in this regard that the problem lexical names pose for the structural theory of meaning was first observed in intensional settings (Bäuerle and Cresswell, 1989, 502f).

Two solutions to the problem raised by *Aphla* and *Ateb* can be drawn from the literature. One of these is based on the views of John Stuart Mill; the other rests on the proposals of Bertrand Russell. According to Mill (1843), proper names refer *directly* to their bearers; they do not refer to them indirectly through any meanings or senses they may be claimed to have. To a Millian, then, the problem posed by *Aphla* and *Ateb* would be solved not by assigning different senses or referents to these names, but rather by developing a theory of linguistic communication which distinguishes the semantic content of a proper name (its reference) from the information associated with a proper name by conversational participants in a particular context of utterance (Soames, 2002, 215). Armed with such a distinction, the Millian would insist that *Aphla* and *Ateb* have the same meaning (or reference); they just happen to convey different information.

Notice that this Millian response to the problem of *Aphla* and *Ateb* is available within *PEST*. It amounts to the claim that proper names are semantically simple. For if they are, then it follows from (9a) above that that the senses of *Aphla* and *Ateb* are nothing but the degenerate trees having |*Aphla*| and |*Ateb*| as their only nodes. Recalling that the meaning $\|x\|$ of an expression x is an ordered pair (N, D) in which N is a set of nodes and D is the relation of dominance in N,

---

[14] They are, after all, the mirror images of the monomorphemic words *alpha* and *beta.*

we have that

(60) $\|Aphla\| = (\{|Aphla|\}, \emptyset)$
(61) $\|Ateb\| = (\{|Ateb|\}, \emptyset)$

Since $|Aphla| = |Ateb|$, the two names in question would have the same sense. Consequently, they would have the same reference as well. As desired by the Millian, the problem posed by *Aphla* and *Ateb* would then be solved by appealing to information associated to a proper name over and above its meaning. This information would be independent of *PEST*—and hence consistent with it.

The Millian solution to the problem raised by *Aphla* and *Ateb* differs sharply from its Russellian counterpart. According to Russell (1918, IV), proper names are abbreviations for definite descriptions. Proper names are therefore synonymous to them, and refer *indirectly* through them. It follows that the Russellian has a ready-made solution to the problem posed by *Aphla* and *Ateb*. All (s)he needs to do is assume that these proper names abbreviate *different* definite descriptions which have the *same* reference. Thus, according to one resilisent version of this theory, proper names abbreviate descriptions which are *metalinguistic* in nature, so a name *N* is just short for *the bearer of the name "N"*. Since one of these descriptions involves the name *Aphla* and the other involves the name *Ateb*, the descriptions involving them would indeed be different. Yet, they would have the same reference.[15]

Like its Millian counterpart, this Russellian solution to the problem of *Aphla* and *Ateb* can be incorporated into *PEST*. All that we would need to do is interpret the names in question as indicated in (62) and (63).

(62) $\|Aphla\| = \|\text{the bearer of the name "Aphla"}\|$
(63) $\|Ateb\| = \|\text{the bearer of the name "Ateb"}\|$

In other words, all we would need to do is resort, once again, to lexical decomposition.[16]

It would take us too far afield to try and choose between the Millian and the Russellian solutions to the problem of *Aphla* and *Ateb*. For, notice that the Millian solution cannot be properly considered such until the theory of linguistic communication it requires has been detailed and evaluated. And, although cer-

---

[15] See Russell (1918, IV), Kneale (1962), Loar (1976, 1980), Katz (1977, 1990, 1994, 2001) and, most notably, Bach (1981, 1994, 2002).

[16] Or to *morphemic decomposition* if lexical items are by definition indecomposable.

tain Russellian theories of proper names have been subjected to devastating crit-
icism by Donellan, Kripke and Searle (Lycan, 2000, Part I), it is an open question
whether this criticism extends to the metalinguistic version thereof. Fortunately,
present purposes do not demand that we face these thorny issues and choose
between Millians and Russellians. As we have seen, *PEST* can seize either one of
their solutions and make it its own.

And *PEST* has options beyond the Millian and the Russellian. Suppose we
were to interpret monomorphemic names as trees consisting solely of roots and
leaves, where the leaves are the names themselves and the roots are their refer-
ents. Suppose, in other words, that the senses of *Aphla* and *Ateb* were as indicated
in (64) and (65).

(64) $\|Aphla\|$ = ({*Aphla*, |*Aphla*|}, {(*Aphla*, |*Aphla*|)})
(65) $\|Ateb\|$ = ({*Ateb*, |*Ateb*|}, {(*Ateb*, |*Ateb*|)})

These senses are the trees diagrammed in (64') and (65').

(64') |*Aphla*| (65') |*Ateb*|

      | |

    Aphla   *Ateb*

Since the senses of *Aphla* and *Ateb* are two different trees with the same roots,
what we have here are two names that differ in sense but agree in reference. This
would of course solve the problem posed by these names.

The preceding solution amounts to the claim that proper names are neces-
sary and sufficient for the construction of their reference (think of the construc-
tion of the reference of a proper name *N* from *N* as the application of a function
$\lambda x[|x|]$ to *N*). The interpretation of proper names thus contrasts with that of most
other expressions, where form is neither necessary nor sufficient for reference.
Take for example the English noun *couch*. Since its reference could have been
built from the English noun *sofa*, its form is unnecessary for the construction of
its reference. Or take the form [bank]. It may refer either to a financial institution
or to either side of a river. This form is therefore insufficient for the construction
of its reference.

The solution which we have just described differs both from the Millian
solution and from the Russellian. It differs from the Millian's in that it requires
proper names to be *necessary* for the construction of their reference. It differs

from the Russellian's in that it requires them to be *sufficient*. Whether this solution inherits the problems of its predecessors—or creates new ones of its own—are issues that must be left for future research.

§8. Addressing possible objections: Formal issues

It might be claimed that *PEST* cannot assign proper interpretations to sentences with self-referential predicates like the ones in (66).

(66) a. *Bill knows that Hillary knows it.*
      b. *John saw that Mary saw it.*
      c. *He said that she said it.*
      d. *p implies that q implies r.*
      e. *X-rays cause Y-rays to cause Z-rays.*

Let us adopt the abbreviations in (67), where the ordered pair (N, D) in (67b) is a tree with nodes N and dominance relation D (see §2).

(67) a. $p = \|\text{that Hillary knows it}\|$
      b. $(N, D) = \|\text{that Hillary knows it}\|$
      c. $N = \{|\text{that}|, |\text{Hillary}|, |\text{knows}|, |\text{it}|, \ldots\}$
      d. $q = |\text{knows that } p|$

Next let us recall that the standard set-theoretical definition of an ordered pair (x, y) is as indicated in (68).

(68) $(x, y) = \{\{x\}, \{x, y\}\}$

Finally, let us assume (69) and (70). (69) says that $|\text{knows}|$ is a function that assigns, to each true proposition, the set of entities that know it; (70) says that (66a) is true. The first of these assumptions is widely accepted; the latter is warranted for appropriate interpretations of *Bill*, *Hillary*, and *it*.

(69) $|\text{knows}| = \{(x, y): x \text{ is a true proposition} \wedge y \text{ is the set of entities that know } x\}$
(70) $|\text{Bill knows that Hillary knows it}| = \text{Truth}$

It now follows that:[17]

---

[17] I am indebted to Ede Zimmerman for the essentials of this argument.

$\|knows\| \in N$	by (67c)
$\in \{N\}$	by definition of $\{N\}$
$\in \{\{N\}, \{N,D\}\}$	by definition of $\{\{N\}, \{N,D\}\}$
$\in (N, D)$	by (68)
$= \|\text{\textit{that Hillary knows it}}\|$	by (67b)
$= p$	by (67a)
$= \{p\}$	by definition of $\{p\}$
$\in \{\{p\}, \{p, q\}\}$	by definition of $\{\{p\}, \{p, q\}\}$
$= (p, q)$	by (68)
$\in \|knows\|$	by (70), (69), (67a), (67d)

In short, it follows that $\|knows\|$ is an element of *itself*. Diagrammatically, $\|knows\|$ is part of the *membership cycle* in (71), where a clockwise edge from a to b represents the fact that $a \in b$.

(71)
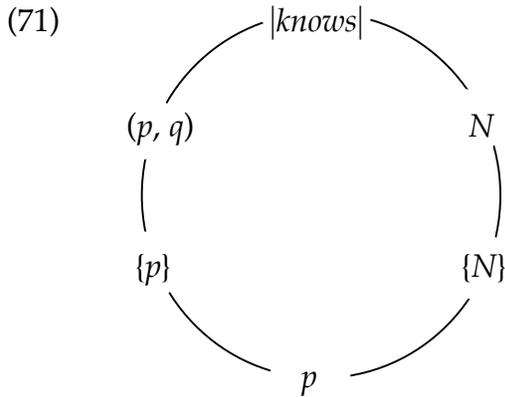
But membership cycles are inconsistent with (72), which is a standard axiom of set theory. It is the *Axiom of Foundation* (as formulated originally by von Neumann in 1925).[18]

(72)   Sets do not form infinitely descending chains of membership.

For, reading (71) counterclockwise, $\|knows\|$ initiates an infinitely descending chain of membership of the form $\ldots \in A_2 \in A_1 \in A_0$ in which $\|knows\| = A_N$ for all $N \equiv 0 \pmod 6$.

It follows that the most natural interpretation of (66a) *PEST* can provide is simply inadmissible on set-theoretical grounds. And the same holds of all the other sentences given above in (66). This is a major problem for the structural

---

[18] The Axiom of Foundation is known also as the Axiom of Regularity. See Aczel (1988, 106f), and Barwise and Moss (1996, §2.5).

theory of meaning—perhaps even the major one (Bäuerle and Cresswell 1989, 501). I hasten to add that this difficulty does not afflict only extensional theories of meaning; it affects intensional ones as well. In fact, the clash between (72) and the *iterated attitudes* in (66) was first noticed, decades ago, in a purely intensional context (see Cresswell 1975). So a simple return to possible-worlds semantics will not get us out of this thicket.

Notice that the problem we have just described arises only when the sentences in (66) are true. Could it be that these sentences never are? Not at all. It is easy to construct models which satisfy them all (once their pronouns are interpreted). For the same reason, we cannot claim that the sentences in question have nonclassical/nonexistent truth values in every model. The fact that the sentences in (66) are uninterpretable under (72) is not an *observation* that these sentences are false or uninterpretable; it is only a *prediction* that they are (and an incorrect one, at that).

One time-honored way to try to solve this problem is to claim that the propositional relations involved in (66) are ambiguous as to *logical type*, so that the propositional relations embedded in (66) belong to different (lower) types than the embedding ones. But this will not do either. For, notice that we would have to claim that the propositional relations in (66) are all *infinitely* ambiguous as to logical type (as shown in (73), there is no upper bound to the number of attitudes we could iterate in (66a)).

(73)   *Bill knows that Hillary knows it.*
         Hillary knows that Bill knows that Hillary knows it.
         Bill knows that Hillary knows that Bill knows that Hillary knows it.

Notice also that there is no independent evidence for this massive ambiguity. In fact, there is good evidence *against* it. (74) shows, for example, that a type-two instance of *knows* can be retrieved under identity with a type-zero instance thereof, so that the elliptical second conjunct of (74) amounts to *I know that he knows that she knows it*.

(74)   *He knows that she is good—and I, that he knows that she knows it.*

As is well known, such retrievals should be impossible if we had truly ambiguous instances of *knows*.

In light of this discussion, it seems that if we want to interpret the sentences in (66) within a structural theory of meaning, then we must bite the bullet

and abandon the Axiom of Foundation (henceforth **AF**). In other words, we must move from a set theory that is *well-founded* into one that is not. Or, to follow recent usage, leave set theory and head for *hyperset* theory instead (Barwise and Moss 1996). Dramatic as this move may be, it is not without precedent. It has been made by a number of researchers in order to provide a natural mathematical account of "circular phenomena" like streams in computer science, terminological cycles in artificial intelligence, and philosophical theories of truth, convention, and, indeed, self-reference. See for example Forti and Honsell (1983), Aczel (1988), and Barwise and Moss (1996).

It might be objected that **AF** is hardly something we can give up, as it is part and parcel of efforts to avoid the antinomies that rocked set theory a hundred years ago. Anyone trying to do away with **AF** must therefore show that set theory without this axiom is consistent (or at least is as consistent as set theory with this axiom). Fortunately, this has been done, as a theorem of *relative consistency* has been proven which states that set theory without **AF** is consistent if set theory with it is. So if some contradiction is derivable in the new set theory, then it is derivable in the old one as well. In fact, this relative consistency result holds even if we adopt an *anti-foundation* axiom **AF'** to the effect that membership cycles are not only possible, but necessary as well.[19]

But it might be feared that setting **AF** aside will lead to the loss of sets needed for mathematics. It turns out that this fear, too, is unfounded. The universe of sets admitted by set theory without **AF** is an *extension* of the universe of sets admitted by set theory with **AF** (Devlin 1993, 144). This means that giving up **AF** will not lead to the loss of any of the sets needed for mathematics (or the mathematical sciences). And the same holds if we were to add **AF'** to set theory. As Moschovakis (1994, 179) has pointed out, we can develop classical mathematics and all the set theory needed for it as if all mathematical objects were licensed by a set theory that gives up **AF**.

Having said this, it must be acknowledged that abandoning **AF** does come at a price. The price is giving up the *cumulative* conception of sets, which is the notion that every set is built from the simplest of sets through a succession of well-defined operations. As a matter of fact, there is a precise sense in which **AF** is nothing more and nothing less than the cumulative (or *iterative*) conception of sets (Devlin 1993, 44). But the cumulative conception of sets is so appealing that it has become dominant in modern set theory (Barwise and Moss 1996, 302).

---

[19] This is not the place to prove this relative consistency result; all we can do here is appeal to it and refer the interested reader to Barwise and Moss (1996, Chapter 9) for its proof.

Moreover, it provides a satisfying explanation for why the infamous class of all sets is "too large to be a set" (and hence for why the Russell Paradox does not arise in set theory). The explanation is that this notorious class would have to be built at the end of the set-building process. But there can be no end to this process, as there is no upper limit to the number of times the set-building operations may apply to their own outputs. This is, indeed, a satisfying explanation. So much so that once we've lost it in hyperset theory, "it is hard not to feel that the set-theoretic paradoxes have only been avoided, not really understood" (Barwise and Moss 1996, 320).

PEST is forced to pay this price, as it must provide an account of the self-reference in (66), and giving up **AF** seems to be the only way to go. It should be added that *PEST* is not alone in this. As far as I can see, all structural theories of meaning in the literature, including possible-worlds semantics, are in the same boat as *PEST*. It should also be added that non-wellfounded attempts to find intuitive conceptions of sets and satisfying constraints on their size are well underway as well (see for example Barwise and Moss 1996, Chapter 20).

§9. Conclusion: Advantages of PEST

It seems that *PEST* can overcome the difficulties that have hitherto plagued the extensional theory of meaning. As we have seen in the course of this paper, *PEST* can account for the variable informativeness of identity statements, the failure of substitution in opaque contexts, the compositional interpretation of modal verbs and adverbs, the non-trivial nature of counterfactuals, and the nonsynonymy of vacuous predicates.

In fact, there are instances in which the accounts of these facts provided by *PEST* are actually better than the ones provided by possible-worlds semantics. For one thing, *PEST* accounts are by and large simpler, less abstract, and more intuitive than those issuing from the intensional account. In addition, only the *PEST* account of the nonsynonymy of vacuous predicates extends to predicates that are necessarily empty (or empty on logical grounds), only the *PEST* account of counterfactual conditionals extends to conditionals that are necessarily counterfactual (or counterfactual on logical grounds), and only the *PEST* account predicts the failure of counterfactual contraposition.

But *PEST* is *intrinsically* simpler, less abstract, and more intuitive than possible-worlds semantics. While possible-worlds semantics requires a nondenumerable set of nondenumerably complex entities (possible worlds), *PEST* calls

only for a denumerable set of finitely complex entities (referential structures). While possible-worlds semantics construes meaning as potential reference, *PEST* construes it only as actual reference—or rather as the *construction* of actual reference. This construction is, of course, invoked by all compositional frameworks. All *PEST* does is keep track of it. As to intuitiveness, note that only *PEST*, not possible-worlds semantics, formalizes the disarming Fregean intuition that the sense of an expression is the way in which that expression presents its reference.

Possible-worlds semantics invokes structural theories of meaning to solve the problem of necessarily coextensional expressions. In doing so, it might have gone overboard. All it needed were structures that described the construction of reference, not the construction of meaning. Meaning is in the construction.

Appendix 1: An Implementation of PEST

There are many ways to implement *PEST* in a grammar. Here we present one. We do this to make *PEST* concrete and to show that this theory of meaning can be implemented in a natural way.

The implementation we envisage is a grammar that incorporates a *rule-by-rule* approach to semantic interpretation. This means that the phrasal component of the grammar will consist of two paired sets of rules. One of these sets will build forms (syntactic trees); the other will build contents (semantic trees). To underscore the strict parallelism that holds between the syntactic and the semantic rules, we will write them side by side, with syntactic rules on the left and semantic rules on the right:

1. $S \rightarrow ADV_1 \quad S$ 1. $|S| = |ADV_1|(|S|)$

2. $S \rightarrow ADV_2 \quad S$ 2. $|S| = |ADV_2|(\|S\|)$

3. $S \rightarrow NP \quad VP$ 3. $|S| = |VP|(|NP|)$

4. $VP \rightarrow V_1 \quad NP$ 4. $|VP| = |V_1|(|NP|)$

5. $VP \rightarrow V_2 \quad S$ 5. $|VP| = |V_2|(\|S\|)$

6. $NP \rightarrow Aphla$ 6. $|NP| = |Aphla|$

7. $NP \rightarrow Ateb$ 7. $|NP| = |Ateb|$

8. $NP \rightarrow Ptolemy$ 8. $|NP| = |Ptolemy|$

9. $V_1 \rightarrow is$

10. $V_2 \rightarrow believes$

11. $ADV_1 \rightarrow$ *it-is-true-that*

12. $ADV_1 \rightarrow$ *it-is-false-that*

13. $ADV_2 \rightarrow$ *it-is-necessary-that*

14. $ADV_2 \rightarrow$ *it-is-possible-that*

9. $|V_1| = |is|$

10. $|V_2| = |believes|$

11. $|ADV_1| = |it\text{-}is\text{-}true\text{-}that|$

12. $|ADV_1| = |it\text{-}is\text{-}false\text{-}that|$

13. $|ADV_2| = |it\text{-}is\text{-}necessary\text{-}that|$

14. $|ADV_2| = |it\text{-}is\text{-}possible\text{-}that|$

The rules on the left are context-free rules.[20] They govern the way in which forms combine to yield forms. We may therefore call them rules of *syntactic composition*. The rules on the right indicate the way in which meanings combine to yield meanings. We may therefore call them rules of *semantic composition*.

Rules of semantic composition describe the way in which the reference of a complex expression is built from the senses of its constituents. Yet, as pointed out in §2 above, the reference of a complex expression will not usually need full access to the senses of its consitituents; only to the referents these senses determine. To keep formalization to a minimum, our rules of semantic composition mention those referents directly (and not indirectly as a function of the senses that determine them). There are some cases, however, where the reference of a complex expression will depend crucially on the full senses of its parts. Unsurprisingly, these are the cases that involve modality operators and propositional attitudes. See rules (2) and (5) of semantic composition, where mention of constituent senses is unavoidable (and, incidentally, straightforward).

The phrasal component above is to be supplemented by the lexical component below.

15. *Aphla*

16. *Ateb*

17. *believes*

18. *is*

19. *it-is-false-that*

15. *Aphla*

16. *Ateb*

17. $|believes|$

18. $|is|$

19. $|it\text{-}is\text{-}false\text{-}that|$

---

[20] We ask the reader to pretend with us that *it is true that*, *it is false that*, etc. are adverbs. No harm will come from these useful fictions, as they can be replaced by more accurate analyses once our goals turn from exposition to description.

20. *it-is-necessary-that*   20. *|it-is-necessary-that|*

21. *it-is-possible-that*   21. *|it-is-possible-that|*

22. *it-is-true-that*   22. *|it-is-true-that|*

23. *Ptolemy*   23. *Ptolemy*

With one important class of exceptions, the lexical component of our grammar is a pairing of simple, logically primitive, forms, with simple, logically primitive, referents. The class of exceptions consists entirely of proper names, which are not paired with their referents, but only with themselves.

Notice that this does not mean that proper names like (15), (16), (23) are lexically self-referential. To be lexically self-referential, proper names would have to be paired to their referents in the lexical component, and we have just issued a disclaimer to this effect. True, we still need to say what the reference of a proper name is. But, as we shall soon see, this will be done not by the grammar of English, but rather by a *model* for the interpretation of English (so knowledge of the reference of a proper name is not part of knowledge of English). It should be pointed out that this is not the only purpose of models; models will also identify the referents explicitly mentioned in the lexical component.

Why we should want to analyze proper names in this roundabout way will become clear in a moment. In the meantime, let me claim that the grammar we have just developed will generate *paired trees*. The leaves of these trees will be licensed by the lexical component of the grammar; their nonleaves will be admitted by its phrasal component. The generated trees are *paired* in the sense that they are generated by paired lexical items and, if necessary, by paired rules as well.

To illustrate, notice that the following are two paired trees generated by our grammar (as an aid to the reader, I will henceforth annotate tree nodes with the numbers of the rules or lexical items that license them).

$$NP_{[6]}$$
$$|$$
$$Aphla_{[15]}$$

$$|Aphla|_{[6]}$$
$$|$$
$$Aphla_{[15]}$$

And the following are also paired trees generated by the grammar.

$$NP_{[7]}$$
$$|$$
$$Ateb_{[16]}$$

$$|Ateb|_{[7]}$$
$$|$$
$$Ateb_{[16]}$$
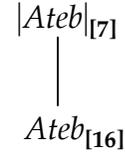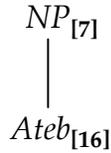
But suppose now that we were to adopt *PEST*. It would follow that the trees to the right would be the *senses* of *Aphla* and *Ateb* (see (9) in §2). And then a glance at (52′) and (53′) above would reveal that the grammar we have developed incorporates the interpretation of proper names presented at the end of §6. It was indeed to incorporate this account that we adopted the lexical entries in (15), (16), (23) above.

So our grammar can distinguish the senses of *Aphla* and *Ateb* (one makes crucial reference to the name *Aphla* while the other, to the name *Ateb*). And we can distinguish between these names even if they had the same reference. For, consider a model for the interpretation of English which stipulates that $|Aphla|$ = $|Ateb|$ = Kilimanjaro. Relative to this model, the senses of the names *Aphla* and *Ateb* would be as follows.

Kilimanjaro
$|$
Aphla

Kilimanjaro
$|$
Ateb
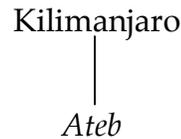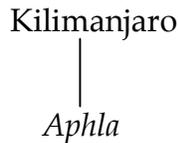
Notice that the semantic difference between *Aphla* and *Ateb* extends to a distinction between the propositions expressed by *Aphla is Aphla* and *Aphla is Ateb*. For note that our grammar will assign the following propositions to these sentences.

$$|is|(|Aphla|)(|Aphla|)_{[3]}$$
$$|Aphla|_{[6]} \qquad |is|(|Aphla|)_{[4]}$$
$$|is|_{[9]} \qquad |Aphla|_{[6]}$$
$$Aphla_{[15]} \quad |is|_{[18]} \qquad Aphla_{[15]}$$

$$|is|(|Ateb|)(|Aphla|)_{[3]}$$
$$|Aphla|_{[6]} \qquad |is|(|Ateb|)_{[4]}$$
$$|is|_{[9]} \qquad |Ateb|_{[7]}$$
$$Aphla_{[15]} \quad |is|_{[18]} \qquad Ateb_{[16]}$$
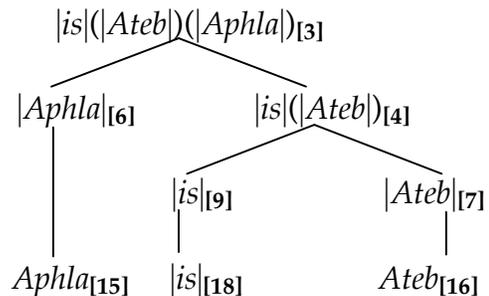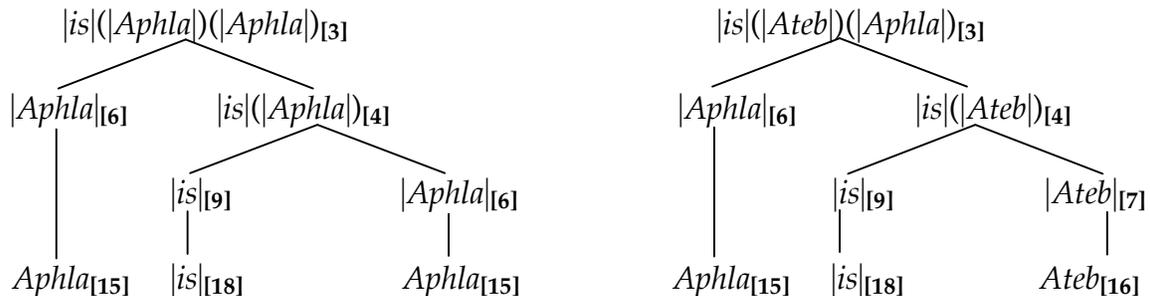
These are different trees. They therefore correspond to different senses. And these senses may even determine different referents. Consider for example a model in which $|is|$ is the function defined as follows (*T* and *F* stand here, of course, for truth and falsity).

$$
|is| = \begin{bmatrix} \text{Kilimanjaro} \rightarrow \begin{bmatrix} \text{Kilimanjaro} \rightarrow \text{T} \\ \text{Ptolomaeus} \rightarrow \text{F} \end{bmatrix} \\ \text{Ptolomaeus} \rightarrow \begin{bmatrix} \text{Kilimanjaro} \rightarrow \text{F} \\ \text{Ptolomaeus} \rightarrow \text{T} \end{bmatrix} \end{bmatrix}
$$

Both of our sentences would be true relative to this model. Yet, the propositions they express would be different. *Aphla is Aphla* and *Aphla is Ateb* may therefore differ in sense when they do not differ in reference.

But our grammar does not only distinguish between these propositions. It can also *use* them as arguments to propositional attitudes like *believe*. This yields properties like |*believes Aphla is Aphla*| and |*believes Aphla is Ateb*|:

|*believes*|(‖*Aphla is Aphla*‖)[5]

|*believes*|[10]   ‖*Aphla is Aphla*‖[3,4,6,9,15,18]

|*believes*|[17]

|*believes*|(‖*Aphla is Ateb*‖)[5]

|*believes*|[10]   ‖*Aphla is Ateb*‖[3,4,6,7,9,15,16,18]
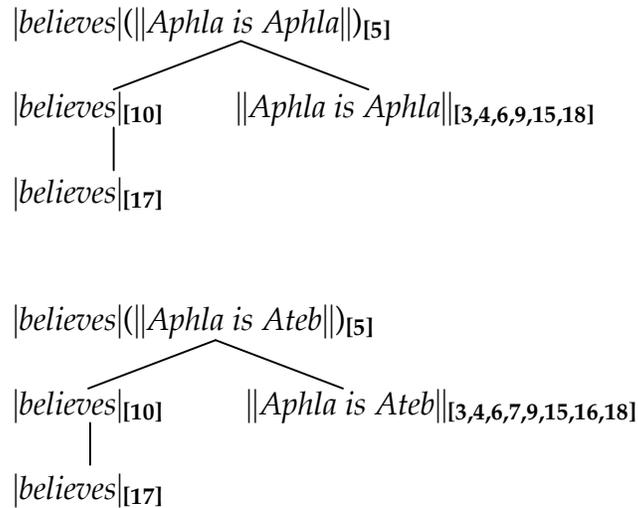
|*believes*|[17]

Since these are different trees, the properties they correspond to will be different as well. And these properties may even differ in extension. Consider for example a model in which *believes* refers to the following function.

$$|believes| = \begin{bmatrix} ||Aphla\ is\ Aphla|| & \rightarrow & \begin{bmatrix} Ptolomaeus & \rightarrow & T \\ \\ KIlimanjaro & \rightarrow & F \end{bmatrix} \\ \\ ||Aphla\ is\ Ateb|| & \rightarrow & \begin{bmatrix} Ptolomaeus & \rightarrow & F \\ \\ Kilimanjaro & \rightarrow & F \end{bmatrix} \end{bmatrix}$$

Suppose further that $|Ptolemy|$ = Ptolomaeus in this model. Relative to it, the first of the following two sentences will be true, while the second one will be false.

$|believes|(||Aphla\ is\ Aphla||)(|Ptolemy|)_{[3]}$

$|Ptolemy|_{[8]}$ $\quad$ $|believes|(||Aphla\ is\ Aphla||)_{[5]}$

$|believes|_{[10]}$ $\quad$ $||Aphla\ is\ Aphla||_{[3,4,6,9,15,18]}$

$Ptolemy_{[23]}$ $\quad$ $|believes|_{[17]}$

$|believes|(||Aphla\ is\ Ateb||)(|Ptolemy|)_{[3]}$

$|Ptolemy|_{[8]}$ $\quad$ $|believes|(||Aphla\ is\ Ateb||)_{[5]}$

$|believes|_{[10]}$ $\quad$ $||Aphla\ is\ Ateb||_{[3,4,6,7,9,15,16,18]}$

$Ptolemy_{[23]}$ $\quad$ $|believes|_{[17]}$
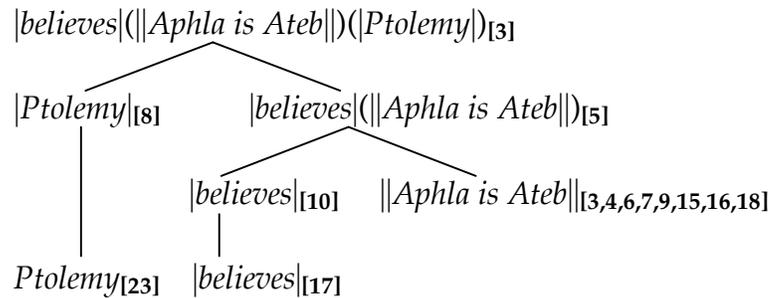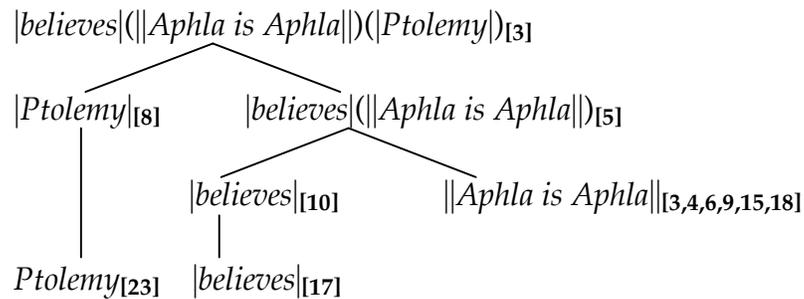
So Ptolomaeus may believe that Aphla is Aphla without believing that Aphla is Ateb. Our properties may therefore differ in extension as well as in "intension".

Finally, notice that our grammar can account for the fact that some expressions are sensitive to a sense while others are sensitive to the reference it determines. Take for example $|it\text{-}is\text{-}true\text{-}that|$. It is a function of truth values. Consequently, it only requires access to the reference of a sentence, not to its sense. The grammar above will provide this function with the argument it needs as follows.

$|it\text{-}is\text{-}true\text{-}that|(|is|(|Ateb|)(|Aphla|))_{[1]}$

$|it\text{-}is\text{-}true\text{-}that|_{[11]}$ $|is|(|Ateb|)(|Aphla|)_{[3]}$

$|Aphla|_{[6]}$ $|is|(|Ateb|)_{[4]}$

$|is|_{[9]}$ $|Ateb|_{[7]}$

$|it\text{-}is\text{-}true\text{-}that|_{[22]}$ $Aphla_{[15]}$ $|is|_{[18]}$ $Ateb_{[16]}$

Compare this with *|it-is-necessary-that|*. This is a function of propositions. Consequently, our grammar requires it to combine with the sense of a sentence, not just with its reference. Our grammar will provide it with the richer argument it needs as indicated below.

$|it\text{-}is\text{-}necessary\text{-}that|(\|Aphla\ is\ Ateb\|)_{[2]}$

$|it\text{-}is\text{-}necessary\text{-}that|_{[13]}$ $\|Aphla\ is\ Ateb\|_{[3,4,6,7,9,15,16,18]}$

$|it\text{-}is\text{-}necessary\text{-}that|_{[20]}$

Note, by the way, that the first of these sentences will be true and the latter false whenever they are evaluated in a model in which:

$$|it\text{-}is\text{-}true\text{-}that| = \begin{bmatrix} T \to T \\ F \to F \end{bmatrix} \qquad |it\text{-}is\text{-}necessary| = \begin{bmatrix} \|Aphla\ is\ Aphla\| \to T \\ \|Aphla\ is\ Ateb\| \to F \end{bmatrix}$$

This means that the effects of these operators on their arguments can differ to the point of producing a difference in reference or truth value. And the same points can be made with *It is false that Aphla is Ateb* and *It is possible that Aphla is Ateb* relative to models in which:

$$|it\text{-}is\text{-}false\text{-}that| = \begin{bmatrix} T \to F \\ F \to T \end{bmatrix} \qquad |it\text{-}is\text{-}possible| = \begin{bmatrix} \|Aphla\ is\ Aphla\| \to T \\ \|Aphla\ is\ Ateb\| \to T \end{bmatrix}$$

Appendix 2: PEST and Entailment

Any semantic theory that aspires to descriptive adequacy must account for intuitions of entailment. This is especially true for *PEST*, where modality is, in effect, an assertion of entailment. The purpose of this appendix is to show that *PEST* can account for intuitions of entailment—at least in principle. To do so we will describe some of the interpretations *PEST* can assign to first-order sentences. Then we will show that these interpretations are rich enough to decide issues of entailment among them.

Suppose *PEST* were to assign the tree in (1) to the existentially quantified sentence *Something is good*.

(1) $\exists x[|good|(|x|)]$
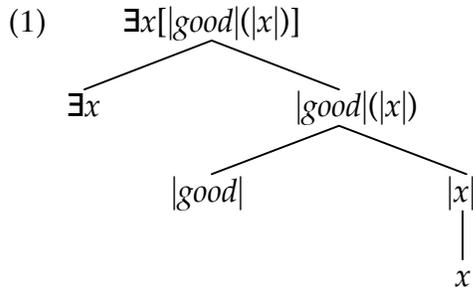
Note that (1) contains both a variable x and its referent $|x|$. Following standard model-theoretic practice, $|x|$ is the image of x under a randomly chosen function g from the set of variables to the set of entities of the model. Note also that we take $|good|$ to be, unsurprisingly, a function from the set of entities of the model to the set of truth values. $|good|(|x|)$ is, therefore, a truth value, and the subtree it roots is, by definition, a proposition. Moving on to $\exists x$, we shall take it to be a function that maps propositions of *PEST* into truth values. This function is defined as follows, where p is a proposition of *PEST*.

> $\exists x(p)$ = Truth if and only if there exists a g' such that the root of p' is Truth, where p' is the result of replacing every instance of $g(x)$ in p by $g'(x)$, and where x is a variable that is free in p.

(following common model-theoretic practice, we let g' stand for a function from the set of variables to the set of entities of the model—a function that may only differ from g in the value it assigns to x).

It follows that $\exists x[|good|(|x|)]$ is a truth value. Consequently (1) is a proposition.

It should not escape the reader that the variable in (1) plays an essential therein.

Indeed, if it weren't for this variable, then one would not be able to identify what variables or positions (if any) are bound by ∃x. All we would have would be an entity and a function, and we wouldn't know what variable contributed this entity. Indeed, we wouldn't even be able to tell whether this variable was contributed by a variable or a constant.

It should also be noted that ∃x is a function that applies to senses (trees) rather than to referents (roots). In fact, it must apply to propositions rather than to truth values. ∃x is therefore like the propositional attitude verb *believe* or the modal operators *it is necessary that* and *it is possible that* of Appendix 1.

Note finally that *PEST* in effect regards a variable x as an expression that has both a sense $\|x\|$ and a reference $|x|$, where

$$\|x\| = \begin{array}{c} |x| \\ | \\ x \end{array}$$

Variables can therefore be interpreted in the same way as proper names under the metalinguistic Russellian analysis presented above. Under this analysis, every term, be it constant or variable, would therefore be interpreted as a two-node tree whose leaf is the term itself and whose root is its reference.

Suppose next that *PEST* were to assign the proposition in (2) to the universally quantified sentence *Everything is good*.

(2) ∀x[|good|(|x|)]
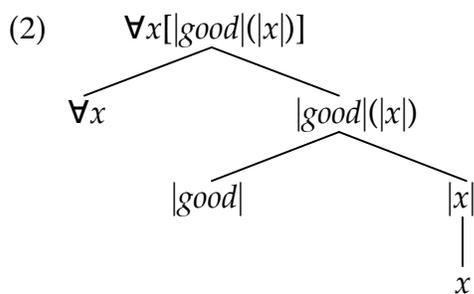
Everything in (2) would be the same as in (1), except that ∀x would be a function from propositions to truth values defined as follows, where g' is interpreted as above and p is a proposition of *PEST*.

∀x(p) = Truth if and only if every g' is such that the root of p' is Truth, where p' is the result of replacing every instance of $g(x)$ in p by $g'(x)$, and where x is a variable that is free in p.
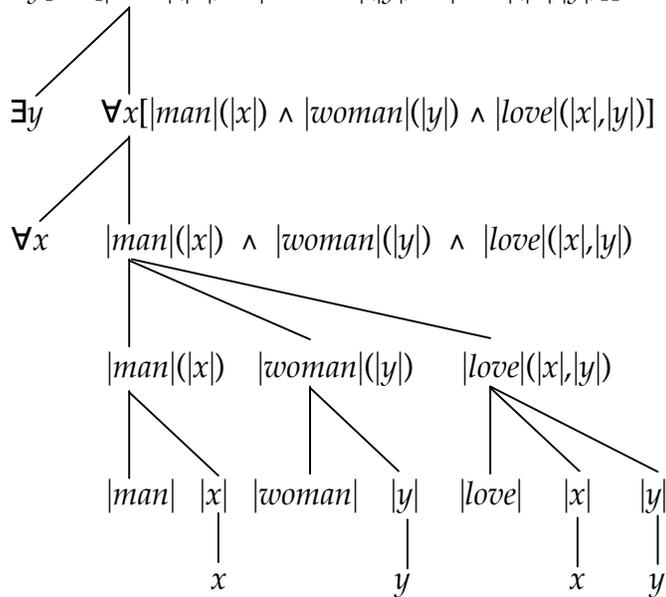
Once again, $\forall x[|good|(|x|)]$ would be a truth value, and the tree in (2) would be a proposition.

And *PEST* can afford the propositions in (3) and (4) for the two senses of *Every man loves a woman*.

(3) $\forall x[\exists y[|man|(|x|) \wedge |woman|(|y|) \wedge |love|(|x|,|y|)]]$

$\quad\quad \forall x \quad \exists y[|man|(|x|) \wedge |woman|(|y|) \wedge |love|(|x|,|y|)]$

$\quad\quad\quad \exists y \quad |man|(|x|) \wedge |woman|(|y|) \wedge |love|(|x|,|y|)$

$\quad\quad\quad\quad |man|(|x|) \quad |woman|(|y|) \quad |love|(|x|,|y|)$

$\quad\quad\quad\quad |man| \quad |x| \quad |woman| \quad |y| \quad |love| \quad |x| \quad |y|$

$\quad\quad\quad\quad\quad\quad x \quad\quad\quad\quad y \quad\quad\quad\quad x \quad y$

(4) $\exists y[\forall x[|man|(|x|) \wedge |woman|(|y|) \wedge |love|(|x|,|y|)]]$

$\quad\quad \exists y \quad \forall x[|man|(|x|) \wedge |woman|(|y|) \wedge |love|(|x|,|y|)]$

$\quad\quad\quad \forall x \quad |man|(|x|) \wedge |woman|(|y|) \wedge |love|(|x|,|y|)$

$\quad\quad\quad\quad |man|(|x|) \quad |woman|(|y|) \quad |love|(|x|,|y|)$
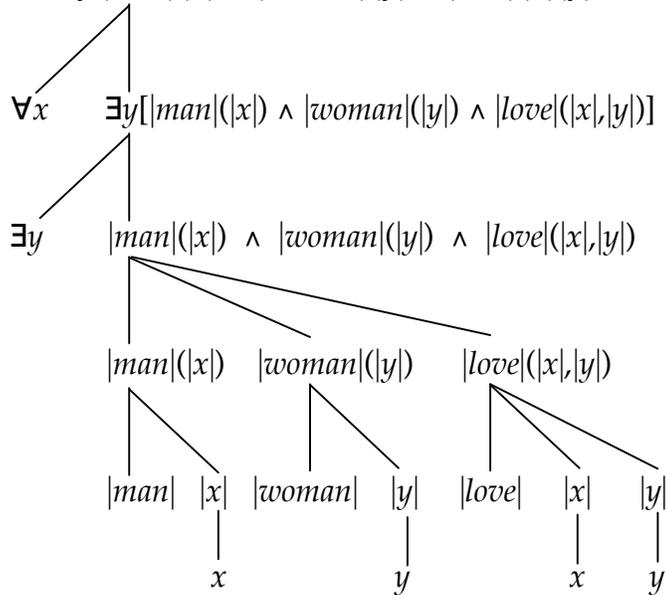
$\quad\quad\quad\quad |man| \quad |x| \quad |woman| \quad |y| \quad |love| \quad |x| \quad |y|$

$\quad\quad\quad\quad\quad\quad x \quad\quad\quad\quad y \quad\quad\quad\quad x \quad y$

It should be clear from the discussion so far that *PEST* will be able to assign propositions to first-order sentences of arbitrary complexity. It should also be

clear that these propositions and the first-order formulas we have used to label their roots can be placed in a one-to-one correspondence. But these first-order formulas are precisely what entailments pertain to, be they the semantic entailments supported by model theories or the deductive entailments sanctioned by proof theories. Thus, any claim of entailment between these formulas can be translated into a claim of entailment between *PEST* propositions. And, conversely, if these are all the claims of entailment we afford *PEST*, then every claim of entailment between *PEST* propositions can be translated into a claim of entailment between first-order formulas. *PEST* can therefore account for entailments—at least in principle.

References

Aczel, Peter (1988) *Non-Well-Founded Sets* [= *CSLI Lecture Notes, 14*]. Stanford, Center for the Study of Language and Information, Stanford University.

Bach, Kent (1981) "What's in a name?" *Australasian Journal of Philosophy* 59: 371–386.

Bach, Kent (1994) *Thought and Reference.* Revised with postscript. Oxford University Press.

Bach, Kent (2002) "Giorgione was so-called because of his name". *Philosophical Perspectives* 16: 73-103.

Barwise, Jon and Etchemandy, John (1990) *The Language of First Order Logic* [*CSLI Lecture Notes, 23*]. Stanford, Center for the Study of Language and Information, Stanford University.

Barwise, Jon and Moss, Lawrence (1996) *Vicious Circles: On the Mathematics of Non-Wellfounded Phenomena* [= *CSLI Lecture Notes, 60*]. Stanford, Center for the Study of Language and Information, Stanford University.

Bäuerle, Rainer, and Cresswell, Max (1989) "Propositional Attitudes". In Gabbay, Dov and Guenthner, Franz (Eds.) *Handbook of Philosophical Logic. Volume IV: Topics in the Philosophy of Language* [= *Synthese Library, 167*]. Dordrecht, D. Reidel Publishing Company.

Boolos, George (1993) *The Logic of Provability*. Cambridge, Cambridge University Press.

Carnap, Rudolf (1947) *Meaning and Necessity.* Chicago, University of Chicago Press.

Cresswell, Max (1975) "Hyperintensional logic". *Studia Logica* 34: 25-38.

Cresswell, Max (1985) *Stuctured Meanings.* Cambridge, Cambridge University Press.

Devlin, Keith (1993) *The Joy of Sets. Fundamentals of Contemporary Set Theory.* Second Edition. New York, Springer-Verlag.

Dowty, David, et al. (1981) *Introduction to Montague Semantics.* Dordrecht, Reidel.

Dunn, Michael (1973) "A truth value semantics for modal logic". In Leblanc, Hughes (Ed.) *Truth, Syntax, and Modality. Proceedings of the Temple University Conference on Alternative Semantics* [= *Studies in Logic and the Foundations of Mathematics*, 68]. Amsterdam, North-Holland Publishing Company.

Forti, Marco and Honsell, Furio (1983) "Set theory with free construction principles". *Annali Scuola Normale Superiore di Pisa, Classe di Scienze* 10: 493-522.

Forbes, Graeme (1989) *Languages of Possibility. An Essay in Philosophical Logic* [= *Aristotelian Society Series, Volume 9*]. Oxford, Basil Blackwell Ltd.

Frege, Gottlob (1879) *Begriffsschrift*. Halle. Translated in Jan van Heijenoort (Ed.) *From Frege to Gödel. Two Fundamental Texts in Mathematical Logic.* Cambridge, Harvard University Press, 1967.

Frege, Gottlob (1892) "Über Sinn und Bedeutung." *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50. Translated by Max Black as "On sense and reference," in Black, Max and Peter Geach (Eds.) *Translations from the Philosophical Writings of Gottlob Frege*. 1960, Oxford, Basil Blackwell, pp. 56-78.

Frege, Gottlob (1980) *Philosophical and Mathematical Correspondence*. Edited by Gottfried Gabriel et al., abridged for the English edition by Brian McGuiness, and translated by Hans Kaal. Chicago, University of Chicago Press.

Garson, James (1993) "Modal Logic". In Zalta, Edward (Ed.) *The Stanford Encyclopedia of Philosophy. Winter 2003 Edition*. <http://plato.stanford.edu/archives/win2003/entries/logic-modal/>.

Goodman, Nelson (1947) "The problem of counterfactual conditionals". *Journal of Philosophy* 44: 113-128.

Hacking, Ian (1994) "What is Logic?". In Gabbay, Dov (Ed.) *What is a Logical System?* [= *Studies in Logic and Computation, 4*]. Oxford, Clarendon Press.

Hamlyn, D. (1967) "Contingent and necessary statements". In Edwards, Paul (Ed.) *The Encyclopedia of Philosophy*. Collier-Macmillan. New York and London.

Hintikka, Jaakko (1961) "Modality and quantification". *Theoria* 27: 110-128.

Hughes, George and Cresswell, Max (1996) *A New Introduction to Modal Logic*. London, Routledge.

Katz, Jerrold (1977) "A proper theory of names". *Philosophical Studies* 31: 1-80.

Katz, Jerrold (1990) "Has the description theory of proper names been refuted?" In George Boolos (Ed.) *Meaning and Method: Essays in Honor of Hilary Putnam*. Cambridge, Cambridge University Press.

Katz, Jerrold (1994) "Names without bearers". *Philosophical Review* 103: 1–39.

Katz, Jerrold (2001) "The end of Millianism". *Journal of Philosophy* 98: 137-166.

Kearns, John (1981) "Modal semantics without possible worlds". *Journal of Symbolic Logic* 46: 77-86.

King, Jeffrey (1995) "Structured propositions and complex predicates". *Noûs* 29: 516-535.

Kneale, William (1962) "Modality, de dicto and de re". In Nagel, Ernest et al. (Eds.) *Logic, Methodology, and the Philosophy of Science: Proceedings of the 1960 International Conference*. Stanford, Stanford University Press.

Kneale, William and Kneale, Martha (1962) *The Development of Logic*. London, Oxford University Press.

Kripke, Saul (1959) "A completeness theorem in modal logic". *Journal of Symbolic Logic* 24: 1-14.

Kripke, Saul (1963a) "Semantical considerations on modal logic". *Acta Philosophica Fennica. Modal and Many-Valued Logics*, pp. 83-94.

Kripke, Saul (1963b) "Semantical analysis of modal logic I: Normal modal propositional calculi". *Zeitschrift für mathematische Logik und Grundlagen der*

Mathematik 9: 67-96.

Larson, Richard & Ludlow, Peter (1993) "Interpreted Logical Form". *Synthese* 95: 305-355.

Lewis, Clarence (1944) "Modes of meaning". *Philosophical and Phenomenological Research* 4: 236-250.

Lewis, Clarence and C. Langford (1946) *Symbolic Logic*. New York, Dover.

Lewis, David (1972) "General Semantics". In Davison, Donald and Gilbert Harman (Eds.) *Semantics of Natural Language.* Dordrecht, D. Reidel Publishing Company, pp. 169-218.

Lewis, David (1973) *Counterfactuals*. Cambridge, Harvard University Press.

Loar, Brian (1976) "The semantics of singular names". *Philosophical Studies* 30: 353-377.

Loar, Brian (1980) "Names and descriptions: a reply to Michael Devitt". *Philosophical Studies* 38: 85-89.

McCawley, James (1993) *Everything Linguists Have Always Wanted to Know About Logic.*[*] Chicago, University of Chicago Press.

Mill, John Stuart (1843) *A System of Logic.* London.

Montague, Richard (1974) *Formal Philosophy. Selected Papers of Richard Montague.* Edited and with an introduction by Richmond H. Thomason. New Haven, Yale University Press.

Moschovakis, Yiannis (1993) "Sense and denotation as algorithm and value". In Oikkonen, J. and J. Väänänen (Eds.) *Logic Colloquium '90* [= *Lecture Notes in Logic*, 2]. Berlin, Springer-Verlag.

Moschovakis, Yiannis (1994) *Notes on Set Theory*. New York, Springer-Verlag.

Partee, Barbara et al. (1990) *Mathematical Methods in Linguistics.* Dordrecht, Kluwer Academic Publishers

Prior, A (1967) "Logic, Modal". In Edwards, Paul (Ed.) *The Encyclopedia of Philosophy*. Collier-Macmillan. New York and London.

Rescher, Nicholas (1964) *Hypothetical Reasoning*. Amsterdam, North Holland.

Richard, Mark (1990) *Propositional Attitudes. An Essay on Thoughts and How we Ascribe Them.* Cambridge, Cambridge University Press.

Russell, Bertrand (1905) "On denoting". *Mind* 14: 479-93.

Russell, Bertrand (1918) "The philosophy of logical atomism". In Russell, Bertrand, *Logic and Knowledge.* London, Allen and Unwin, 1956.

Soames, Scott (1987) "Direct reference, propositional attitudes, and semantic content". *Philosophical Topics* 15: 47-87.

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

Sowa, John (2002) "Laws, facts and contexts: Foundations for multimodal

---

[*] *but were ashamed to ask.*

reasoning". Paper presented at the conference on *Dimensions of Epistemic Logic.* Danish Network for Philosophical Logic and its Applications. Roskilde University, May 2-4, 2002.

Version: 10/30/05