How to Think About Nothing

Introduction: A certain direction in cognitive science has been to try to "ground" public language statements in some species of mental representation. A central tenet of this trend is that communication – that is, public language – succeeds (when it does) because the elements of this public language are in some way correlated with mental items of both the speaker and the audience so that the mental state evoked in the audience by the use of that piece of public language is the one that the speaker wanted to evoke. The "meaning", therefore, of an utterance – and of the parts of an utterance, such as individual sentences and their parts, the individual words, etc. – is, in this view, some mental item.

Such a view of meaning is variously called "internalist" or "cognitive" or "subjectivist" or "solipsistic" or (sometimes) "representationalist" (these terms having, however, further connotations which set them apart from one another in other ways), and can be found in a wide variety of writers who do not agree on many other things. It is opposed to views that take the meaning of an utterance to be an item of "reality," however defined. In different writers this latter view is called "externalist" or "objectivist" or "realist," always with the idea that there is something other (or at least, more) than the mental state of speakers and hearers that determines meaning. The literature is rife with arguments between the two on such topics as "truth" and "synonymy" and "twin earth" and "arthritis" (to mention only a few). I do not plan to engage this literature here. Instead I wish to point to a different problem: a problem concerning the type of internal, mental entities that are invoked as meanings for the public language, a problem that seems to infect a wide range of, perhaps all, internalist theories.

There are different types of cognitivist theories of meaning, of course, and when trying to describe the underlying problem I will attempt to remain neutral as to what the mental items are that imbue public terms with meaning. If I sometimes slip into a "mental language of thought" idiom, or a "mental images" idiom, or a "mental models" idiom, etc., attempt to read the point being made in the broadest way possible. Although I think that none of the cognitivist theories of meaning have in fact adequately addressed the problem to be introduced, I do think that some of the theories

could mount reasonable answers to this issue, and so I will try to show how some cognitivist theories could provide an answer to the problem.

How Cognitivism Deals with Problems in Naïve Semantics: Naïve semantics is the view that expressions of natural language get their meaning from the aspect of reality that they name or describe. A sentence, for example, might describe an event, or an action, or a state of affairs. The described entity is what counts as the meaning of this sentence, and the sentence is true if and only if the event, action, state of affairs actually obtains. An immediate problem with naïve semantics concerns names that do not denote anything, such as 'The Easter Bunny' and 'Vulcan.' Since there is no rabbit that doles out colored eggs, nor any planet orbiting between Mercury and the sun, there are no events or actions or states of affairs involving The Easter Bunny or Vulcan; from which it follows that any sentence using such names is meaningless–which seems to fly in the face of common sense about the meaning of these kinds of sentences.

Naïve semantics is, of course, a kind of realist semantic theory. The problem presented, therefore, is really only a problem for certain realist semantic theories; for, it is only they who might view meaning as some class of items in "reality." But it can be seen that the central pre-supposition of naïve semantics–that *something* is correlated with each sentence to give it meaning –is a principle that holds more widely than just for realist theories. In fact it is presup-posed by every "non-holistic" cognitive semantic theory. And it is these non-holistic cognitivist theories in which I am interested here. I wish to see whether it's really true, as cognitivist theorists believe, that these problems do not apply to their theories.

What gives a public sentence meaning, these non-holistic cognitivist theories say, is a mental image or other representation (such as a statement in a "mental language") of the event or state of affairs envisaged by the sentence. And this holds regardless of whether the various items mentioned in the sentence exist in "reality," because what counts according to these theories is that they exist "in the mind." So long as the speaker and audience have mental correlates of the item,

then the public sentence has an appropriate grounding for its meaning.[1] So a sentence about The Easter Bunny has meaning for a speaker if s/he has a mental item correlated with this term, and communication with an audience succeeds only if they both have a representation which is occasioned in the correct manner.

There is a second, related problem brought out by naïve semantics which perhaps does not find so convenient an account in cognitivist views on meaning, and that is the issue of "neg-ative existentials." A sentence like 'Vulcan does not exist' is self-refuting in a naïve semantics, for if there is something in reality that answers to the name 'Vulcan' then the sentence is false, while if there is not anything answering to the name then there is nothing about which we can say that it does not exist. So on naïve semantics negative existentials are either false or mean-ingless. And any semantic theory that admits there to be sentences whose truth presupposes their own meaninglessness is surely a self-refuting semantic theory. Cognitive theories of meaning need to provide some explanation of how negative existentials can be meaningful.

What can a cognitivist view say about this problem? The expedient of simply deferring the meaning to some mental structure in the same way as done with the first problem does not work without some modification; for, although such a theorist might want to say that they have a mental image or other item that means/represents 'Vulcan,' we need in addition to have some sort of distinction drawn between terms that represent non-existing items and terms that represent existing items. Otherwise we will be unable to distinguish the meanings of 'Vulcan exists' and 'Vulcan does not exist.' (Of course, in an internalist semantics, this is a matter of whether the speaker *thinks* it exists or doesn't exist, not whether "in reality" it does or does not exist). So, the mental model of the world that the speaker grounds his/her meaning in must contain a distinction between items that (are believed to) exist and those that do not. In effect a negative existential says that the subject-term of the sentence is correlated with an object in the "doesn't exist" portion of the mental

[1] Again, we do not pause here to consider the critiques leveled by realists against this view of meaning. My goal here is instead to explain the theory.

theory, while a positive existential says that an object in the "existence" portion of the mental story is being indicated.

Can this general account can be extended? First, can it be extended to hold of negative existential *general* statements, such as 'Unicorns do not exist'? Presumably the cognitivist theory takes the mental item corresponding to 'unicorn' and somehow places it into the "does not exist" area. We return later to what this might mean, but it is not at all clear how the same move can be made with sentences like 'Round squares do not exist,' for under some versions of what these mental correlates are there cannot be one for items that can't *possibly* exist.

Whatever the particular solution to general negative existentials is chosen, we see the strat-egy being pursued. The cognitivist theory attempts to restate certain troublesome sentences of public language by a paraphrase that does not involve this new restatement in the same dif-ficulty that the naïve semantic theory posed for the original public statement. In a (non-holistic) cognitive semantic theory one shows systematically which mental items are correlated with each public sentence, and to do it in such a way that the resulting system of mental repre-sentation of public language does not fall afoul of the semantic puzzles raised by naïve sem-antics. The method to be employed when one of these puzzles seems to come into play is to *paraphrase the offending portion of the public language* by some inoffensive portion of the mental correlate.

In my explanation of one of these puzzles I will often use the term "mentalese" or "mental language" for what might more neutrally (but awkwardly) be called "the system of mental representation of public language which gives meaning to the public language." And I will often talk as if it is a matter of "translation" of the public language into this mentalese. Much of this is in accordance with common usage in the cognitive science literature, but I do not wish for anyone to think that the resulting mental system has to be language-like in the way envisaged by certain believers in The Language of Thought. For, exactly the same points hold (I contend) for those who take the mental system to be a system of imagistic items, or categories of prototypes plus rules, or combinations of any of the well-known candidates. What is relevant to the issue is merely that the cognitive theory hold there to be *some* mental structure which gives meaning to any

particular public sentence. And whatever the relation that holds here between the public and the mental is what I call "translation"; and what I call "paraphrase" occurs when the straightforward application of the translation method does not yield an appropriate mental meaning.

Predicate Negation: We have discussed negation in conjunction with existence, generating naïve semantic puzzles for negative existentials of various types. Similar puzzles can also be generated for naïve semantics by negating predicative expressions, and they do not seem to have similar solutions. Sentences such as 'Kim is not plump' employs a predicate negation, which intuitively has the effect of saying that the predicate in question does not apply to the referent of the subject of the sentence. The reason for its lack of application is different in dif-ferent sentences: a matter of fact about the world (as here), an issue in the taxonomy of concepts, a matter of mathematical necessity, and so on. Note that, on the surface, there is no item or group of items which are said not to exist, unlike the case with negative existentials. Within a realist framework, a sentence describes an event or state of affairs or some such thing. Yet this seems self-refuting in the case of predicate negation, as it did in the case of negative existentials, because the very sentence asserts that the relevant event or state of affairs is *not* occurring…and so the relevant meaning couldn't be there if the sentence were true.

When a cognitivist attempts to give the same response to this problem as was given to the problem of negative existentials, it is soon discovered that there are other factors that come into play with predicative negation. The main difference comes with the category of item that is to give meaning to these two different constructions. In the case of negative existentials where the subject term is a singular term, such as 'The Easter Bunny', the corresponding mental item is a "singular item"....just what a singular item amounts to depends on the mental theory, but it might be an idea or image of an individual, or a name in a Language of Thought. In either case it seems reasonably straightforward simply to put this mental item onto the "doesn't exist" section of one's mental theory. When the negative existential is about a general term, such as 'unicorn', the corresponding mental item is a "general"....just what a general item amounts to depends, once again, on the mental theory, but it might be an idea or concept of a class of things, or a general term in a

Language of Thought. In either case the placement of this type of mental item into the "doesn't exist" section of one's mental theory amounts to saying *there are no instances (in my mental theory) corresponding to this concept*. (Of course a concept's being in the "doesn't exist" portion of the mental theory does *not* amount to saying that the concept itself does not exist!)

But sentences are different. To them corresponds some *combination* of concepts, ideas, images or terms. In cognitivist theories that invoke a language-like mental representation, this combination is some syntactic-like nexus of the representations of the items that correspond to the parts of this combination. In such theories, the mental language contains some "mental glue" to make these parts form a whole in much the same way that the syntactic combining of words/phrases employs a "syntactic glue" to form a sentence of the public language out of a mere list of words. Imagistic cognitivist theories likewise envision a meaning assigned to a public sentence by a combination, but this time it is by combining the various "images" that correspond somehow to the parts of the public sentence to form them into an image which represents the entire sentence.

When this picture of sentence meaning encounters negative predications there will be a problem. For, the meaning of a negation is made up somehow from the meaning of its parts, which includes the positive of the sentence. By hypothesis the positive has no meaning: al-though the mental items corresponding to the concepts, ideas, images, and terms mentioned in the sentence are presupposed to exist, they happen not to be combined in the way the (positive of the) sentence mandates. Yet the *combination* is not the kind of item that can be placed in the "doesn't exist" section of the mental model: it is neither an individual nor a general concept.

Two Inadequate Answers: Consider our sentence 'Kim is not plump.' What could a cog-nitivist theory assign as a meaning of this sentence? Naïve semantics says that the meaning of a negation amounts to saying the positive does not occur–there is no state of affairs of Kim's being plump; it then points out that this is precisely to say that the negative sentence has no meaning. How can a cognitivist theory accommodate negation, and the idea that negative sentences are a

function of the meanings of the components of the underlying positive sentence, without encountering this naïve semantics puzzle?

One way might be to posit the existence of items exactly corresponding to these negations. So not only does there exist a combination the Kim-concept and the Intelligent-concept to give a meaning to the public sentence 'Kim is intelligent', but also there exists a combination of the Kim-concept and the Plump-concept and the Not-concept to give a meaning for the public sentence 'Kim is not plump'. This does not seem to be much of an advance in understanding, but merely a redescription of the initial problem. The initial problem was to explain how a *non*-combination (there is no combination of the Kim-concept and the Plump-concept) could be said to exist so that it can serve as the meaning of a negative public sentence. And to answer that this non-combination suddenly becomes a combination when the separate and uncombined parts are mixed with the Not-concept does not seem to be much of an explanation. Another way of putting this objection to the miraculous invocation of a Not-concept is to say that the notion of "not" cannot straightforwardly be invoked in a semantic theory which alleges that the meanings of the unnegated sentence is an *object* – regardless of whether this object is a piece of "the world" (as in the original naïve semantics) or whether it is a piece of one's mind (as in cognitivist theories). The negation of an *object* is not an independently well-defined notion; and it needs to be further explicated before it can be used. Negation is defined semantically only for truth-conditions: a negation is true if and only if the positive is false. But cognitivists of the sort we are considering deny that this truth-conditional aspect of the public language is relevant to their viewpoint. So, any cognitivist theory…even one that posits a Language of Thought…is required to give some explanation of their negation to show what sort of function it is, and what sort of values it takes when applied to the objects that the theory holds to be meanings. The mere *naming* of a piece of the mental machinery as 'NOT' or as '¬' simply gets us nowhere until we know how such items actually function in the mental economy…or indeed, whether it is even distinct from any other concepts. They do *not* work like their identically-named apparatus of traditional externalist semantic language, and it is an inadequate cognitivist theory that does not offer any account of the operation of such a sign.

A second way to try to introduce something to be the referent for negations of public language sentences would be simply to increase the number of concepts or primitive terms in the mental language. As before, each positive predicate term of the public language is correlated with some mental item, but now each negation of a predicate term of the public language finds an independent mental item correlated with it, a term which is unrelated to the positive one. There is a mental item (concept, term, etc.) corresponding to 'plump' and likewise one corresponding to 'not-plump'. Positive sentences such as 'Kim is plump' and negative ones like 'Kim is not plump' do *not* make use of the same predicate concept. Rather the positive predicates and the negative predicates are correlated with entirely distinct and logically unrelated mental items. It is to be emphasized that this theory does not posit *any* specific relationship between a positive and (what we would intuitively think of as) its negative counterpart, for this view wishes to evade the task of showing how it is that a mental item correlated with negation operates on a concept to form the correlated negative concept.

It seems clear that this last attempt to (not) answer the problem generated by negative predications is a desperate one, and should only be embraced if no other satisfactory account of the interaction of negation with mental items can be found. For one thing it seems perfectly plain that ordinary people *do* think of their positives and negatives as being related to each other; and it is well known that semantic priming effects can be elicited for a positive by priming with the negative, and conversely. The present proposal in effect says that the mental operation of negation has no systematic result when it is applied to different mental items. And this is simply a view that we should hope not to have to adopt. We would want there to be a "negation" operation that applies to predicates, but what can it be in cognitivist theories? As we've seen, the traditional notion of negation does not apply to objects – even if they are objects that are correlated with predicates.

Possible Mental Negations: It seems to me that there are two general directions that cognitivist theories might take in their quest to give some sort of theoretically adequate grounding to the problem of predicate negation. Recall that one of the difficulties with invoking a mental NOT

operator and expecting it to do duty as a mental correlate for the public language negation is that the items of the mental language are by hypothesis objects, and that the operation of neg-ation in the public language is just not defined to be applied to objects. So we need to posit some mental operation that *does* apply to objects, and which we can use to generate the appro-priate behavior of negation (the behavior given by its operation in the public language).

One relation that holds between pairs of objects is the relation of *nonidentity*, and I propose to use it to give one type of account of predicate negation. However suspicious we might be about a mental theory helping itself to such a concept, at least it is a possible property of (pairs of) mental items…arguably even the most basic one…unlike the impossible NOT. I offer it to cognitivists who wish to give some adequate account of predicate negation. There is still some work to be done, because (mental) ≠ is not the same as predicate negation. To use ≠ in an explanation of predicate negation, the cognitivist theory also needs to have the resources of quantification. This is perhaps not so much of a miraculous invention, for any adequate mentalese must be able to accommodate quantification in one way or another. In the formulations I give below I will not dwell on how a cognitivist theory should do this. I expect that different styles of cognitivist theories will find different ways to bring quantification into their theory. Instead I will just write a formula that looks rather like quantified logic with the intent that the quantification that is indicated in the formula will have to be suitably accounted for in the cognitivist theory. Recall also that our different cognitivist theories invoke different types of mental structures to correspond to items of public language. In order to stay neutral as to what these mental items are, I will use boldface to stand for whatever mental item the public language term stands for. Thus, if 'F' or 'a' are terms of the public language, then '**F**' and '**a**' are their correlates in the mental theory being developed.

There are two different ways that a cognitivist theory could invoke ≠. The first claims that (simple) negative sentences of the public language having the form "a is not F" correlate with:

(1) $\forall x(\mathbf{F}x \supset x \neq \mathbf{a})$

That is, the predicate negation amounts to saying that whatever manifests **F** is distinct from **a**. One should be wary about the use of the phrase '**F**x' in this formulation, because '**F**' here stands for

some concept (or other mental item) whereas the universally quantified '**F**x' stands for all the concepts (or other mental items) in which something is being **F**. And it is not immed-iately obvious that the simple **F**-concept will necessarily play a role in the more complex concepts that involve also a subconcept of an individual. But this is an issue that must be faced by any cognitivist theory: how are items that correspond to complex pieces of public language constructed? Are they made in a compositional manner, being constructed out of the mental items that correspond to the public language parts of the public language whole? This is indeed a central issue for cognitivist theories to face, but for the purposes of the present discussion I will assume that they have somehow answered these questions so that the formula in (1) makes sense, indeed, the "natural" sense that was stated immediately after the formula. In this natural sense, it appears that (1) is perfectly fit to be a mental correlate of a simple predicate negation.

The public language sentence that most naturally corresponds to (1) is "a is distinct from all those things which are F". This is not the only way to employ non-identity in an analysis of predicate negation. One could instead say "every property that a exhibits is distinct from F":

(2) $\forall G(G\mathbf{a} \supset G \neq \mathbf{F})$

Although this appears to express a formula of a second-order language in its quantification over predicates, we should remember that in cognitivist theories the predicates are in fact (mental) objects: concepts or images or symbols, etc. Once again, the non-identity is said to hold between *objects*, making it well-formed and a suitable candidate to replace the inappropriate NOT operator that does not hold of objects. The real mystery in both (1) and (2) comes from the quantification. In (1) we need to make sense, somehow, of what "all the objects that manifest **F**" while in (2) we need to make sense, somehow, of what "all the properties that **a** manifests." And in the mentalese of various theories these notions will be given different explanations; the explanation of these quantified phrases will in turn depend on how the theory analyzes simple positive predications. □But this is not an issue to be discussed here.

The second direction that cognitivist theories might move in seeking an analysis of predicate negation is to employ a primitive notion of "incompatibility" rather than of "nonidentity". Again

perhaps critics might think the cognitivist who employs incompatibility as a primitive notion is unfairly availing him- or herself of negation under some other name. But at least incompatibility, like nonidentity but unlike negation, is a relation that *can* hold of objects. Of course, it does not hold of *all* objects, but in the sense to be employed, sense can be given to its holding of the type of objects required for the analysis of negation.

The fundamental idea here is that mental entities that correspond to predicates come in "applicability groups." This means that there groups of these mental entities such that, for any (mental correlate of an) object, at most one member of that group can be manifested by that object. A simple example might be the mental concepts corresponding to the public language color predicates. They form such an applicability group, so that given any (mental item corresponding to an) object, it can manifest at most one member of this group – that is, it can have at most one (overall) color. It seems clear that a cognitive theory would be entitled to adopt such an organization of its mental economy.

Let us use the symbol ⊗ to signify belonging to the same applicability group, so that **F**⊗**G** means that **F** and **G** are incompatible. Then we could paraphrase "a is not F" as

(3) $\exists G(G \otimes \mathbf{F} \,\&\, G\mathbf{a})$

or in other words that there is some concept which **a** manifests that is incompatible with **F**. Once again, it seems that this analysis captures the semantic features of negation in the public language, and does it without invoking an incoherent notion of NOT in a mental theory. As with analysis (2), it appears that (3) is a second-order formulation involving quantification over predicates. But again I remark that the mental correlates of the predicates are *objects* in these cognitivist theories, and so looks to the contrary notwithstanding, (3) does not in fact involve quantification over second order properties, but only over ordinary (mental) objects.

Conclusion: I have presented three possible analyses of predicate negation that a cognitivist theory of language could avail itself. Each of the analyses avoids the mistake of trying to direct-ly import the public language NOT, which is not applicable to the mental items of a cognitivist theory. Yet each of them accounts for the basic features of the negation in public language, and each of

them avoids the puzzles brought out by naïve semantics. For, none of them tries to ground a negation in the non-existence of the very object that is required to give it meaning.

Which of the analyses (1), (2), and (3) that is adopted by any given cognitivist theory will depend on other features of the theory that determine whether one or another of them better meshes with what else is in the theory. It seems to me that cognitivist theories that postulate an internal, mental economy that is to account for the meaning of public language should do more toward addressing these foundational issues than has been done in the past. Even if one postulates a mental theory which is just another language having more or less the same syntactic form as the public language, there are differences between the two that need to be carefully considered. Such theories are not free just to adopt semantic devices from the public language, for the public devices were designed to apply to a different aspect of the world than is available in a strictly internalist cognitive theory. A theorist who claims that the mental items in his/her mental economy are semantically akin to items in public language, so that notions such as NOT can be directly incorporated, is not employing a strictly internalist account of issue. Such theories are not considered in this article, but one wonders whether the miraculous invocation of semantic properties to mental items doesn't just make the mental items otiose. Why duplicate what is already in the public language?