# Possible worlds semantics

PHIL 93914
Jeff Speaks

January 16, 2008

## 1   From extensions to intensions

One response to the incompleteness of the theory of reference is to blame its incompleteness on the fact that the theory focuses exclusively on the reference of expressions with respect to the actual world. The meaning of an expression, on this view, comprises not just the expression's reference in the actual world, but also what the expression would have referred to, had the actual world been different.

This is the guiding idea of possible worlds semantics. If we use *extension* as a label for the reference of an expression — so that the extensions of names are objects, and the extensions of simple predicates are functions from objects to truth-values — we can introduce *intension* as a label for an expression's reference across possible worlds. So, for example:

- The intension of a name is a function from possible worlds to objects.

- The intension of a predicate is a function from possible worlds to functions from objects to truth-values.

- The intension of a sentence is a function from possible worlds to truth-values.

- What would you give as the intension of a quantifier like 'Nobody'? (See Lewis, 'General semantics', §VII)

('Intension' is often used as a synonym for 'meaning.' But, strictly, it is a function from possible worlds to extensions; it is a matter of substantive debate whether meanings are intensions.)

Let's see how this theory fares with three of the limitations of theories of reference:

1. It assigns propositions to sentences which are distinct from their truth-values.

2. It helps with explaining the behavior of expressions in non-extensional contexts; for example, we can now explain the truth value of sentences of the form ⌜Necessarily, $S$⌝. We could treat 'Necessarily' as a function from functions from worlds to truth values to functions from worlds to truth values — i.e., as a function from sentence intensions to sentence intensions. In other words, 'Necessarily' would have as its reference a function which took as argument the intension, rather than the extension, of '$S$.' (Which such function would it be?)

   It also provides some help with attitude ascriptions. Perhaps we can say that an ascription ⌜$A$ believes that $S$⌝ says that $A$ bears the belief relation to the intension of '$S$'. (So 'believes' might be a function from ordered pairs of objects and sentence intensions to sentence intensions.) This explains why materially equivalent sentences cannot be substituted salve veritate in the complements of attitude ascriptions.

3. It appears to help with the problem of saying what speakers know when the understand a sentence. For, even if a speaker who understands a sentence needn't know its truth-value, she arguably must know what its truth-value would have been had the world been this or that way. And this is in some way to grasp the relevant function from worlds to truth-values.

4. It is an account of meaning on which the meaning of an expression straightforwardly determines its reference. Meanings are functions from worlds to extension; so, given as argument the relevant world, the meaning of an expression delivers a reference as value.

## 2    The problem posed by propositional attitude ascriptions

However, propositional attitude ascriptions pose a problem which many see as the fundamental problem with possible worlds semantics. Let's make two assumptions explicit:

> *The naive relational theory of attitude ascriptions*: ⌜$A$ believes that $S$⌝ is true iff $A$ bears the belief relation to the meaning of $S$. (We're ignoring context-sensitivity for now.)

> *The distribution of belief over conjunction*: if ⌜$A$ believes that $S$ and $T$⌝ is true, then so is ⌜$A$ believes that $S$⌝ and ⌜$A$ believes that $T$⌝.

We can then argue as follows. Let '$S$' be some sentence, and let '$T$' be a necessary consequence of '$S$'. Then the intension of the conjunction of '$S$' and '$T$' will be the same as the intension of '$S$.' But then, from the view that meanings are intensions and the naive relational theory of attitude ascriptions, it follows that if ⌜$A$ believes that $S$⌝ is true, so is ⌜$A$ believes that $S$ and $T$⌝. But, by the distribution of belief over conjunction, it follows that if ⌜$A$ believes that $S$ and $T$⌝ is true, so is ⌜$A$ believes that $T$⌝. So every agent believes all of the necessary consequences of everything that he believes.

This should strike you as bad. We can further bring out the badness in two ways: (1) Necessary truths are necessary consequences of everything. So, anyone who has any beliefs at all believes all necessary truths. (2) Every proposition is a necessary consequence of a necessary falsehood. So, if anyone believed a necessary falsehood, they would believe every proposition. But no one believes every proposition. So no one believes any necessary falsehoods. (For more detail on this argument, see Soames, 'Direct reference, propositional attitudes, and semantic content.')

## 2.1 Stalnaker's way out

Stalnaker's central response to this problem (in *Inquiry*) is to deny the naive relational theory of attitude ascriptions: the view that an ascription $\ulcorner \alpha$ believes that $\sigma \urcorner$ is true just in case the referent of the value of '$\alpha$' bears the belief relation to the semantic content of the value of '$\sigma$' (in the context of the ascription). Instead, Stalnaker thinks, such ascriptions sometimes report a relation to a meta-linguistic proposition about the truth of the sentence in the complement clause of the ascription. Because this proposition will always be contingent, and the possible worlds account of the objects of belief runs into trouble precisely with necessarily true and necessarily false propositions, this meta-linguistic reinterpretation promises to deliver a more intuitive assignment of truth-conditions to attitude ascriptions than the unmodified possible worlds theory.[1]

The main problem with this analysis is not so much that it is implausible as that it does very little to palliate the counter-intuitive consequences of Stalnaker's theory. Consider the sentence, "No whole number raised to a power greater than two is equal to the sum of two other whole numbers, each raised to that power." This is an example of a sentence which poses problems for the view of the objects of belief as sets of possible worlds, because (i) since it expresses a necessary proposition, it follows from the closure of belief under necessary consequence that any agent who has any beliefs at all believes what it says, and yet (ii) there is no difficulty in finding an example of an agent $A$ such that the sentence

[1] $A$ believes that no whole number raised to a power greater than
two is equal to the sum of two other whole numbers, each raised
to that power.

seems clearly false. Intuitively, many agents have beliefs without believing Fermat's last theorem. The meta-linguistic strategy is designed to block our having to treat [1] as true in these cases by interpreting it as attributing to $A$, not belief in the necessary proposition expressed by

[2] No whole number raised to a power greater than two is equal to
the sum of two other whole numbers, each raised to that power.

---

[1] Note that Stalnaker does not deny that, for example, anyone who has any beliefs at all bears the belief relation to the (one and only) necessary proposition, expressed by, among many other sentences, "Arithmetic is incomplete"; what he denies is that, in all such cases, an ascription $\ulcorner \alpha$ believes that arithmetic is incomplete $\urcorner$ will be true.

but rather belief in the contingent meta-linguistic proposition expressed by the sentence

[3] "No whole number raised to a power greater than two is equal to
the sum of two other whole numbers, each raised to that power"
is true.

Since the proposition expressed by [3] is contingent, closure under necessary consequence doesn't entail that $A$ believes it; hence Stalnaker's semantics for belief ascriptions seems to make room for the wanted result that [1] is not true.

A problem with this strategy of systematically reinterpreting attitude ascriptions is that agents often in the problematic cases have beliefs about the meaning of the relevant sentence. For example, we can suppose that the agent of [1] believes the proposition expressed by [4]:

[4] "No whole number raised to a power greater than two is equal to
the sum of two other whole numbers, each raised to that power"
means that no whole number raised to a power greater than two
is equal to the sum of two other whole numbers, each raised to
that power.

The problem is that, given that Fermat's Last Theorem is a necessary truth, [4] has [3] as a necessary consequence. So, by the closure of belief under necessary consequence, if Stalnaker grants that the agent believes what is expressed by [4], he must grant that the agent believes what is expressed by [3]. But, if the agent believes what is expressed by [3], [1] is true on the meta-linguistic reinterpretation of the ascription — which is the result that the reinterpretation was introduced to avoid.

## 2.2  Structured intensions

Lewis recognizes a version of this problem in §V of 'General semantics', where he notes that we are inclined to make distinctions in meaning which do not correspond to any difference in sentence intension. He suggests that these intuitions can be accommodated in his framework by thinking of meanings as 'structured intensions'. On this view, the meaning of a sentence is not just a function from worlds to truth-values, but rather a structured object which has as its constituents the intensions of the semantically simple units which make up the sentence. (Think of the structure as what is contributed by the syntax of the sentence.)

We could use this view to solve the above problem if we took propositional attitude ascriptions to relate agents to the structured intension of the complement sentence, rather than its intension simpliciter. To take this way out is to significantly change the possible worlds theory. One basic question about propositions is whether they are structured things which have constituents. Many have worried that the relevant sense of 'constituent' is obscure. But for our purposes the relevant point is that the suggested way of making the required distinction between meanings pushes the possible worlds theorist from a view of propositions as unstructured to one on which they are structured, and have constituents.