# Thomas and Substitution

The leading "Quantifier-Reinterpreting" strategy for dealing with quantification over nonexistents was, you remember, Ruth Marcus' substitution-interpretation approach, discussed at the end of *M&M* Ch. 1. I reproduce the latter discussion here (minus footnotes):

Ruth Marcus has proposed that Meinongian quantification be understood as *substitutional* quantification.

This suggestion has a good deal of intuitive appeal. It is quite plausible in the case of "There are things that don't exist": when I utter that sentence aloud, I feel a tendency to continue by listing true negative existentials ("...you know: the round square doesn't exist, Macbeth doesn't exist, the free lunch doesn't exist, and so on"). And there certainly is plenty of overt quantification in English that is substitutional. It might be thought that Marcus' proposal fails in the case of possible worlds on the grounds that worlds in general do not have names at all (the substitution class would be far too small); but we may easily generate a system of canonical names for possible worlds from existing resources: Each world, we may suppose, is correctly described by a maximally consistent set of sentences {*P*, *Q*, *R*, ...}; to obtain a name of a world in which P, Q, R, ..., simply form a definite description from the latter indefinite one: "($\iota w$)$\text{In}_w$(*P* & *Q* & *R* & ...)."

This proposal will inherit the usual sorts of problems that philosophers have raised for substitution interpretations of more familiar quantifiers. The most obvious of these is the "not enough names" problem: First, given that almost any real-valued physical magnitude characterizing our world will have nondenumerably many nonactual worlds corresponding to it, the cardinality of the set of all worlds--if that notion is not undefined or paradoxical--will be inconceivably high. But there are only denumerably many names of the sort exemplified above; thus, it seems, universal quantifications over worlds will be verified more easily than we would like. Second, a number of philosophers have argued that the substitution interpretation somehow collapses into the standard interpretation when incorporated into a full-scale truth theory. Kripke (1976) refuted at least the most salient versions of this charge, though I shall argue in Chapter 9 below that there is a further, somewhat related difficulty which impugns the usefulness of the substitution interpretation for the truth-conditional analysis of *natural* languages.

As I see it, the main problem for Marcus' proposal is the same that arose for the counterfactual approach: How will Marcus reinterpret set abstraction and all the other operations that will need to be applied to names of "worlds"? (Certainly there have been attempts at metalinguistic reinterpretations of set abstraction, but they have concentrated on reinterpreting the abstractor itself, not the variable it binds; that is, they have concentrated on detoxifying mention of the *sets* in question, not mention of the sets' members.)

<>

Let's take the "usual sorts of problems" one by one and see whether they afflict Thomas' proposed semantics for number, property etc. talk. N.b., as he said, he did *not* join Marcus and extend his semantics to quantification over nonexistents.

<u>"Not enough names"</u>

Over the years this has proven to be a hairy issue, because there is no consensus as to what is required for something to count as a "name." We can say right now on the substitutionalist's behalf that (a) "name" doesn't mean *proper* name as opposed to any other sort of singular term, and (b) it doesn't mean anything like "*genuine*" singular term, because as Thomas says, in applying the substitution interpretation we want to include syntactic positions that are only superficially singular-term-like. But there's a further persistent issue about, so to speak, *how actual* a name must be. (1) Must there be an actual token *of it*? Then the substitution interpretation is hopeless. (2) Must there be a compositional structure constituting it, which structure itself may be abstract but where each of the primitives has actually been tokened? That would mean the set of such abstract names would still be denumerable, so couldn't handle real-valued domains or pseudo-domains.

In his prize-winning paper, "Inexpressible Properties and Propositions," Thomas makes some useful further distinctions. Then in his own proposal for what's to count as a "name," he appeals to an infinite class of (what are syntactically but not pragmatically) demonstratives. I haven't worked carefully through the apparatus, and Thomas should correct me if what I'm about to say is wrong, but it looks to me as though he's talking about *merely possible tokens of* the demonstratives.

For the purpose of dealing with numbers, properties or propositions, that may be no objection; whether it is depends on some other things, such as whether Thomas would then want (as I do) to appeal to properties in explicating "mere possibilia." (And Sellarsians have made inspired use of merely possible name-tokens; see especially Mark Lance, "Quantification, Substitution and Conceptual Content," *Noûs* 30 (1996): 481-507.) But for purposes of explaining quantification over mere possibilia themselves, it plainly would not do!

<>An option mentioned by Lewis (*Plurality*, sec. 3.2) is the use of "Lagadonian" languages in which things serve as their own names; but a proponent of Thomas' view can hardly propose to let properties and propositions serve as their own names, because his idea is to do entirely without a domain of properties or propositions in the first place. (Similarly and even more obviously, Marcus cannot suggest letting actual possible worlds serve as their own names.) <>

<u>Collapse into objectuality</u>

This is a complicated and technical issue. What I had in mind is a 1971 argument of John Wallace's to the effect that the substitutional interpretation clause in a Tarskian truth definition for a predicate calculus is not sufficient to yield the required T-sentence, "True('(Ex)Fx') <—> (Ex)Fx." More to the point, not even adding all the axioms "True('Able is F') **<-–>** Able is F," "True('Baker is F') <--> Baker is F," "True('Charlie is F') **<-–>** Charlie is F," etc. will afford the derivation. What would be sufficient, and is the obvious candidate, is to add (i) a notion of denotation **Δ**, (ii) the recursive clause "⌐*s* is F ¬ is true <--> F(**Δ**(*s*))," and (iii) each of the axioms "**Δ**('Able') = Able," "**Δ**('Baker') = Baker," etc. Assuming Wallace is right at least on the first point, that would predict that we don't really understand bare substitutionally quantified formulas, since (I take it) understanding requires at least the tacit ability to derive pure T-sentences.<**1**>

On the basis of that and some other case studies, Wallace draws the general moral that "[objectual] quantification theory reads itself into any language it interprets." I'm not sure how I stand on that broader and more abstract claim.

But in any case, Kripke has taken sharp issue with this material and related arguments of Wallace and the late Leslie Tharp. See his monograph, "Is There a Problem About Substitutional Quantification?," in Evans and McDowell (eds.), *Truth and Meaning* (OUP, 1976)—a fascinating discussion, and required reading for anyone who wants to understand the relevant philosophical issues.

<u>Semantic defectiveness</u>

I argued in 1979, and (independently) Peter van Inwagen argued in 1981, that there is a semantic problem about interpreting *natural-language* expressions substitutionally. Very crudely: Although substitutionally interpreted quantification *is* interpreted and its truth-conditional contribution is clear, that does not suffice to determine a meaning for it, *and* there is broadly syntactic reason for doubting that it has a meaning in the sense of saying anything in particular. Here's the *M&M* version of that material (which quickly explains the basics of the substitution interpretation if you want to review them). Van Inwagen's paper is "Why I Don't Understand Substitutional Quantification," *Philosophical Studies* 39 (1981): 281-85, online only to Springer subscribers.

But I think our argument does not, or anyway need not, apply to Thomas' formulation in terms of infinite disjunctions, because it's clear enough that an infinite disjunction does say something in particular, and indeed it's clear what the disjunction says.

(The corresponding substitutionally quantified formula is indeed truth-conditionally equivalent to the infinite disjunction, but part of my original argument is that truth-conditional equivalence does not show sameness of meaning, nor does it show that if S says something and T is truth-conditionally equivalent to S, then T says something.)

<u>Thoroughness of the paraphrastic program</u>

<>Here again was my complaint about Kripke's counterfactual analysis, complete with the footnote about virtual classes.

A more promising Paraphrastic program would be to understand "possible-world" talk counterfactually, as has been suggested by Kripke…. This is quite a natural suggestion and does much to make talk of possible worlds more homey. An antic sentence such as "In some possible world distinct from our own, Richard Nixon is a Black Panther" might be paraphrased as "Had things been otherwise, Richard Nixon might have been a Black Panther," a sentence which we all more or less understand or at least would not balk at in ordinary conversation.

The counterfactual approach is inadequate in two serious ways, I think. First: It is not enough to provide a sample paraphrase or two. The counterfactual theorist would have to work out a systematic and rigorous *formula* for paraphrasing formal, model-theoretic sentences concerning possible worlds, and in such a way as to preserve all the theorems of our logical theory and all the advantages of each of the modal logics or modal semantics under analysis. It is hard to imagine how this would go.[27] The difficulty becomes critical when we note that any adequate modal semantics will require many *sets of* possible worlds, sets of sets of worlds, and so on…. The counterfactual approach is not allowed to leave set abstraction on worlds undefined. (This seems to me to be a crucial point, one that I have never heard a Paraphrastic theorist address.) Even if we have provided a satisfactory system of eliminative contextual definitions for *quantification* over nonexistent possibles, this system would have to be extrapolated to cover set abstraction as well, and no way of doing this in terms of counterfactuals comes to mind. (We might try invoking "ways things might have been" and abstracting on them, but to do that would be to reify the "*ways*" and leave us with all the same problems we had before.)...

———————————

<>27. The point I am making here is very easily overlooked by philosophers seeking paraphrastic eliminations of the metaphysically dubious entities assumed by some formal and highly technical theory. Consider a simple (and mythical) example: It seems that sentences of the form "$x\ \varepsilon\ \hat{y}(Fy)$" can be paraphrased simply as "*Fx*," as Quine has observed. Thus, some class abstracts may be regarded as *façons de parler*. A philosopher who lacked Quine's own mathematical sophistication might well come to think that nominalism had been achieved, in that Quine had hit upon a program where class abstraction and talk of classes could be paraphrased away. What this naive philosopher would be overlooking is that simple class abstraction is not the only technical operation that occurs essentially in set theory. In this case, as Quine points out, one would not even be able to explicate talk of classes of classes, since his "virtual class" device leaves undefined any construction in which a bound variable occurs immediately to the right of "$\varepsilon$."

A nonmythical example of this optimistic sort of fallacy is some philosophers' reaction to Wilfrid Sellars' approach to abstract entities. Sellars, mobilizing his ingenious device of dot quotation, has offered some very plausible paraphrases for simple talk of properties, propositions, sets, and so on. Philosophers justly impressed by the cleverness and by the naturalness of these paraphrases have taken Sellars to have offered an acceptable nominalistic *theory of* abstract entities generally. But Sellars has given us no reason at all for thinking that the rarefied operations of (e.g.) graph theory, the integral calculus, differential geometry, or other areas of higher mathematics can be explicated in terms of dot quotation, since he has provided no directions in which his original paraphrases of simple and relatively nontechnical constructions are to be extrapolated. (Sicha [1974]) has tried to provide at least one such direction, but the mathematics he is able to treat is very elementary indeed.)

Now, Thomas has given us one sort of sample paraphrase, via the "trivial" inferences. E.g., "Fido has the property of being a dog" can be reduced to just "Fido is a dog," so the apparent reference to a "property" can be seen as mere misleading surface grammar. (Likewise, "The proposition that Fido is a dog is true" goes to "It is true that Fido is a dog.") But how are we to extrapolate those simple paraphrases into an entire paraphrastic program covering all the complex higher-order references to properties and/or propositions that various theorists have made?

Notice that Thomas' same sort of "trivial inference" / "merely a focus construction" appeal could be made on behalf of virtual classes themselves. The following sounds "trivial" in Thomas' way: "Fido is a dog. Thus: Fido is a member of the class of dogs. Thus: There is a class of which Fido is a member, namely that of dogs. Thus: There are classes, among them that of dogs." All well and good (not really—see the next section); but for the reason Quine gives, the availability of these simple paraphrases give us no reason to think that all or even any further references to classes or sets can be paraphrased away. In fact, we know they cannot be.

<> A pertinent case of rich and demanding appeal to properties is Shapiro's and my own Ersatzing construction of "worlds" out of structured sets of structured sets of…properties and individuals. Set abstraction on top of property abstraction on top of set abstraction on top of property abstraction for a hundred years. There may be reasons why Thomas' simple paraphrase won't extrapolate that are exactly parallel to the case of virtual classes (I haven't checked); but even if not, there will be reasons of the same kind.

So: My first three objections to Marcus either remain debatable or plainly do not extrapolate back to Thomas' cases or both. The fourth, I think, does.

And now:

<u>My more basic objection to Thomas' line</u>

The reason I don't really find the "Puzzle" puzzling is that the original inferences, which Steve Schiffer aptly called "something-from-nothing" inferences, don't sound trivial to me—precisely because they give you something from nothing. In fact, when I read them in Schiffer in the 1990s, I thought, "How can he say this? It's just the sort of thing Quine warned us against, viz., incurring a quiet little existential commitment and pretending it isn't one." Perhaps that was too quick a Quinean reflex, since all parties agree that ontological commitment can't simply be read off surface grammar. But we can see a more substantive reason why the inferences are not trivial.

<> It is true that Thomas' "metaphysically loaded counterparts" are focus constructions. Doubtless they result syntactically from focus-introducing processes. But it does not follow that they are trivially equivalent, or equivalent at all, to the respective simpler "innocent statements." As we shall see shortly, a focus-introducing process can add new material as well as merely inducing focus.

Notice carefully that in Thomas' one uncontroversial example of structural focus, Clefting, no common noun or other substantive term is introduced at the surface. What are introduced are only either "expletive" traces or relative pronouns coreferring with other terms already in the sentence: "<u>It</u> is Johan that likes soccer," "<u>It</u>'s Johan <u>who</u> likes soccer," "Johan is <u>the one</u> <u>who</u> likes soccer," etc.) (The same is true of other focal constructions, such as simple Topic Fronting as in "Soccer Johan likes.") So even if we agree that the metaphysically loaded counterparts are focus constructions, we should acknowledge that they have elements additional to those normally found in focus constructions.

Consider these:

(1) Oswald killed Kennedy.

(2) It was Oswald that killed Kennedy.

(3) Oswald was the man who killed Kennedy.

(4) It was Kennedy Oswald killed.

(5) Kennedy is the one Oswald killed.

(6) Killed was what Oswald did to Kennedy.

(7) Killed was what was done to Kennedy by Oswald.

<>(7) is particularly interesting in that it's nested: focus within focus, ranked. The main focus is on the type of act, killing, but a secondary focus is on the victim as opposed to the protagonist.

Now, did you notice? I sort of hope not, because my purpose is to show how easily you can be fooled by data of this sort. (But if you did notice, good for you.) Look at (3). (3) may sound as redundant as the rest of (2)-(7), but it certainly is not. (3) entails, as (1) does not, that Kennedy was killed by a man. Which makes it a far from trivial inference from (1), however routine it may have sounded as it went by. (Thanks to John Roberts for this example.)

What inferences sound trivial depends on what we're assuming. In 1963, anyway, it was simply tacitly assumed that the assassin was male.

As Thomas notes (p. 263), Hartry Field maintains that in the case of Jupiter's moons, the "trivial" equivalence holds only assuming there are numbers. I think Field is exactly right. The equivalence of (3) with (1) holds only assuming that the assassin was a man. The equivalence of "Fido has the property of being a dog" with "Fido is a dog" holds only assuming that *there are properties*. That of "The proposition that Fido is a dog is true" with "It is true that Fido is a dog" holds only assuming that *there are propositions*. And so on. <>

I don't say dogmatically that Field is right and Thomas is wrong; it's an empirical question. I do take Field's side until given more reason to switch to Thomas'.

## Note

1. A "pure" T-sentence is one from whose RHS all semantical terms have been purged through Tarskian derivation. Thus, "'Bill is athletic and Thomas is handsome' is true iff 'Bill is athletic' is true and 'Thomas is handsome' is true" is an impure T-sentence; the corresponding pure T-sentence is just "'Bill is athletic and Thomas is handsome' is true iff Bill is athletic and Thomas is handsome."