

## The Varieties of Extensionalism

Andrew Ward  
University of Kansas

In The Language of Thought, Fodor claims that

A typical thesis of positivistic philosophy of science is that all true theories in the special sciences should reduce to physical theories 'in the long run'.<sup>1</sup>

For those who ascribe to such a "unity of science thesis", the correlate to this claim in the philosophy of language is that any sentence in a non-extensional language is translatable (without loss of meaning) into a sentence in some extensional language.<sup>2</sup> While such a reductionist thesis may appear quite radical (indeed, let us refer to it as "radical" extensionalism<sup>3</sup>), a lot depends upon what is to count as an extensional language. Following the insights of Carnap's Meaning and Necessity,<sup>4</sup> one means of defining an extensional language (or, in Carnap's words, an extensional system) is to say that a language is extensional just in case all of its sentences are extensional. With regards to the definition of an extensional sentence, Carnap claims that a sentence is extensional just in case the replacement of any designative component<sup>5</sup> of that sentence by some other appropriately related (e.g., co-referential) designative component leaves the truth-value of the sentence unchanged. Accepting this as an account of what an extensional language is, it follows that the rationale of one's ascribing to any sort of reductionist thesis (for the philosophy of language) turns on the rationale for claiming that there is some relation such that the substitution of appropriately related designative components does not alter the truth value of the sentence in which the substitution takes place. For Carnap this is complicated by the recognition of (at least) four types of designators: (declarative) sentences, predicates, functors, and singular terms. Carnap's complexities aside, the assumption of this paper is that the function of singular terms in extensional sentences is paradigmatic of the function of other types of designators. Hence, if one can understand both (i) what the adoption of radical extensionalism in the case of singular terms amounts to

and (ii) what the "obligation" is that some feel concerning the adoption of radical extensionalism for singular terms, then one will understand the motivation for claiming that radical extensionalism is applicable in the case of other types of designators. Accordingly, the first thing to do is to formulate an unproblematic version of radical extensionalism as it applies to singular terms. To this end the paper adopts the lead of Michael Dummett (and Frege, depending upon how accurate Dummett's interpretation of Frege is on this point) and offers the following "Dummett-like" formulation:

(ET) If both 'a' and 'b'<sup>6</sup> are singular terms and F( ) is any predicate, then from the truth of both 'a is equivalent to b' and 'F(a)', the sentence 'F(b)' is logically entailed.<sup>7</sup>

Although seemingly innocuous, there is an important sense in which (ET) is doubly ambiguous. First, (ET) does not itself provide an explanation of what it means to be a singular term, and second, (ET) does not specify the nature of the equivalence relation connecting 'a' and 'b'. With regards to the first matter, more will be said later. At present an intuitive understanding of what it means to be a singular term will suffice. That is to say, let us suppose that by 'singular term' we mean names, pronouns, definite descriptions, etc.<sup>8</sup> The second matter, however, needs immediate attention. Without any definite specification, the relation, denoted by the phrase 'is equivalent to', which binds 'a' and 'b', could be any relation satisfying the requirements of reflexivity, symmetry, and transitivity. For instance, we surely do not want this relation (call it 'R') to be '...has the same orthographic structure as...', and yet this relation is an equivalence relation. As Ruth Marcus has pointed out,<sup>9</sup> the problem is that there is no relation which intuitively presents itself as the sole viable candidate for R. Instead, there seem to be a number of relations which, prima facie, could sanction the entailment demanded by (ET). What distinguishes these relations from one another is not the ability to sanction the entailment in (ET), rather the distinction rests with the variation in the constraints they impose on what set of sentences are entailed. Since radical extensionalism claims that all sentences in a language are extensional (or translatable into an extensional sentence), then the specification of R must be such that given the truth of 'F(a)' and 'aRb', the truth of any 'F(b)' necessarily follows.

Classically, the most common specification of R would seem to be:

(ER) '...has the same denotation as...'

Thus, let us for the moment suppose that the relation, R, in (ET) is (ER). If, in addition, we suppose that the above adequately captures the thesis of radical extensionalism as it applies to singular terms and that 'Eq.' is to be used in what follows for (ER), then the following represents an acceptable proof schema:

p 1. 'A' Eq. 'B'

p 2. F(A)

Hence 3. F(B) by (ET)

Indeed, this proof schema does seem to hold for such simple instances as:

p 1.' 'John' Eq. 'the butler'

p 2.' John is tall.

Hence 3.' The butler is tall. by (ET)

But consider the following instance of the proof schema:

p 1." 'John' Eq. 'the butler'

p 2." Frank believes that John is rich.

Hence 3." Frank believes that the butler is rich. by (ET)

Russell referred to the failure of 3." to follow from 1." and 2." as the problem of propositional attitudes.<sup>10</sup> The problem is that while it might very well be true that 'John' and 'the butler' have the same denotation and that Frank believes that John is rich, (and so both 1." and 2." be true), Frank might nonetheless fail to believe (perhaps because the butler leads a double life that is unknown to Frank) that the butler is rich (and so 3." be false). However, if it is possible that both 1." and 2." are true and 3." false, then the truth of 3." does not follow from the truth of 1." and 2." Yet in as much as 1.", 2.", and 3." are instances of 1, 2 and 3 respectively, then according to (ET) if 1." and 2." are both true, then 3." must also be true. Clearly, something has to be given up. The question, of course, is "what should we give up?". Reflection suggests that there are two distinct alternatives.

First, one could claim that it is false that both 1." and 2." could be true and 3." false. On the face of it, however, this sort of claim requires us to distinguish (at least) two sorts of belief. Specifically, we are led to say that while it may be true that it is

false that Frank (actively) "entertains" the belief that the butler is rich, this is not what 3." is about. Rather, 3." is making the weaker claim that Frank "has" the belief that the butler is rich.<sup>11</sup> The point is that while all of us (presumably) "have" the belief that cows do not play chess, it is also probably true that very few of us "entertain" the belief unless they need it as a philosophical example. Thus, in the case of Frank and his beliefs, while:

3a" Frank entertains the belief that the butler is rich.

does not follow from (in conjunction with the truth of 1.") either Frank having the belief that John is rich or Frank entertaining the belief that John is rich,

3b" Frank has the belief that the butler is rich.

does follow from (in conjunction with the truth of 1.") either Frank having the belief that John is rich or Frank entertaining the belief that John is rich. Although this distinction seems efficacious in the case of Frank and his beliefs, it is not, in itself, sufficient to resolve the problem with all propositional attitude contexts. To see this, imagine the following:

- 1."" Set theory is equivalent to set theory plus the Banach-Tarski theorem.<sup>12</sup>
- 2."" Frank believes that set theory is true.
- 3."" Frank believes that set theory plus Banach-Tarski is true.

From the examination of 3a" above, it seems unproblematic that:

3a"" Frank entertains the belief that set theory plus Banach-Tarski is true.

does not follow from (in conjunction with the truth of 1."" ) either

2a"" Frank has the belief that set theory is true.

or

2b"" Frank entertains the belief that set theory is true.

Moreover, it seems clear that, because of the extremely counter-intuitive nature of the Banach-Tarski theorem, not only does 3a"" not follow from either 2a"" or 2b"", but that

3b'' Frank has the belief that set theory plus Banach-Tarski is true.

does not follow from (in conjunction with the truth of 1.'') either 2a'' or 2b''. After all, it seems odd to say that someone has a belief that a proposition (or set of propositions) is true while (staunchly) refusing to assent to its truth. According to Field, however, this is precisely what happens for nearly everyone (not knowing Banach-Tarski to be a theorem) for whom the theorem is formulated. Hence, it seems that while 3b'' does follow from the truth 1." and 2.", it is not, in general, true that the truth of 'A believes (has the belief) that p' follows from the truth of both 'a Eq. b' and 'A believes (has or entertains the belief) that q', where 'p' is the same as 'q' except for the replacement of occurrences of 'a' in 'q' by 'b' in 'p'. Perhaps another "type" of belief could be found which would rescue us from the case of 3.''; if so, I do not see it.

Now the presumption of the above was that by discovering a way to "fix up" 2." in an appropriate manner the problems that propositional attitude contexts created for (ET) could be solved. However, with the apparant failure of this attempted fix and no other intuitive resolutions to the problem available (to me, at any rate), it seems as if a re-examination of (ER), with respect to the specification of R, is in order. In other words, rather than attempting to delineate finer and finer shades of belief, R can be respecified as a relation stronger than identity of denotation. Such a move is, in effect, tantamount to claiming that the contradiction (ET) seemingly gives rise to is due to an improper specification of the equivalence relation that relates the two singular terms. Indeed, if the scope of (ET) is universal, then the failure of the first attempt at a fix together with the lack of any intuitive means of delineating finer shades of beliefs seems to naturally lead to the attempt to specify R.

Now, as has already been noted, there are a number of possible (re)specifications of R. Indeed, for anyone familiar with the (more philosophical) writings of Frege, the failure of (ER) should not have been wholly unexpected. It is a problem Frege recognized and attempted to meet in "On Sense and Meaning".<sup>13</sup> In this work Frege claims that in order to account for a posteriori identities one must refer not only to a singular term's referent, but to its sense as well. To use Frege's own example, the identification of the referent of 'the morning star' with the referent of 'the evening star' is informative precisely because there is a sense to 'the morning star' which differs from the sense of 'the evening star'. In contrast, in an priori identity such as 'the morning star is the morning star', both the sense and the reference of the

referring expressions are the same. In as much as identity is, in Frege's account, a function of both sense and reference, it follows that to claim that 'a is equivalent to b' is true just in case 'a' and 'b' both denote the same object and have the same sense is stronger than the claim made by (ER). Indeed, as Holland notes, identity of sense (and so too, reference) is so strong that it is a sufficient condition "to license substitution in indirect speech."<sup>14</sup> Hence, we can strengthen (ER) by the following respecification:

(ER)' '...has the same sense as...'

Moreover, because, according to Frege, the function of a sign's sense is to identify (pick out) that sign's reference (if it has one), (ER)' can be reworded as:

(ER)" '...identifies its referent in the same way as...'

If (ER)" seems intuitively inadequate as a specification of R, it is because of the ambiguity attached to the relation of identification utilized in (ER)". While the relation of denoting seems to be clearly defined, the relation of identification does not.<sup>15</sup> For instance, do 'unmarried male' and 'bachelor' have the same sense?; i.e., does 'unmarried male' identify its referent (a class I take it) in the same way as 'bachelor'? The answer to this question, for Frege at least, seems to rest with Dummett's observation that:

The sense of an expression is...that part of its meaning which is relevant to the determination of the truth-value of sentences in which the expression occurs.<sup>16</sup>

In other words, any two expressions, for example 'unmarried male' and 'bachelor', have the same sense (i.e., identify their referents in the same way) only if the first expression may uniformly replace every occurrence of the second expression in every sentence of the language in which the second expression occurs without altering the truth-value of that sentence, and visa versa. If Dummett is correct in his interpretation of Frege (and I am correct in my interpretation of Dummett), then (ER)" not only characterizes Frege's views on identity but also Mates' definition of synonymy.<sup>17</sup> In this case (ER)" can be rewritten as:

(ER)'" '...is synonymous (in a Language L) with...'

The effect of adopting (ER)'" is, in the case of 1." - 3.", to make 1." false. That is, while it may be true that 'John' and 'the Butler' both have the same

referent, it is pretty clearly not true that 'John' is synonymous with 'the butler' (in English, at any rate). If 1." is false, however, then it is no longer necessary that 3." be true. As far as it goes, this is precisely what we wanted. The contradiction which seemed to follow from adopting (ER) is avoided. Notice, however, that if the adoption of (ER)"" means that 1." is no longer true, then it follows that 1.' is no longer true either. Hence, not only is it no longer necessary that 3." be true, it is also not necessary that 3.' be true. What this suggests, I think, is that while (ER)"" does (apparently) solve the problem of propositional attitudes, the price of the solution (for English, at any rate) may well be too high. It appears as if the specification of R in (ER)"" is so strong that no two (orthographically) different singular terms are synonymous; at least this seems to be the moral of the set theory plus Banach-Tarski example. Not only is this counter-intuitive, but, more importantly, it suggests an even deeper problem with (ER)"". The adoption of (ER)"" as a specification of R means that the truth or falsity of the sentence 'a is equivalent to b' depends on the purely syntactical matter of whether 'F(b)' is or is not (respectively) entailed by 'F(a)' for any predicate F in the language. Recall, however, that the function of specifying the equivalence condition in a particular way was to both justify and place some constraint on the entailment of 'F(b)' from 'F(a)' for any predicate F in the language. In other words, while (ET) says that one can derive 'F(b)' if and only if both 'F(a)' and 'a is equivalent to b' are true, (ER)"", the specification of the equivalence relation binding 'a' and 'b' together, says that 'a is equivalent to b' if and only if 'F(b)' is entailed by 'F(a)' for all predicates F in the (object) language. The problem, of course, is that on this account, (ET) is attempting to set conditions on 'F(b)'s being entailed by 'F(a)' and 'a is equivalent to b' only by presupposing that the equivalence of 'a' and 'b' is a function of 'F(b)' 's being entailed by 'F(a)'. Given this, it seems clear that (ER)"" could not be a correct specification of R.

There are, of course, different ways of trying to fix-up (ER) that do not (obviously) involve the circularity of (ER)"". One way, suggested (though later rejected) by Mates<sup>18</sup> in analyzing the notion of synonymy, is to weaken the constraint on synonymy to something like Carnap's intensional isomorphism. In this case (ER) becomes:

(ER)"" '...is intensionally isomorphic to...'

According to Mates, the problem with (ER)"" is there are counter-examples to the weakened constraint which

undercut its use as a solution to the propositional attitude problem. Mates' remarks:

...let "D" and "D'" be abbreviations for two intensionally isomorphic sentences. Then [because "D" and "D'" are intensionally isomorphic] the following sentences are also intensionally isomorphic:

- (14) Whoever believes that D believes that D.
- (15) Whoever believes that D believes that D'.

But nobody doubts that whoever believes that D believes that D. Therefore, nobody doubts that whoever believes that D believes that D'. This seems to suggest that, for any pair of intensionally isomorphic sentences--let them be abbreviated by "D" and "D'",--if anyone even doubts that whoever believes that D believes that D', then Carnap's explication is incorrect.<sup>19</sup>

Mates' criticism can be illustrated by considering one of Carnap's own examples. In Meaning and Necessity Carnap claims that the expressions, ' $5 > 3$ ' and ' $\text{Gr}(v, III)$ ' are intensionally isomorphic. In this case, though, it follows that "whoever believes that  $5 > 3$  believes that  $5 > 3$ " is intensionally isomorphic to "whoever believes that  $5 > 3$  believes that  $\text{Gr}(v, III)$ ". However, it is easy to imagine someone (say a 1st or 2nd grader) who does not doubt that "whoever believes that  $5 > 3$  believes that  $5 > 3$ " but does doubt (perhaps because they think that for the Romans 5 was not greater than 3) that "whoever believes that  $5 > 3$  believes that  $\text{Gr}(v, III)$ ". It follows that the truth value of "whoever believes that  $5 > 3$  believes that  $5 > 3$ " may be different than that of "whoever believes that  $5 > 3$  believes that  $\text{Gr}(v, III)$ ". Hence, the two sentences cannot be intensionally isomorphic. If this counter-example is sound, then it is clear that intensional isomorphism is not strong enough as a specification for (ER). Of course one may claim that ' $5 > 3$ ' is not intensionally isomorphic to ' $\text{Gr}(v, III)$ ', but if they are not then it is difficult to see what would count as a genuine instance of intensional isomorphism. Surely, in other words, ' $5 > 3$ ' and ' $\text{Gr}(v, III)$ ' are L-equivalent since ' $(5 > 3) = (\text{Gr}(v, III))$ ' is L-true in English. In fact, the example using ' $5 > 3$ ' and ' $\text{Gr}(v, III)$ ' seems to meet the even more rigorous standard Putnam proposed in 1954<sup>20</sup> as a modification of Carnap's intensional isomorphism.<sup>21</sup> Now while the above counterexample used (for the sake of simplicity) the sentences ' $5 > 3$ ' and ' $\text{Gr}(v, III)$ ', it can be reproduced at the level of singular terms since, for Carnap, the concepts of L-equivalence and intensional isomorphism also apply to singular terms.



Hence, it follows that (ER)"" is inadequate as a specification of R.

Not surprisingly, Putnam is not alone in thinking that Mates' criticisms of (ER)"" can be met by an appropriate reconstrual of intensional isomorphism. Probably<sup>22</sup> the other two most detailed responses to Mates have come from Church and Sellars. In Sellars' case,<sup>23</sup> the response to Mates amounts to the claim that one must distinguish two different uses of expressions in oblique belief contexts: (i) a pure using sense, and (ii) an autonomous self-referential use. That is, there is a distinction between the occurrence of 'Gr(v,III)' in 'John believes that whoever believes that 5>3 believes that Gr(v,III)' and the occurrence of 'Gr(v,III)' in "The sentence 'whoever believes that 5>3 believes that Gr(v,III)' expresses something John believes" since, in this latter occurrence, 'Gr(v,III)' is part of a sentence which is to be understood as John uses it. Sellars' point is that Mates' counterexamples turn on a systematic confusion of these two contexts. As Burge notes, however:

The trouble with Sellars' rebuttal to Mates' argument is that it is too pat. [The one believing] may not have any particular counterexample to the substitution in mind. His doubt may be the result of a perfectly general caution about substitution in belief contexts.<sup>24</sup>

In addition, it is not intuitively obvious that Mates' counterexamples apply only to direct-discourse occurrences of D and D'. In Church's case,<sup>25</sup> Mates' counterexamples are met by (in contrast to Putnam) strengthening Carnap's intensional isomorphism to synonymous isomorphism. That is, the condition of L-equivalence, necessary for isomorphism, is replaced by synonymy. However, as both Scheffler<sup>26</sup> and Burge point out, there is reason to believe that this attempted answer to Mates' problem is, like Sellars' insufficient. Moreover, the problems with synonymy that arose in connection with (ER)"" would, ostensibly, arise here. That is, what does it mean to say that the two singular-terms are synonymous?

Although it may begin to appear as if the search for an adequate specification of R is like the search for the Holy Grail, there are still alternatives not yet considered. We can, for instance, follow the lead of philosophers such as Hintikka and suppose that the proper specification for (ER) depends upon the appropriate adoption of a possible world semantics/ontology. Concerning this, Hintikka says:

My basic assumption (slightly oversimplified) is that an attribution of any propositional attitude to the person in question involves a division of

all the possible worlds (more precisely, all the possible worlds which we can distinguish in the part of the language we use in making the attribution) into two classes: into those possible worlds which are [doxastically] in accordance with the attitude in question and into those which are [doxastically] incompatible with it.<sup>27</sup>

This means that any sentence of the form 'a believes that p' must be translated into 'in all possible worlds [doxastically] compatible with what a believes, it is the case that p'. The ramifications of this move are many: not only must (ER) be appropriately modified by an explicit reference to possible worlds, but in addition, the notion of what it means to be a singular term must be nailed down. In particular, since belief is relativized to possible worlds, the referent of any genuine singular term about which there is some belief (e.g., John in 2.) cannot simply be some individual object in the actual world. Rather, Hintikka (and other possible world theorists) must, in some sense, distinguish actual from possible (albeit real) objects. The point is that while two singular terms 'a' and 'b' may, from the point of view of possible world semantics, refer to the same individual in the actual world (i.e.,  $\phi(a, \mu_0) = \phi(b, \mu_0)$  where  $\mu_0$  denotes/designates the actual world), it is nonetheless possible that in some possible world  $\mu_1$ , 'a' and 'b' refer to different individuals (i.e.,  $\phi(a, \mu_1) \neq \phi(b, \mu_1)$ ). This "refinement" is reflected in Hintikka's definition of what it means to say that an instance of 'a believes that p' is true, viz.:

'a believes that p' is true in a possible world  $\mu_0$  if and only if p is true in every member of  $\phi_B(a, \mu_0)$ , where  $\phi_B(a, \mu)$  is the set of all possible worlds doxastically compatible with a's belief, and  $\phi_B$  is the two-argument function of referentially opaque belief.<sup>28</sup>

The upshot is that Hintikka, by strengthening the conditions under which two singular referring expressions are truly identical, has undercut the entailment of 3." from 1." and 2." by showing 1." (with 'Eq.' reconstructed along Hintikka's lines) to be false. In order for the inference from 1." and 2." to 3." to work, 1." would have to have been:

- 1." In all possible worlds, it is the case that the referent of 'John' is the same as the referent of 'the butler'. For all  $\mu \in \Omega$ ,  $\phi('John', \mu) = \phi('the butler', \mu)$ .

Now it may be that the adoption of a possible world semantics (though perhaps not Hintikka's<sup>29</sup>) really does

"save" (ET). Certainly W. Lycan, who holds a view of possible worlds similar to Plantinga, thinks it does. Lycan says:

By positing nonactual worlds as states of affairs, we may achieve our familiar but still remarkable reduction of alethic modalities to quantifiers, formulate Tarski-style semantics for propositional attitudes and hosts of other troublesome constructions...<sup>30</sup>

But even if the adoption of a possible world semantics does insure the universality of (ET)'s application (and I am inclined to think it does), it is only at a certain cost. For instance, if one's conception of possible worlds is like David Lewis', then possible worlds are viewed as "...respectable entities in their own right."<sup>31</sup> This means that accepting Lewis' view of possible worlds requires that a certain amount of "ontological baggage" (Lewis' "Ontological Slum" in Haack's words<sup>32</sup>) must also be countenanced in exchange for "saving" (ET). If, in contrast to Lewis, one thinks of possible worlds along the lines of Kripke, i.e., as purely formal mathematical model structures, then while one need not, perhaps, worry about any "extra" ontological entities, there are still certain metaphysical assumptions that one seems forced to accept. Specifically, essentialism seems to unavoidably follow from Kripke's definition of a quantificational model requiring individuals being identifiable across possible worlds.<sup>33</sup> Thus, the one reformulation of (ER) which seems to offer the promise of a solution to the problem of propositional attitude contexts requires either a considerable widening of one's ontology, or an acceptance of metaphysical "doctrines" such as essentialism, or both. This suggests that there is a (perhaps not so subtle) dependency of the applicability of (ET) on prior metaphysical and ontological commitments.

But now let us pause. What I have tried to make clear in the above is that even those logics traditionally called intensional, Frege's and Hintikka's for example, are really just attempts to save extensionalism. Indeed, it seems clear that if an intensional logic is supposed to differ from an extensional logic, ostensibly by either denying (ET) or offering some alternative to (ET), then so-called intensional logics such as Frege's and Hintikka's are simply not intensional at all.<sup>34</sup> This differs from Marcus' view that "A language is explicitly intensional to the degree to which it does not equate the identity relation with some weaker form of equivalence."<sup>35</sup> The whole point is that the equivalence relation simpliciter, Eq., is uninterpreted. Are we, for instance, interested in equivalence qua sameness of reference in

the actual world or in equivalence qua something else. It seems to me that, in contrast to the traditional view which claims that a context is intensional just in case the salva veritate substitutibility of co-referring expressions does not universally hold,<sup>36</sup> a better way to view the extensional/intensional distinction is to adopt Marcus' own view that extensionality comes in degrees. In this case, a (traditionally) intensional language is simply a more restrictive (with regards to the specification of R) extensional language.

This suggests, I think, that the impetus for adopting a possible world semantics lies with the demand that (ET)'s applicability be universal. However, the unwillingness of many to countenance any talk of possible worlds raises an important question which, up to this point, has been lurking in the background, viz., "Why should we feel obligated to continue to attempt to patch up (ET) in such a way that (ET) should retain its universal applicability for the language of the language user?". In other words, why not simply stop with the original specification of (ER) and say that (ET)'s applicability does not range over the entire language but simply over some well-defined portion of it? In a way this seems to be precisely what Quine does in Word and Object. Here Quine is apparently unwilling to countenance either the individuals (e.g., intensions) or the metaphysics necessary to save the universal applicability of (ET) and so simply restricts its applicability. Indeed, this seems to be precisely the position of Quine's that Marcus had in mind when she says that in Quine's view, propositional attitude contexts are "consigned to a shelf labeled 'referential opacity'...and are disposed of."<sup>37</sup> Of course this does not really settle the issue. How we are to decide upon the range of (ET)'s applicability lies, after all, in an answer to the question "Which is most important, ontological/metaphysical austerity or universal applicability of (ET)?" To answer this, however, presupposes that we understand the motivation for holding (ET) (whatever the limits of its applicability) in the first place. Accordingly, it is to this issue that I at last turn.

The suggestion of this paper is that the reason there has been such a reluctance to either limit the applicability of (ET) or reject it entirely rests with the crucial role it is deemed to play in any attempt to provide a theory of meaning. Now by a theory of meaning I mean, at the very least, a theory which will, in some appropriate sense, do two jobs: first, it will "give the meaning" of each sentence (of the object language), and second, it will give an account of how the meanings of the sentences (of the object language) depend upon the meanings of the words that compose them.<sup>38</sup> In the simplest case, what the first requirement on a theory of meaning amounts to is that any ade-

quate theory of meaning must generate (i.e., entail) an instance of the schema

(M) 's means that p'

for each sentence in the object language. Given this as a response to the first demand of a theory of meaning, what now of the second? For reasons given by Davidson,<sup>39</sup> the appeal to meanings of sentences (as entities, e.g. Facts) is of no help in solving the second of the dual tasks required of a theory of meaning. Hence, 'p', as it occurs in the M-schema, cannot be replaced by the name of some entity/thing. Accordingly, Davidson suggests that the natural move to make is to imagine 'p', in (M), replaced by a sentence (in the metalanguage). Since sentences prefixed with 'that' don't name facts, (unless we choose to allow them to), the problem now changes and becomes that of understanding precisely how the (apparently intensional) 'means that', as it occurs in (M), is to be construed so as to allow the second requirement of a theory of meaning to be met.

If we look backwards in the history of philosophy to Wittgenstein (of the *Tractatus*), we find the view expressed that 'means that' means something like 'expresses agreement and disagreement with the possibilities of combinations of the existence and non-existence of atomic facts'. This rather long winded explanation, however, simply means that it is the truth conditions of a sentence we are asking for when we ask for its meaning. Thus, to say in the Wittgensteinian vernacular that a sentence  $S$  has a particular meaning is to say that the sentence  $S$  referentially pictures a particular combination of the existence and non-existence of the elements of some subset of the set of all atomic facts. But in this case, any two sentences,  $S_1$  and  $S_2$  have identical meaning just in case they referentially picture the same particular combination of the existence and non-existence of the elements of some subset of the set of all atomic facts. Putting the point differently, if  $S_1$  referentially pictures  $\langle a_1, a_2, \dots, a_n \rangle$  where  $\{a_1, a_2, \dots, a_n\}$  is a subset of all atomic facts, then the sentence  $S_2$  means  $S_1$  if and only if  $S_2$  also referentially pictures  $\langle a_1, a_2, \dots, a_n \rangle$ . As suggested above, however, this amounts to the claim that  $S_1$  has the same meaning as  $S_2$  if and only if the truth-conditions of  $S_1$  are the same as the truth-conditions of  $S_2$ . It is important to notice here that we already have an implicit acceptance of (ET) operating in the theory of meaning. In particular, by saying that  $S_1$  means  $S_2$  just in case they referentially picture a particular combination of the existence and non-existence of the elements of the same subset of the set of all atomic facts, one is implicitly saying that if both  $S_1$  is a sentential component of some larger

(molecular) sentence SENT and  $S_1$  means  $S_2$ , then substitution of  $S_2$  for  $S_1$  in SENT will not alter the meaning of SENT. After all, SENT will change meaning only if it referentially pictures a different combination of the existence and non-existence of the elements of some subset of the set of all atomic facts. The substitution of  $S_2$  for  $S_1$  in SENT, however, cannot cause any alteration in SENT's referential picturing since the particular combination of the existence and non-existence of the elements of some subset of the set of all atomic facts is precisely the same for both  $S_1$  and  $S_2$ . Wittgenstein, however, goes further with his analysis and claims both that corresponding to every atomic fact there is an atomic proposition (an atomic sentence if you will) which pictures it and every (meaningful) atomic proposition pictures an atomic fact. Hence, if the sentence  $S_1$  is a molecular sentence, then the manner in which one determines what the truth-conditions of that sentence are is by uncovering its deep logical form; i.e., by determining both the constituent atomic propositions of  $S_1$  and their relative order. Accordingly, if the sentences that make up SENT are themselves both meaningful and molecular, then they must, in principle, be able to be broken down into their constituent sentences. Indeed the process only stops when the analysis encounters nothing but atomic sentences. Hence, SENT is nothing but a particular ordering of atomic sentences. This accords with Dummett's claim that a theory of meaning requires a molecular (in his sense, not Wittgenstein's) conception of language.<sup>40</sup> Moreover, there is a curious sort of parallel between this view and the methodology Davidson employs in coming to understand (interpret) another language. For instance, Davidson says:

The process of devising a theory of truth for an unknown native tongue might in crude outline go as follows. First we look for the best way to fit our logic, to the extent required to get a theory satisfying Convention-T onto the language; this means reading the logical vocabulary of first order quantification theory (plus identity) into the language, not taking the logical constraints one by one, but treating this much of logic as a grid to be fitted onto the language in one fell swoop.<sup>41</sup>

Davidson's point is that a theory of meaning for a language presupposes that the language either has a readily observable logical structure or else must have one read into it. But why, we may ask, is this necessary? The answer seems to be that Davidson believes that it is by using a Tarski-type truth definition that one supplies a theory of meaning for sentences of the language. A Tarski-type truth definition is, however,

a definition which defines the truth of sentences by specifying the sentence's truth-conditions. Hence, Davidson's holism notwithstanding, the theory of meaning advocated by Davidson appears to be a sophisticated variant of a Wittgensteinian theory wherein the meaning of a sentence is given when the truth-conditions of that sentence, and all sentences built-up from it or its open-sentence equivalent, are given. In addition, for Tarski the truth-conditions of a sentence ultimately depend upon the notion of satisfaction, and, when satisfaction is used to provide a theory of meaning (as opposed to just a definition of truth as in Tarski), the satisfaction relation is a (realistic) correspondence relation.<sup>42</sup> In other words, the meaning of a non-atomic sentence is a function of both the satisfaction conditions of its atomic constituents and the (logical) order of its atomic constituents. But this means that the Davidsonian theory of meaning rests upon the same basic assumptions that Wittgenstein's theory did. In addition, it should be clear that both theories defend (ET) just because of their reliance upon the correspondence theory which grounds the truth-conditions and which, in turn, are utilized by both in constructing a theory of meaning.

To return to the view of Wittgenstein, which shows so well the presuppositions of more contemporary theories of meaning, we can formulate the following predecessor of (ET):

(ET)\* If ' $S_1$ ' and ' $P_1$ ' are both atomic propositions and ' $S_1$ ' and ' $P_1$ ' both referentially picture  $\langle a_1, a_2, \dots, a_n \rangle$ , then if ' $F(S_1)$ ' is true, then ' $F(P_1)$ ' is true.

As suggested above, the upshot of (ET)\* is that the definition of the meaning of a sentence as its truth-conditions demands that (ET) hold. But now suppose, paralleling the original problem with 2." and 3.", that ' $F(S_1)$ ' is true and ' $F(P_1)$ ' is false. If it is granted that this situation could occur, then clearly ' $F(S_1)$ ' cannot mean the same thing as ' $F(P_1)$ ' since meaning is defined as identity of truth-conditions. But what has failed? It seems, as was pointed out in the discussion of (ET), that the only answer is that it is the specifications of ' $\dots$ meaning the same as...' as ' $\dots$ picturing the same state-of-affairs as...' which has failed.

This failure notwithstanding, it seems that the general conception that the second requirement of an adequate theory of meaning can be met by an appeal to truth-conditions has been so pervasive that nearly everyone in philosophy of language has either explicitly (e.g., Davidson, Quine) or implicitly (e.g., Hintikka) adopted it. What I have tried to suggest, though, is that by its very nature, anyone who seriously adopts

such a presupposition about a theory of meaning is inevitably led to adopt (ET) as well. Hence, to give up the universality of (ET) for a language is tantamount, on this view, to giving up the quest for providing a single theory of meaning for that language. It may be, however, that it is not to give up the search for a theory of meaning for some proper subset of the set of sentences of a language. What I have in mind is a variation of Quine's view concerning the possibility of alternate translation manuals for a language; the suggestion being that a translation manual using (ET) might be appropriate for some limited portion of a language, but inappropriate for another.

In the final analysis, of course, it should be recognized that the theory of meaning Wittgenstein so well epitomizes rests upon a particular view concerning the relation of the world and language. While there is a temptation to see (only) the correspondance theory of truth behind the adoption of (ET), one must also ask why one would, in the first place, adopt a correspondance theory. A possible answer to this is that it is the acceptance of a realist metaphysics/ontology which motivates the acceptance of a correspondance theory. In other words, it seems that, at the deepest level, it is an acceptance of a realist metaphysics which leads to the adoption of (ET). Indeed, this seems to accord well with the anti-realist's reluctance to grant (ET) universal applicability. That is, while (ET) may hold for that set of sentences which are either true or false, it does not hold for that set of sentences which are neither true nor false.<sup>43</sup> Hence, questions about the applicability of (ET) are, perhaps, best deferred until the knotty metaphysical/ontological problems of realism are dealt with.



## NOTES

<sup>1</sup>J. A. Fodor, The Language of Thought (New York: Harvard University Press, 1978), p. 9.

<sup>2</sup>See R. Carnap, The Logical Syntax of Language (New York: Harcourt, Brace, 1937), p. 320; also see J. Cornman, "Intentionality and Intensionality," in Intentionality, Mind and Language, ed. by A. Marras (Urbana, IL: University of Illinois Press, 1972), p. 55ff. For a comparison to the methodologies underlying reductionist materialist theories of mind see Materialism and the Mind-Body Problem, ed. D. Rosenthal (Englewood Cliffs, NJ: Prentice-Hall, Inc., 1971), pp. 1-17. For a comparison to reductionist theses in general see J. J. Katz, "Recent Criticisms of Intensionalism," in Minnesota Studies in the Philosophy of Science, Volume 7: Language Mind and Knowledge, ed. by K. Gunderson (Minneapolis: University of Minnesota Press, 1975), pp. 36-130.

<sup>3</sup>Radical in the sense that the scope is universal; i.e., all sentences either are, or can be translated into extensional sentences without loss of meaning.

<sup>4</sup>R. Carnap, Meaning and Necessity, 2nd ed. (Chicago: University of Chicago Press, 1956), p. 48.

<sup>5</sup>For Carnap's explanation of what a designative component of a sentence is, see Meaning and Necessity, pp. 6-7.

<sup>6</sup>For the use of quasi-quotes, see W.V. Quine, Mathematical Logic, revised edition (Cambridge, MA: Harvard University Press, 1976), pp. 33ff.

<sup>7</sup>cf. M. Dummett, Frege: Philosophy of Language, 2nd ed. (Cambridge, MA: Harvard University Press, 1981), pp. 207, 208. The major difference between the version provided in this paper and Dummett's is that where I use the expression 'a is equivalent to b', Dummett uses, what seems to me, the less perspicuous 'a is the same as b'. Also, see G. Harman, "Deep Structure as Logical Form," in Semantics of Natural Language, ed. by Donald Davidson and Gilbert Harman (Dordrecht: D. Reidel Publishing Co., 1972), pp. 35-38; I. Linsky, Referring (Atlantic Highlands, NJ: Humanities Press, 1980), esp. chapter 5. For a somewhat different formulation see G. Bergman, "Conditions for an Extensional Elementaristic Language," in The Metaphysics of Logical Positivism (Madison: University of Wisconsin Press, 1967), p. 232.

<sup>8</sup>See, e.g., W.V. Quine, "Logic as a Source of Syntactic Insights," and "The Scope and Language of

Science," both in The Ways of Paradox, revised ed. (Cambridge, MA: Harvard University Press, 1976), pp. 44-49. Also, W.V. Quine, Word and Object (Cambridge, MA: The MIT Press, 1960), esp. sections 19 and 20; and, P.F. Strawson, "Singular Terms and Predication," in Logico-Linguistic Papers (London: Methuen and Co., Ltd., 1974), pp. 53-79.

<sup>9</sup>See R. Marcus, "Extensionality," in Reference and Modality, ed. by L. Linsky (Oxford: Oxford University Press, 1977), p. 44 for a formulation of an extensionalist thesis similar to (ET). This suggests, I think, that (ET)'s formulation is genuinely uncontroversial. Also see A.N. Prior, Time and Modality (Oxford: Clarendon Press, 1957), pp. 70ff. For a slightly different formulation of the extensionalist position, see J. J. Katz, Language and Other Abstract Objects (Totowa, NJ: Rowman and Littlefield, 1981), pp. 134ff.

<sup>10</sup>B. Russell, An Inquiry into Meaning and Truth (New York: Norton, 1940), p. 210; Also, see Hans Reichenbach, Elements of Symbolic Logic (New York: Macmillan and Co., 1947), pp. 277ff.

<sup>11</sup>See, e.g., D. Dennett, "A Cure for the Common Code," in Brainstorms (Hassocks, Sussex: Harvester Press, 1978), p. 104.

<sup>12</sup>As formulated here, 1. " means that finitely axiomatized von Neumann set theory (together with the axiom of choice) and its theorems are the same as finitely axiomatized von Neumann set theory plus its theorems plus the Banach-Tarski theorem. Indeed, the point is that the Banach-Tarski theorem is a theorem of finitely axiomatized von Neumann set theory (with the axiom of choice). What the example trades on is the possibility that someone would be willing to accept (as true) von Neumann set theory (axioms and theorems) while denying (the truth of) the Banach-Tarski theorem (and so denying that the Banach-Tarski theorem is a theorem in von Neumann set theory). See David Lewis, "General Semantics," in Semantics of Natural Language, ed. by D. Davidson and G. Harman (Dordrecht: D. Reidel Publishing Co., 1972), pp. 182ff; H. Field, "Mental Representation," in Readings in Philosophy of Psychology, ed. by N. Block (Cambridge, MA: Harvard University Press, 1981), pp. 503ff.

<sup>13</sup>G. Frege, "On Sense and Meaning," in Translations from the Philosophical Writings of Gottlob Frege, ed. by Geach and Black (Totowa, NJ: Rowman and Littlefield, 1980), pp. 56-78.

<sup>14</sup>A. Holland, "Carnap on Frege on Indirect Reference," Analysis, v.38, n.1 (January 1978), p. 29.

<sup>15</sup>See M. Devitt, Designation (New York: Columbia University Press, 1981), pp. 7ff.

<sup>16</sup>M. Dummett, Frege: Philosophy of Language, p. 89.

<sup>17</sup>See B. Mates, "Synonymity," in Semantics and the Philosophy of Language, ed. by L. Linsky (Urbana, Illinois: The University of Illinois Press, 1952), pp. 111-138.

<sup>18</sup>Ibid., p. 125.

<sup>19</sup>Ibid., p. 125.

<sup>20</sup>H. Putnam, "Synonymity, and the Analysis of Belief Sentences," Analysis, v.14 (1954), pp. 114-119.

<sup>21</sup>As Barbara Partee points out, Carnap tries to get around the Mates-type counterexamples by narrowly redefining what will count as evidence for the presence of a belief. For the details, see her paper "The Semantics of Belief-Sentences," in Rudolf Carnap, Logical Empiricist, ed. by J. Hintikka (Dordrecht: D. Reidel Publishing Co., 1975), pp. 248ff.

<sup>22</sup>See T. Burge, "Belief and Synonymy," The Journal of Philosophy, v. 75, n. 3 (March, 1978), p. 120.

<sup>23</sup>W. Sellars, "Putnam on Synonymity and Belief," Analysis, v.15, n.5 (April, 1955), pp. 117-120.

<sup>24</sup>T. Burge, "Belief and Synonymy," p. 121.

<sup>25</sup>A. Church, "Intensional Isomorphism and Identity of Belief," Philosophical Studies, v.5, n.5 (October, 1954), pp. 65-73.

<sup>26</sup>I. Scheffler, "On Synonymy and Indirect Discourse," Philosophy of Science, v. 22, n. 1 (January, 1955), pp. 39, 40.

<sup>27</sup>J. Hintikka, "Semantics for Propositional Attitudes," in Philosophical Logic, ed. by Davis, Hockney, and Wilson (Dordrecht, D. Reidel Publishing Co., 1969), p. 25.

<sup>28</sup>Ibid., p. 28.

<sup>29</sup>For Hintikka there appears to be a problem when the identity (corresponding to 1.) is a necessary identity. Whether this problem (assuming it's genuine) is unique to Hintikka's modal theory of semantic struc-

tures is not wholly clear. Hintikka, for instance, seems to believe he can avoid the problem by distinguishing logically possible worlds from epistemically possible worlds. See, e.g., J. Hintikka, "Impossible Possible Worlds Vindicated," Journal of Philosophical Logic, v. 4, n. 4 (November, 1975), pp. 475-484. F. M. Katz and J. J. Katz seem to think that this problem is genuine and undercuts all modal theories of natural languages. See F. M. Katz and J. J. Katz, "Is Necessity the Mother of Intension?," The Philosophical Review, v. 86, n. 1 (January, 1977), pp. 70-90; esp. part III.

<sup>30</sup>W. Lycan, "The Trouble With Possible Worlds," in The Possible and the Actual, ed. by M. Loux (Ithaca, NY: Cornell University Press, 1979), p. 274.

<sup>31</sup>D. Lewis, "Possible Worlds," in The Possible and the Actual, p. 183.

<sup>32</sup>S. Haack, "Lewis' Ontological Slum," The Review of Metaphysics, v. 33 (1977), pp. 415-429.

<sup>33</sup>See L. Linsky, Names and Descriptions (Chicago: The University of Chicago Press, 1980), Chapter 7.

<sup>34</sup>For suggestive remarks in this direction see D. Davidson, "In Defense of Convention T," in Truth, Syntax and Modality ed. by L. Leblanc (Amsterdam: North-Holland, 1973), p. 77 ff.

<sup>35</sup>R. Marcus, "Modalities and Intensional Languages," in Contemporary Philosophical Logic, ed. by Copi and Gould (New York: St. Martin's Press, 1978), p. 258.

<sup>36</sup>See, e.g., R. Grossman, "Acts and Relations in Brentano," Analysis, v. 21, n.1 (October, 1960), p. 3.

<sup>37</sup>R. Marcus, "Modalities and Intensional Language," p. 260; In particular, in Word and Object (p. 151), Quine says:

I do not disallow failure of substitutivity, but only take it as evidence of non-referential position; nor do I envisage shifts of reference under opaque construction.

<sup>38</sup>See, for instance, D. Davidson "Truth and Meaning" in Readings in the Philosophy of Language, ed. by J. F. Rosenberg and C. Travis (Englewood Cliffs, N.J.: Prentice-Hall, 1971), p. 450.

<sup>39</sup>Ibid., pp. 452-454.

<sup>40</sup>See M. Dummett, "What is a Theory of Meaning?," in Mind and Language, ed. by S. Guttenplan (Oxford: Clarendon Press, 1975), p. 123; and, "What is a Theory of Meaning (II)?," in Truth and Meaning, ed. by G. Evans and S. McDowell (Oxford: Oxford University Press, 1977), p. 72.

<sup>41</sup>D. Davidson, "Radical Interpretation," Dialectica, v. 27, nos. 3-4 (1973), p. 323.

<sup>42</sup>These are, perhaps, not wholly uncontroversial claims. The entire issue seems to turn on precisely what is going to count as a correspondance theory of truth. For discussion of these matters see: H. Field, "Tarski's Theory of Truth," The Journal of Philosophy, v. 69, n. 13 (1972), pp. 347-375; H. Putnam, Meaning and Moral Sciences (London: Routledge and Kegan Paul, 1978), esp. lectures I and II; H. Keuth, "Tarski's Definition of Truth and the Correspondance Theory," Philosophy of Science, v. 45, n. 3 (September, 1978), pp. 420-430.

<sup>43</sup>Clearly more needs to be said. See, e.g., Colin McGinn, "Truth and Use," and Mark Platts, "Moral Reality and the End of Desire," both in Reference Truth and Reality (London: Routledge and Kegan Paul, 1980), pp. 19-40, 69-82 respectively. At the very least, I want to suggest that the desire to either do philosophy of language apart from metaphysics or reduce metaphysics to philosophy of language is misguided.