Review of Kit Fine, 'Model Theory for Modal Logic. Parts I-III'

(Journal of Philosophical Logic, 7 (1978), pp. 125-156, 277-306; and 10 (1981), pp. 293-307)

Saul A. Kripke

This is the published version of Kripke, S. A. (1985), 'Kit Fine. Model Theory for Modal Logic. Part I—the de re/de dicto distinction. *Journal of Philosophical Logic*, vol. 7, (1978), pp. 125-156. -Kit Fine. Model Theory for Modal Logic—Part II. The elimination of de re modality. *Journal of Philosophical Logic*, vol. 7 (1978), pp. 277-306. -Kit Fine. Model Theory for Modal Logic—Part III. Existence and Predication. *Journal of Philosophical Logic*, vol. 10 (1981), pp. 293-307.' *Journal of Symbolic Logic*, 50(4): 1083-1093, which can be obtained from the publisher at https://doi.org/10.2307/2274002. It is reproduced here by permission of the Association for Symbolic Logic which holds the copyright.

© Association for Symbolic Logic



Review

Reviewed Work(s): Model Theory for Modal Logic. Part I--the de Re/De Dicto Distinction by Kit Fine; Model Theory for Modal Logic--Part II. The Elimination of de re Modality by Kit Fine; Model Theory for Modal Logic--Part III. Existence and Predication by Kit Fine

Review by: Saul A. Kripke

Source: The Journal of Symbolic Logic, Vol. 50, No. 4 (Dec., 1985), pp. 1083-1093

Published by: Association for Symbolic Logic

Stable URL: https://www.jstor.org/stable/2274002

Accessed: 12-04-2019 16:00 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms



 $Association\ for\ Symbolic\ Logic\ is\ collaborating\ with\ JSTOR\ to\ digitize,\ preserve\ and\ extend\ access\ to\ The\ Journal\ of\ Symbolic\ Logic$

 $d(\alpha, \beta)$ in this model, given in terms of the formulas $A_i(\varepsilon)$ and $B_j(\delta)$. Now if we choose any valuation on CODE(\mathcal{F}), and $D(\varepsilon, \delta)$ is falsified at a point x, then x must have a "point of view" similar to $d(\varepsilon, \delta)$; to be exact, x = d(u, v), where $(\alpha \varepsilon, \delta \beta)$ is derivable from (u, v) in \mathcal{F} , for some α and β . Now if (ε', δ') is derivable from (ε, δ) in \mathcal{F} then $x < a(\alpha \varepsilon', \delta'\beta)$, so we can find points in the canvas that show $D(\varepsilon', \delta')$ false at x. Thus $L(\mathcal{F})$ is validated by CODE(\mathcal{F}). Conversely, any formula $D(\alpha, \beta) \supset D(\varepsilon, \delta)$ where (α, β) is not derivable from (ε, δ) is refutable in the canonical model, so the undecidability result follows. The author shows admirable ingenuity in solving a long-standing problem.

Popov's paper, by contrast with Šehtman's, is long and crammed full of idiosyncratic notation and techniques. The author employs a kind of normal form theorem for a calculus somewhat resembling a natural deduction system, but the details are so peculiar that it is difficult in many places even to guess at the author's intentions. Like the previous author, Popov starts from a system of semi-Thue productions. The logic defined contains an axiom scheme for each production, together with an added scheme representing a fixed word V_0 . In his main theorem the author claims that a word V is derivable from V_0 by using the productions if and only if a certain formula is a theorem of the logic (p. 455). However, the proof seems to contain an error. Let us suppose that the word V_0 is the single letter a_1 (clearly, there is no loss of generality in supposing this). The axiom scheme representing the word V_0 is built up from formulas $\tilde{Z}_6, \dots, \tilde{Z}_{16}$ (pp. 444-445). According to the author's definitions we have $\tilde{Z}_6 = (X_7^{m+1} \vee X_6^{m+1}), \ \tilde{Z}_7 = (X_7^{m+1} \supset X_6^{m+1}), \ \tilde{Z}_8 = (X_9^{m+1} \vee X_8^{m+1}), \ \tilde{Z}_9 = (X_9^{m+1} \supset X_8^{m+1}), \ \tilde{Z}_{10} = X_{10}^{m+1}, \ \tilde{Z}_{11} = X_{11}^{m+1}, \ \tilde{Z}_{12} = X_{12}^{m+1}, \ \tilde{Z}_{13} = X_{13}^{m+1} \times X_{15}^{m+1}, \ \tilde{Z}_{14} = X_{14}^{m+1}, \ \tilde{Z}_{15} = X_{15}^{m+1}, \ \tilde{Z}_{16} = X_{16}^{m+1}, \text{ where each of the expressions } X_8^x \text{ is a variable. It follows that by substituting in this axiom scheme, we can deduce the corresponding axiom scheme for an arbitrary word <math>V$. This contradicts the author's theorem.

The following additional corrections should be noted for Šehtman's paper. In the second paragraph on page 656, the third occurrence of T_2 should be replaced by R_2 . In the proof of Lemma 7 on page 659, the variable λ' should be replaced by δ' .

ALASDAIR URQUHART

KIT FINE. Model theory for modal logic. Part I—the de re/de dicto distinction. **Journal of philosophical logic**, vol. 7 (1978), pp. 125–156.

KIT FINE. Model theory for modal logic—part II. The elimination of de re modality. Ibid., pp. 277-306.

KIT FINE. Model theory for modal logic—part III. Existence and predication. Ibid., vol. 10 (1981), pp. 293–307.

The author's interesting project is to prove *philosophically* significant theorems about modal logic similar to the preservation theorems of classical model theory. Certain philosophical positions demand that only those sentences preserved under certain mappings are meaningful. The problem is to characterize the sentences preserved as those equivalent to sentences in a certain syntactically defined class, as Tarski characterized the sentences preserved under substructures as those equivalent to universal sentences. Often the syntactically defined class turns out to be independently motivated in terms of the philosophical position in question. Sometimes the author includes interesting technical results naturally suggested by the mathematics, but not necessarily by the philosophical position.

The usual terminology of (the reviewer's semantics for) modal logic will be used. A (quantificational) frame (a model structure) is a quadruple (G, K, R, ψ) , with K a set of possible worlds, G a distinguished element of K (the real world), R a binary relation on K, and ψ a function assigning a set (domain) to each element of K. Let U be the union of all the domains. A model based on (G, K, R, ψ) (or, in the author's terminology, a "modal structure") assigns a subset of U^n , $\phi(P_i, H)$, to each n-ary predicate letter P_i of a given language and each H in K. Given a modal structure, there is a natural classical (first-order) structure $\phi(H)$ associated with each world H, obtained by restricting the relations ϕ associates with each P_i in H to $\psi(H)$. (In a natural abuse of terminology, this structure can be identified with the world itself.) The entire modal structure can also be viewed as a classical (first-order) structure. Ordinarily the author is concerned only with S5; then R can be dropped, since we can assume that all worlds are related to each other. It will be assumed that we are dealing with S5 in the sequel unless the contrary is specified.

Anti-Haecceitism holds that all identifications of individuals across possible worlds are arbitrary (meaningless). Call two modal structures ϕ_1 and ϕ_2 based on the same frame locally isomorphic iff for each H in K, $\phi_1(H)$ and $\phi_2(H)$ are isomorphic structures. The anti-Haecceitist cannot accept distinctions between locally isomorphic modal structures—only those sentences that are preserved under local isomorphism can be meaningful. Two modal structures are weakly locally isomorphic iff their real worlds are isomorphic (as classical structures) and exactly the same isomorphism types are realized on worlds of

the one as on worlds of the other. Essentially, locally isomorphic structures are weakly isomorphic structures that realize each isomorphism type the same number of times. It is easy to show that sentences preserved under local isomorphism are preserved under weak local isomorphism.

On the other hand, Quine, apparently basing himself on anti-Haecceitist premisses, objected to "quantifying in" and to "essentialism." Quine's strictures are violated whenever a sentence has a well-formed part of the form $\Box A$, where A contains a free variable. (Assume the language has no constants.) On this basis, Quine condemned all quantified modal logic; but of course many sentences of quantified modal logic—the "de dicto" sentences—are free of the alleged problem. Clearly de dicto sentences are preserved under local isomorphism (observed by Pavel Tichy, Journal of philosophical logic, vol. 2 (1973), pp. 387–392, and others). The author proves, conversely, that every sentence preserved under local isomorphism is equivalent to a de dicto sentence. Actually, he proves the result even if attention is confined to models satisfying a given modal theory T. The author states that J. Broido (unpublished dissertation, University of Pittsburgh, 1974) had also proved the result for theories invariant under local isomorphism (which obviously includes the empty theory). The philosophical moral is that the sentences involving de re modality are precisely those that ought to be objectionable to the Haecceitist. The largely weaker (Broido) theorem requires much less model-theoretic technique than the author's version (see below).

The following sketches a short proof of the result by saturated models. If A is not equivalent (in all models satisfying T) to a fixed de dicto sentence, then by a standard argument using two applications of the compactness theorem, there are two countable models ϕ_1 and ϕ_2 satisfying T that satisfy the same de dicto sentences (in the real world) such that A is true in (the real world of) ϕ_1 but false in (the real world of) ϕ_2 . Both of these models are elementary subsystems (as classical structures) of corresponding saturated models of the same uncountable cardinality κ —call these ϕ_1^* and ϕ_2^* . Then ϕ_1^* and ϕ_2^* (considered as modal structures) still satisfy T and the same de dicto sentences (in the real world), while A is true in (the real world of) ϕ_1^* and false in (the real world of) ϕ_2^* . Now in both ϕ_1^* and ϕ_2^* , by saturation, a complete classical first-order theory Δ is realized in some world H of the model iff, for every finite subset $\{A_1,\ldots,A_n\}$ of A, $\{A_1 \land \cdots \land A_n\}$ is true in the real world of the model. Since $\{A_1 \land \cdots \land A_n\}$ is always de dicto, and the real worlds of ϕ_1^* and ϕ_2^* satisfy the same de dicto sentences, this means that Δ is realized in some world of ϕ_1^* iff it is realized in some world of ϕ_2^* . Hence, precisely the same elementary (classical) theories are realized in worlds of ϕ_1^* as in worlds of ϕ_2^* . Further, two classical structures that are elementarily equivalent and are associated with worlds of either model necessarily either have the same finite cardinality or are saturated with the same infinite cardinality κ . In either case, such structures will be isomorphic. These observations show that ϕ_1^* and ϕ_2^* are weakly locally isomorphic. Therefore A is not preserved under weak local isomorphism and hence is not preserved under local isomorphism. (Readers who prefer to use recursively saturated model pairs can modify the preceding proof accordingly. The author uses a different method, taking the union of an elementary chain. This technique invokes more elementary machinery, but gives a somewhat more cumbersome proof.)

The method just given can be used to prove various stronger statements: for example, if C is any Δ -elementary class of modal structures (the class of those satisfying a modal theory T is a special case), then a sentence is preserved under weak local isomorphism of structures in C iff it is equivalent (for structures in C) to a *de dicto* sentence. But here, unlike the special case where C is the class of structures satisfying a given modal theory T, 'weak local isomorphism' cannot be replaced by 'local isomorphism.'

This theorem (and any techniques likely to prove it) indeed has exactly the flavor of the preservation theorems of classical model theory. Nevertheless, although the fact that the theorem holds for arbitrary modal theories T is technically interesting, the reviewer finds it difficult to appreciate the general result in terms of the anti-Haecceitist motivation. If the axioms of T are themselves not preserved under local isomorphism, and hence are "meaningless" for the Haecceitist, what significance can be attributed to a theorem about structures satisfying T? It would seem that the philosophically significant result is the weaker one restricted to theories that are themselves preserved under local isomorphism; but then the much easier technique of Broido, which does not use methods with the flavor of classical model theory (see below), is quite adequate. The reviewer is disappointed that the connection between model theory and philosophy is not as strong as one might have hoped. But the reviewer also feels that technical and mathematical motivations should not be dismissed; see the last paragraph of this review.

The preceding gives the main result of the first paper. Aside from the completeness proofs to be discussed below, the rest of the paper discusses refinements. The author considers what happens when constants are allowed in the language; he states that the main result, properly formulated, remains valid

for the usual normal modal logics weaker than S5 (S4, K, T, etc.); he states that the results still hold for "possibilist" (outer) quantifiers. (He also says that the results still hold for quantifiers over individual concepts, but to understand this claim one must realize that quantifiers over an arbitrary family of individual concepts are intended, in which case the claim does not differ much from the case where the quantifiers are over individuals. The claim does not apply to the case of quantifiers over all individual concepts.) Another result states that a sentence is preserved (for models of T) under isomorphisms of the actual world alone iff T implies that it is materially equivalent to a fixed sentence without modal operators. It seems to the reviewer that all these results can be obtained by the method of saturated models above. The author gives an interesting hierarchy of modal and de re complexity of formulas. He also gives results characterizing preservation under local isomorphism for a purely classical language (of modal structures) (with quantification over worlds).

The second paper discusses means by which a "soft" de re skeptic might justify the full language of quantified modal logic after all by interpreting de re sentences as "equivalent" to de dicto ones. He does this, not by considering a direct translation, but by considering an extension L of quantified S5 (or quantified S5 with constant domain—"S5B") with additional axioms. The axioms of L will imply that every sentence (or even every formula) is equivalent to a de dicto sentence (or formula); the author calls this "sentence eliminability" (or "formula eliminability"). The extended system is to be conservative over the original system, as far as de dicto theorems are concerned (but not for arbitrary theorems). The additional axioms are themselves not de dicto; hence the "soft" de re skeptic justifies them as "meaningless" devices used to obtain the translation of arbitrary sentences into de dicto equivalents. (The reviewer believes that the author's results would have been better formulated in terms of a somewhat different type of conservative extension; this will be discussed later.)

On the other hand, there is a corresponding model-theoretic idea: The anti-Haecceitist might select a special class of modal structures C such that every modal structure (or, every structure with constant domain) is weakly locally isomorphic to one in C. The special class (a "normalizing" class) can be thought of as giving a *conventional* determination of identities across possible worlds.

Now the model-theoretic approach may correspond to the axiomatic approach in the following way: Suppose L is a modal theory whose consequences are precisely those sentences valid in all structures of a normalizing class C, and suppose L gives sentence eliminability. Since every modal structure (or every structure with constant domain) is weakly locally isomorphic to one in C and C and C and C sentences are preserved under weak local isomorphism, it follows easily that C is a conservative extension of C (or C Sb) for C and C sentences. Since C holds on all models in C and gives sentence eliminability, clearly for every sentence C there is a C sentence C such that C is C holds on all models in C.

The author mentions three examples of this approach. One is to take \mathbb{C} as the class of structures in which the domains of worlds are disjoint (cf. Leibniz; here the convention is to make no identifications). The author says that this does not seem to lead to a corresponding L that eliminates de re modality, although it will if an operator \square is added, where \square A is true in H iff A is true in all worlds other than H. Since in this paper the author wishes to treat only standard modal languages, he does not develop this case further. In fact, however, the project can be carried out for the ordinary modal language if we consider the class \mathbb{C} of all modal structures such that (i) distinct worlds have disjoint domains; (ii) atomic predicates, other than identity, are true in each world H only of existents (members of $\psi(H)$); (iii) each world in the structure is isomorphic to infinitely many others. (The last clause obviates the need to extend the language.) Then we can easily give an axiom set L whose theorems are precisely those sentences valid in \mathbb{C} . It can be shown that in L every sentence is equivalent to a de dicto sentence (though the reviewer cannot see that formula eliminability holds). Since \mathbb{C} is a normalizing class, L is a conservative extension of quantified S5 for de dicto theorems. (Presumably the argument intended here is identical to the one the author intended, but did not give, for the case of disjointness, except for the additional tricks required to keep within the standard modal language.)

The main idea the author considers is homogeneity. Here only models with constant domain are considered. A modal structure ϕ is homogeneous iff for every world in ϕ all isomorphic structures with the same domain (permutations) are also in ϕ . Let $D(x_1, \ldots, x_n)$ say that x_1, \ldots, x_n are pairwise distinct. Let S5H be the extension of S5B obtained by adding to S5B the axiom schema

$$(H) \qquad \qquad \Box(x_1) \dots (x_n) [D(x_1, \dots, x_n) \supset (\Box A \equiv \Box(x_1) \dots (x_n) (D(x_1, \dots, x_n) \supset A))],$$

where all free variables of A are among x_1, \dots, x_n . (So, roughly, there are no "non-trivial" essential

properties or relations distinguishing between some individuals and others: an *n*-ary relation holds necessarily of an *n*-tuple of distinct objects iff it is necessary that it hold for all *n*-tuples of distinct objects.) The author shows readily that S5H corresponds to the normalizing class of homogeneous models in the way described above. The theorems of S5H are precisely the formulas valid in all homogeneous models, S5H is therefore a conservative extension for *de dicto* theorems of S5B, and S5H gives *formula* eliminability. The author also sketches a similar, but more complicated, system that is conservative over S5 allowing variable domains, and he describes a corresponding class of models. The author states that some of his results on homogeneity were also obtained by Broido and by T. J. McKay (*Journal of philosophical logic*, vol. 4 (1975), pp. 423–438); the reviewer had also done similar work (unpublished). The author's version appears to be the best. The author also discusses the light his work sheds on some ideas of Terence Parsons (*The philosophical review*, vol. 78 (1969), pp. 35–52).

The elimination process that gives a *de dicto* formula f(A) provably equivalent to A in S5H is effective. But then f has an important property relevant even to the underlying system S5B. Namely, f is a general recursive function assigning a *de dicto* formula f(A) to *any* formula A, such that A is (provably) equivalent to f(A) in S5B (not just S5H) if it is (provably) equivalent in S5B to a *de dicto* formula at all. For suppose B is *de dicto*, and $A \equiv B$ is a theorem of S5B. Then since $A \equiv f(A)$ and $A \equiv B$ are theorems of S5H, $B \equiv f(A)$ is a *de dicto* theorem of S5B. Hence since S5H is a conservative extension of S5B for *de dicto* formulas, $B \equiv f(A)$ is provable in S5B, so $A \equiv f(A)$ is also provable in S5B.

We can argue further: If A is a sentence invariant under (weak) local isomorphism, then since every model of S5B is weakly locally isomorphic to a homogeneous model, and since $A \equiv f(A)$ holds in homogeneous models and is invariant under weak local isomorphism, $A \equiv f(A)$ holds in all models (is valid). In S5B, this proves the main result of the author's first paper for T empty. (The proof generalizes to any T invariant under local isomorphism.) Although the author does not state this argument, he obviously knows it. Presumably this, in essence, was Broido's proof of his result stated above. (The reviewer has not seen Broido's dissertation.) Clearly the technique here is much more elementary than the techniques apparently needed for the main result of the author's first paper; the latter techniques alone resemble those used to prove the classical model-theoretic preservation results. Using either the author's modification of homogeneity for S5 allowing variable domains or the disjointness method sketched above in this review, one can extend Broido's methods, including effective eliminability, as sketched in this paragraph, even to quantified S5 with variable domains. The reviewer believes that in this case disjointness method gives somewhat simpler proofs than homogeneity.

The effective eliminability results in the preceding paragraph show that it is not really accurate to think of the Broido method as giving a result "weaker" than the author's. The author's method shows that for any theory T, a formula A preserved under local isomorphism of models of T is equivalent to a de dicto sentence. If T is recursively axiomatizable, then obviously there is a partial recursive function g such that g(A) is defined whenever A is preserved under local isomorphism and is a de dicto formula whose equivalence to A is logically implied by T. However, nothing in the author's proof or the method of saturated models sketched above implies that there is a general recursive f such that f(A) is always defined and de dicto (for any A) and is such that T logically implies its equivalence to A when A is preserved under local isomorphism. But the Broido method shows that the stronger effectiveness claim does hold if T is recursively axiomatized and preserved under local isomorphism (in particular, it holds if T is empty). (One could also formulate a question that is independent of the recursive axiomatizability of T: For arbitrary T, is there a totally defined f with the properties mentioned recursive in the set of Gödel numbers of T?) Analogous questions can be formulated for the classical model-theoretic preservation results. For example, universal sentences are those preserved under substructures, but is there a general recursive f, giving a universal formula f(A) equivalent to A if any such formula exists at all? Here the answer has proved to be negative even if only pure logic is in question (Y. Gurevich, Toward logic tailored for computational complexity, Computation and proof theory, Lecture notes in mathematics, vol. 1104, Springer-Verlag, 1984, pp. 175-216; see p. 189). If asked to guess, the reviewer would conjecture that effectiveness fails similarly in the author's theorem, for some particular recursively axiomatized T. But if this is so, this means that the Broido method, and the similar methods applicable if variable domains are allowed, proves a stronger conclusion (effective eliminability) from a much more restrictive hypothesis. It therefore would be incomparable in strength with the author's result.

As is well known, if quantifiers range over arbitrary individual concepts, $(\exists x) \Box A$, for non-modal A, is equivalent to $\Box(\exists x)A$. The author next wishes to propose a theory containing S5B that eliminates de dicto sentences using something resembling this idea. Such a schema cannot be added to S5B without

collapsing the system, however, so instead the author proposes: $\Box((\exists x \neq \vec{y}) \Box A \equiv \Box(\exists x \neq \vec{y})A)$, where \vec{y} is a list of all the free variables in A. Actually the author adds a stronger schema to S5B, giving a system he calls "S5C," namely all universal closures of

(*)
$$\square[(\exists x \neq \vec{y})(A \land \square B \land \diamondsuit C_1 \land \dots \land \diamondsuit C_n) \cdot \equiv \cdot (\exists x \neq \vec{y})(A \land B) \land \square(\exists x \neq \vec{y})B \land \diamondsuit (\exists x \neq \vec{y})(B \land C_1) \land \dots \land \diamondsuit (\exists x \neq \vec{y})(B \land C_n)]$$

where A, B, C_1, \ldots, C_n are non-modal, and x, \vec{y} is a complete list of all free variables in these formulas. It is easily seen that this schema allows any sentence (not formula) to be converted into a *de dicto* equivalent. The author shows that there are two types of models for S5C. In one type, S5C holds if the structure associated with each possible world is invariant under all permutations of its domain ("flat"). In the other, for any non-modal A, we consider models where the schema "I" (for indiscernibility) holds: $\Box(\vec{y})((\exists x \neq \vec{y})A \supset (\exists_2 x \neq \vec{y})A)$, where x, \vec{y} are all the free variables in A. (The schema I says that no property, even involving individual parameters, is uniquely satisfied. The author could have remarked that flat models with infinite domain also satisfy I, and that all flat models satisfy H above.) The author shows how, starting with any model of S5B satisfying I, an extension preserving the same *de dicto* sentences and satisfying (*) can be obtained. The construction is the union of an increasing chain of models: at successive stages models are expanded by duplicating each world infinitely many times, and then (roughly) adding witnesses to the left-hand side of the equivalence in (*) whenever the right-hand side is true. The construction has a pleasing resemblance to constructions in standard model theory, but nevertheless, as the author acknowledges, the resulting models are rather artificial and the relation to the individual concept interpretation is not entirely clear.

The author adds some general results about when a theory or logic admits the eliminability of *de re* modality, including a result showing that S5H is the unique logic that permits formula eliminability and is a conservative extension of S5B for *de dicto* sentences. (This condition, however, is rather strong; the rather natural disjointness method mentioned above does not satisfy it because of the stipulation that atomic formulas are false of non-existents.)

As was said above, the reviewer prefers a different formulation of the conservative extension results. The author is trying to develop a point of view according to which only de dicto formulas are meaningful. The reviewer sees the author's results as "justifying" a system such as S5H by showing that in it every sentence (or even formula) is equivalent to a de dicto sentence (formula), and that it is a conservative extension for de dicto sentences of a standard system, S5B. Then an arbitrary sentence of S5H can be interpreted as "really" meaning its de dicto translation into S5B, and all the axioms of S5H are thus "justified" under this interpretation. Thus S5H has two interpretations, a model-theoretic one given by the homogeneous models, and another in which its theorems are viewed as "disguised notation" for corresponding de dicto theorems provable in S5B. This second justification of S5H requires supplementation, since the de dicto theorems of S5B are proved in a system that contains many axioms that are not de dicto and thus are "meaningless" from the philosophical point of view being presupposed. S5B (just as must as S5H) has to be viewed as an instrument for proving meaningful theorems via steps that need not themselves be meaningful. This is especially obvious in the case of the Barcan formula and its converse, which are schemata with no de dicto instances whatsoever (unless the universal quantifier involved is vacuous); the corresponding model-theoretic idea, constancy of the domain, is not preserved under local isomorphism and is obviously meaningless on the basis of the philosophical view presupposed. One therefore cannot claim that the de dicto theorems of S5B are "evident" on the basis of their proofs in S5B, and a "justification" of S5H by arguing that its theorems all translate into de dicto theorems of S5B is similarly incomplete.

Why not consider a quantified modal language L whose formation rules are restricted so that only *de dicto* formulas are meaningful, that is, the necessity operator is applicable only to closed formulas? Someone who doubts ordinary quantified modal logic solely on the basis of Quine's rejection of "essentialism" ought not to reject quantified modal logic altogether, but rather should prefer the language L. Suppose we restrict the axiom schemata and rules of ordinary quantified S5 (without the Barcan formula and its converse) to their instances meaningful in the language L; call this system "S5"." (For this purpose, it is best to start with the formulation of quantified S5 along the lines of the reviewer's XXXIV 501, where only closed formulas are theorems; following the author, we are considering the case without individual constants.) Then it turns out that S5" is still semantically complete; what this amounts to in an

ordinary formulation of S5 is that every de dicto theorem has a proof with only de dicto steps. (If we do not worry about interpreting de re formulas, this already gives a justification for the ordinary formulation in the style of Hilbert's program.) If we wish to formulate a system that is formulated in L and analogously yields precisely the de dicto theorems of S5B, "S5B-," we cannot simply add the Barcan schema to S5-, since, as we have seen, this is meaningless from the de dicto point of view. Let A_n say that there are exactly n individuals. Then it turns out that added to either ordinary quantified S5 or S5-, the schema $\square(A_n \supset \square A_n)$ (for all finite n), gives exactly the same de dicto theorems as the Barcan formula (or its converse, which are equivalent in S5). We can thus formulate the system S5B- by adding this schema to S5-. We now "justify" S5H, not as an extension of S5B conservative for de dicto theorems, but as a conservative extension of S5B- in the full (ordinary) sense. Further, every theorem of S5H effectively translates into a theorem of S5B-

This approach has an especial advantage in the case of the elimination motivated by the individual concept interpretation. The author believes (second paper, p. 297) that his S5C, highly artificial as it is, is the only natural way of using anything like $\square((\exists x)\square A \equiv \square(\exists x)A)$ in an elimination. There are two fundamental difficulties, both of which stem from the author's insistence on an extension of ordinary quantified S5. First, as we saw above, $\square((\exists x) \square A \equiv \square(\exists x)A)$ cannot be added to quantified S5; so the author contents himself with $\square((\exists x \neq \vec{y}) \square A \equiv \square(\exists x \neq \vec{y})A)$. Second, it would be usual in a formulation with quantification over individual concepts to interpret '=' as coincidence of individual concepts rather than identity; but this interpretation is excluded already by the presence of $(x)(y)(x = y \supset x)$ $\Box x = y$) in quantified S5. The trouble is that an individual concept interpretation motivates the elimination, but as long as we restrict ourselves to extensions of ordinary quantified S5, we are building on a system from which the usual individual concept interpretation (with '=' as coincidence) has already been excluded. With these restrictions, what is surprising is not that an elimination procedure imitating the individual concept interpretation should be highly artifical; the real surprise is that anything like it should succeed at all. (Indeed, the reviewer still does not really have a proper "feel" for S5C and its relation to individual concepts.) The situation changes when instead we consider S5-. The language L of S5- (the de dicto language) has the feature that the interpretation of its sentences remains entirely the same whether the quantifiers range over individuals or individual concepts, and whether '=' is interpreted as identity of individuals or coincidence of individual concepts. Then we can extend S5 to a system formulated in the full language of modal logic, based on full classical quantification theory (thus proving $\Box(\exists x)(x=x)$), but with the substitution schema for identity restricted to non-modal formulas (so that coincidence and individual concepts is the intended interpretation), plus the Barcan schema, plus all closures of the schema

$$(**) \qquad \Box[(\exists x)(A \land \Box B \land \diamondsuit C_1 \land \dots \land \diamondsuit C_n) . \equiv .(\exists x)(A \land B) \land \Box(\exists x)B$$

$$\land \diamondsuit(\exists x)(B \land C_1) \land \dots \land \diamondsuit(\exists x)(B \land C_n)],$$

where A, B, C_1, \ldots, C_n are non-modal. Note that (**) is just like (*) except that the annoying clauses $\neq \vec{v}$? have been omitted; $\square((\exists x)\square A \equiv \square(\exists x)A)$ is a special case. Call the resulting system "S5-C." It is possible to define a class of models with individual concepts such that the formulas valid in these models are precisely those provable in S5⁻C. To get this result one must restrict not only the models allowed but also the family of individual concepts allowed in each model. The resulting class of models is rather artificial (though not so artificial as the models for S5C, since now genuine individual concepts are involved). Nevertheless, in an appropriate sense the models involved constitute a normalizing class, proving that S5⁻C is a conservative extension of (classical) S5⁻ (with $\square(\exists x)(x=x)$). (Note that S5⁻C therefore does not contain S5B⁻, in spite of the presence of the Barcan schema and its converse!) Further, it is easy to use (**) (and the Barcan schema and its converse) to show that in S5⁻C every sentence is equivalent to a de dicto sentence ("sentence eliminability"). It follows that it is possible to interpret S5⁻C in two ways: either as a system with quantification over individual concepts, or as a reformulation of S5 where the sentences that are not de dicto are used as reformulations of de dicto sentences in disguised notation. On the second interpretation, the quantifiers might as well be viewed as ranging over ordinary individuals! So in a way, the results show that individual concept quantifiers of a certain kind can be regarded, in a sense, as disguised notation for ordinary quantifiers over individuals. In spite of the obvious parallel between S5⁻C and S5C, the reviewer is unclear about any relations between them. The reviewer agrees with the author that it is natural to preserve the customary logic, including identity, and admires the author's ingenuity in obtaining individual-concept-like interpretations satisfying this

condition. But since the *de dicto* language admits interpreting '=' as coincidence, one does not violate "logic" if one adopts this interpretation, which is natural in connection with individual concepts.

The reviewer should mention that the artificiality of the models of S5⁻C is entirely due to the impoverished notation available in the conventional modal language L of S5⁻. If the modal language is enriched by allowing operators $\diamondsuit_n A$, saying that A holds in at least n distinct worlds, or propositional quantifiers (this is the method required if we consider systems weaker than S5), then much more natural elimination results can be proved. The systems involved have as their intended interpretations the class of all models, with quantifiers ranging over all individual concepts. Nevertheless the elimination results, showing that every sentence is equivalent to a de dicto sentence and that the system involved is conservative over an appropriate extension of S5⁻, show that the quantifiers over individual concepts results can be regarded as disguised notation for quantifiers over individuals. (The reviewer had obtained this version prior to the author's papers.)

In his third paper, the author is concerned with the following problem: In the reviewer's semantics for modal logic, the truth value of a sentence can be affected by changes in the extensions of atomic predicates that affect only the applicability of these predicates to non-existents in certain worlds (or the relations of existents to non-existents in these worlds); in each world the extensions of the predicates remain unchanged when restricted to existents. It might be natural to regard this feature of the semantics as objectionable. Could there be a possible world just like the actual one as far as the things that exist are concerned, and as far as the properties of and relations among existing objects are concerned, but somehow differing from the actual one because of what is true in it of objects that exist neither in it nor (therefore) in the actual world? Perhaps one should hold that a world is not really changed if the set of existents and the properties of and relations among existents are unchanged, regardless of what happens to relations among non-existents or between existents and non-existents. If so, something may seem to be wrong if such changes are allowed to affect the truth values of formulas.

The author considers two ways out, though he acknowledges that his survey is far from exhaustive. (As he mentions, he does not deal at all with approaches that invoke truth-value gaps.) One way out admits as really meaningful only those sentences whose truth values are unaffected by changes that leave the domains of each world and the relations among existents in each world unchanged. The author shows that a sentence has this property if and only if it is equivalent to the sentence obtained by replacing each atomic part $P(x_1, \ldots, x_n)$ by $P(x_1, \ldots, x_n) \wedge E(x_1) \wedge \ldots \wedge E(x_n)$, where all free variables in the formula have been listed. The proof is simple: The property holds only if the formula's truth value would remain the same whether or not we impose the convention that atomic formulas are false whenever non-existents are involved, but the extra conjuncts involving existence amount to imposing this stipulation regardless of the truth value assigned to the original atomic formula. The author also shows that this is no longer true if we restrict ourselves to models of an arbitrary modal theory T. Surprisingly, the result is restored, even relative to a modal theory, if the underlying logic is not S5, but some weaker system such as S4, K, or T. The phenomenon appears to be related to analogous phenomena involving the interpolation lemma, which fails for all modal logics between KB and S5 (proved for S5 by the author in XLVIII 486; the review XLVIII 486 gives arguments sufficing for the more general statement); but holds for S4, etc. (with variable or cumulative domains but not constant domain; proved by the author and others; see again XLVIII 486 and the review).

Another approach holds that atomic formulas are false whenever some of the variables are assigned non-existents. In his XXXIV 501, the reviewer preferred not to build this stipulation into the semantics, natural though some people think it is, since then substitution of arbitrary formulas for predicates will not hold; the valid formulas will no longer be valid as *schemata*. However, *if* this is really the *only* reason for rejecting this stipulation, and the idea of a privileged class of genuinely atomic (or "positive") formulas is accepted otherwise, it is natural to consider the system based on this idea (with obvious added axioms for the atomic predicates), and to consider what *schemata* are valid in the system. (If identity is included, whether it should entail existence presents a special problem. For S5, we can follow the author's idea of defining ordinary identity as $\diamondsuit(x = y)$, where '=' is existence-entailing identity.) Any formula valid in the reviewer's system is obviously valid as a schema in the modified system, in the sense that all substitution instances are provable; but the converse is by no means obvious. If the converse were to fail (as far as the reviewer knows, no one has investigated the question), then from this point of view it could be argued that the reviewer's semantics and formal system give too small a class of theorems. Even from this point of view, however, an enlargement of the language to admit other locutions (other intensional operators,

infinitary formulas, etc.) will tend to decrease the class of valid schemata, and it seems almost inevitable that eventually the class will be narrowed to the validities investigated by the reviewer.

The author, however, takes a different point of view. He speaks of "theory interpretability," where a theory T is interpretable in the desired sense if under a fixed substitution of formulas for atomic predicates the consequences of the resulting theory T, in models where atomic formulas are always false of non-existents in every world, consists exactly of the "translations" of the theorems of T. Except perhaps in some very special situations, the reviewer would not have thought this the most natural approach to take; rather the theorems of T should be those that are provable under all substitutions, as in the schematic interpretation of formulas. Given his approach, the author shows that not all theories are interpretable but that some simple sufficient conditions imply theory interpretability. He also shows that theory interpretability always holds if we allow the quantifiers in the translating theory T to be relativized to some formula A(x).

In his first paper, the author describes a method he regards as *the* way to prove the completeness of quantified modal systems, avoiding "the usual inelegancies in axiomatization and proof." This seems highly misleading. The method is very closely related to the reviewer's basic original methods (see XXXI 120, XXXI 276, XXXIV 501, XXXV 135). In comparison it does not change the axiomatizations proved complete at all, and it changes the basic proofs only slightly. Review first the classical situation, using Smullyan's (XL 237 and XL 508) elegant formulation (which we change slightly) simply as a perspicuous way of viewing even versions not explicitly presented in his terms. Suppose \wedge , \sim , and (x) are the only primitives. An analytic (cut-free) consistency property is a property of sets of sentences meeting the conditions: (1) If $A \wedge B$ or $(\sim \sim A)$ belongs to $(\sim \sim A)$ or $(\sim \sim A)$ or $(\sim \sim A)$ or $(\sim \sim A)$ belongs to $(\sim \sim A)$ or $(\sim \sim A)$ or $(\sim \sim A)$ belongs to $(\sim \sim A)$ or $(\sim \sim A)$ or $(\sim \sim A)$ belongs to $(\sim \sim A)$ or $(\sim \sim A)$ or $(\sim \sim A)$ belongs to $(\sim \sim A)$ or $(\sim \sim A)$ belongs to $(\sim \sim A)$ or $(\sim \sim A)$ belongs to $(\sim \sim A)$ or $(\sim \sim A)$ belongs to $(\sim \sim A)$ or $(\sim \sim A)$ belongs to $(\sim \sim A)$ or $(\sim \sim A)$ belongs to $(\sim \sim A)$ or $(\sim \sim A)$ belongs to $(\sim \sim A)$ for each constant $(\sim (\sim A))$ belongs to $(\sim \sim A)$ or $(\sim \sim A)$ finite set $(\sim \sim A)$ is jointly satisfiable.

The proof uses (1) and (2) to extend S_0 to an increasing chain of finite sets in Γ whose union S is closed under conditions corresponding to (1) and (2) (e.g. if $\sim (A \land B) \in S$, either $\sim A \in S$ or $\sim B \in S$), and contains no subset $\{A, \sim A\}$. Then, in a familiar manner, given the closure conditions, a model for S can be "read off" from S itself. (If Γ is closed under unions of chains, which normally can be made to hold, the proof of the fundamental theorem extends straightforwardly to an infinite set S_0 provided there are "enough" constants not in S_0 ; in the uncountable case the ascending chain is uncountable.) Whether or not they explicitly put it this way, "semantic tableau"-type completeness proofs of a formal system Q show that the property Γ of joint consistency in Q (consistency of the conjunction) of a finite set S is an analytic consistency property.

A synthetic (Henkin) consistency property is defined by the conditions: (Cut) If $S \in \Gamma$, either $S \cup \{A\}$ or $S \cup \{ \sim A \} \in \Gamma$. (2*) If $\sim (x)A(x) \in S \in \Gamma$, then $S \cup \{ \sim A(c) \} \in \Gamma$ for each constant c not in S. (3*) If has been added and (3) is strengthened to (3*); but these changes render (1) and half of (2) superfluous. (From Gentzen we are familiar with the idea of adding cut to (1)-(3), which is essentially equivalent.) Fundamental theorem: If Γ is a *synthetic* consistency property, every finite set in Γ is jointly satisfiable. This follows from the previous theorem, since it is easy to show that if Γ is a synthetic consistency property, the property of being contained in a set in (or, if Γ is closed under subsets, as ordinarily holds, Γ itself) is an analytic consistency property. It is traditional in Henkin-style arguments, however, to argue directly, again extending the given set by an increasing chain of sets in Γ , whose union S is now closed under conditions corresponding to cut and (2*). Nevertheless, to complete the proof by "reading off" a model from S one must still argue that S is closed under the conditions corresponding to (1) and (2) (e.g. that if $\sim (A \land B) \in S$, then $\sim A \in S$ or $\sim B \in S$), invoking essentially the same argument used to reduce the synthetic case to the analytic case. Thus, even as formulated before Smullyan's exposition, the difference between tableau (analytic) and Henkin (synthetic) style completeness proofs comes to this: An analytic proof argues by verifying directly that Γ (joint consistency in the system) is closed under certain properties, while a synthetic proof derives some (not all) of these properties from others that are directly verified (cut and (3*)—note that (3*) strengthens (3) in ways corresponding exactly to the omitted conditions). Some important consistency properties are not obviously closed under cut, so the analytic method has greater generality; it postulates the minimum needed (see the review XXII 360 by Craig) for the proof. If Γ is obviously closed under cut, each style of proof can be trivially and mechanically obtained from the other.

In his original completeness proofs (see XXXI 120, XXXI 276, XXXIV 501, XXXV 135), in effect the reviewer extended the notion of an analytic consistency property to modal logic, using an indexed family of sets of formulas, with each index corresponding to a world. Conditions (1)–(3) above apply to each indexed set (world) in the family. The added conditions for necessity are obvious from the possible world semantics and can be thought of as translations of the conditions for the universal quantifier (in the language with quantification over worlds). The proof that every synthetic consistency property implies an appropriate notion of joint satisfiability proceeds, exactly as in the classical case, by a chain of approximations to the model. Given these ideas, the only additional point that is not a routine analogue of the classical argument is the formulation of a particular modal consistency property that implies the completeness of the axiomatic system, namely consistency of the "characteristic formula" of a finite indexed family of sets. For S5, the case the author explicitly considers, this is simply $\Diamond A_1 \wedge \ldots \wedge \Diamond A_n$, where A_i is the conjunction of the formulas in the *i*th world—or, if one wishes to distinguish the real (first) world, $A_1 \wedge \Diamond A_2 \wedge \ldots \wedge \Diamond A_n$. For systems weaker than S5, one must build up the characteristic formula in a more complicated "nested" manner, corresponding to the accessibility relation (see the references above), and for some quantified systems it is best to complicate the definition further still; the details will be omitted.

The author's method (as given in the paper for S5 as an example) is the same, using the same notion of characteristic formula, except that he proves that consistency of the characteristic formula is a synthetic consistency property! As in the classical case, one must verify implicitly or explicitly that a synthetic consistency property is an analytic consistency property. (In agreement with Fitch, this JOURNAL, vol. 31 (1966), p. 152, and Fitting, Tableau methods of proof for modal logics, Notre Dame journal of formal logic, vol. 13 (1972), pp. 237-247, whom he does not mention, the author prefers to consider a single set of indexed formulas, rather than an indexed family of sets. An indexed formula A, can then be interpreted as "A holds in world w." In the present context, this is merely a notational difference, since it is a triviality of universal mathematics that an indexed family of sets corresponds to a single set of index objects; but it may make the connection with the language with quantification over worlds slightly more perspicuous. The reviewer began with this formulation but chose the other one so as to keep close to the modal language.) The reviewer cannot see that completeness proofs, of the same axiomatizations, that are line-for-line identical to proofs by the analytic (tableau) method, except for a simple mechanical transformation well known from the classical case (essentially replacing direct verification of the analytic closure properties by an indirect verification), can constitute as new a departure as the author apparently thinks he has made. There are not-completely-routine modal problems, not mechanically derived from the classical case and varying from system to system, in formulating an appropriate notion of characteristic formula and proving that its consistency in the system is a consistency property; these are essentially the same whether 'consistency property' is taken in the analytic or the synthetic mode (and the easy translation between the two styles is uniformly the same). (Actually, unless one complicates the axiomatization considerably, for various systems, such as S4 and T, it is difficult or impossible directly to prove that the appropriate notion is a synthetic consistency property, while the proof that it is an analytic consistency property is easy. This difficulty can be overcome by restricting the cut condition to a weaker one still implying the analytic conditions, which uniformly postulate the minimum needed.) The author never mentions the mechanical way his proofs relate to the analytic proofs, not even that he uses the notion of characteristic formula. The reviewer can imagine considerations (the same modally and classically) favoring either style of proof, but the styles correspond so closely that neither can be said greatly to simplify the other.

As is well known, classically and modally the analytic conditions suggest the very natural "tableau" proof procedure for unsatisfiability, where one systematically extends a set S_0 by conditions (1) and (2). Because of the condition for $\sim (A \wedge B)$, the procedure branches into alternatives; if every branch eventually violates (3), S_0 is unsatisfiable. A similar proof procedure corresponds to the synthetic method (essentially equivalent to the analytic procedure with cut). A more ambitious result than mere completeness of a formal system, using essentially an effective version of the arguments above, effectively translates a refutation by the proof procedure into a disproof in the system. (Such a procedure must consider the entire proof tree, not just a branch as above.) Thus it not only shows that every unsatisfiable formula is disprovable, but also shows how to exhibit a disproof. The reviewer proved the stronger result, and "linearized" the disproofs obtained by using disjunctions of characteristic formulas, without pointing out that this is not necessary for completeness itself. The simpler-to-present (but not essentially

different) non-effective argument has long been well known for classical logic in both analytic and synthetic styles and is followed by the author. In general, some cumbersome features of the reviewer's original presentation may have obscured the simple basic method. (But the reviewer feels that the reader who understands the method thoroughly should really find the effective argument natural also.) Another difficulty may be that the reviewer, wrongly expecting his method to be so well understood from the published cases that unpublished arguments would easily be reconstructed, published the details for some basic propositional systems, but for only two quantified systems (one of them intuitionistic logic).

The author stresses that his method applies "straightforwardly to uncountable languages." So does the analytic method, as stated above; and the trivial correspondence still holds. In any case, as the reviewer, Mortimer, van Benthem, Smoryński, and probably others have remarked (and see even the author's related XLIII 373(2)), as long as the modal axioms correspond to first-order properties of the accessibility relation, the compactness theorem for a modal system is simply a special case of the compactness theorem for classical first-order logic (regarding the modal structures as classical structures), so that special modal arguments are superfluous. Hence at present the advantage the author states his method (like the original method) has over the "standard" method (see below) has no known substantial application to compactness. Uncountability is obviously irrelevant to (so-called weak) completeness. (Similar remarks apply to the author's observation that his method yields a modal Löwenheim—Skolem theorem.)

The author also emphasizes that his method is a translation of the classical synthetic (Henkin) method as applied to the first-order language with quantifiers over worlds. As opposed to the "standard" method, this indeed has a much better claim as a translation. However, exactly the same relation holds for the analytic (tableau) method; in any event, it is hard to see why it is so important that a modal completeness proof be a translation of a classical proof.

Probably the author primarily compared his method with the "standard" method in the literature (due to Scott, R. Thomason, Makinson, Schütte, and others), normally presented using the family of maximal consistent sets. Its widespread use may have led him to think of it as "usual... axiomatization and proof," wrongly regarding the original methods as obviously superseded. Here his method is not mechanically obtainable from its rival, and some of the published literature might support his impression of its far greater simplicity for axiomatization and proof. Moreover, for some cases the "standard" method does lead to complications avoided (in the same way) by the original method and its synthetic variant. Without space for details here, nevertheless the reviewer feels that the author exaggerates the intrinsic complexity of the "standard" method also. Many others, largely because they ignored quantified modal logic, erred in the opposite direction; the reviewer agrees with the author that the methods set out above (whether in the analytic or in the synthetic mode) are a significant rival to, and somewhat simpler than, the "standard" method. The author has unwittingly discovered that the widespread feeling that the original methods were superseded was unjustified.

Often the "standard" method is called "the Henkin method." One might think that since the author also uses the Henkin method, in this respect he is closer to the "standard" method than to the tableau method. But, as we saw, the original method changes relatively little whether it is based on the tableau (analytic) or Henkin (synthetic) method. On the other hand, though again we do not examine details, the "standard" method is no more closely related to the Henkin method than to any other classical method of proving completeness. Properly set out, it does not imitate classical methods but merely cites classical results. The idea that any fundamental contrast derives from "Henkin" versus "tableau" methods should be abandoned.

The reviewer would reject the author's quest for "the" method of proving completeness. Various methods (not all mentioned above) have various virtues and differing applications. This becomes even clearer if the language is enriched by infinite conjunction, other quantifiers, etc., where not all methods appear to apply equally to all cases. Why not let one hundred flowers bloom?

Finally, although the reviewer admires the author's success in combining model theory and philosophy, he feels that some of the author's initial remarks in favor of his program give an exaggerated impression, belied even by the author's own successful practice elsewhere. Against those who tried to prove modal analogues of classical model-theoretic results, he suggests that the analogues may fail (see the interpolation lemma for some systems) or be devoid of philosophical interest. But the first point only argues that the program of extending classical results should be replaced by one of comparing modal and classical results. In modal logic as in extensions of classical logic, the question of when a property such as the interpolation lemma fails and holds is itself of interest. (And some similar problems are purely modal.

For example, if a modal propositional or quantified logic satisfies completeness and compactness, must it be determined by a first-order class of frames? An answer for quantified logic would help clarify the uncountable situation discussed above.) As to the second point, a problem originally pursued for purely technical reasons may turn out to be related to philosophical questions (compare the author's paper on the interpolation lemma with its review XLVIII 486). Further, what is wrong with the purely technical pursuit of mathematically natural questions even if the original motivation is philosophical? (The author himself has successfully pursued such questions. Possibly he means to protest the mechanical importation of classical problems into modal logic even when they are neither mathematically nor philosophically natural, and one can sympathize with this.) Some of these questions are simple analogues of classical results. Others are more peculiarly modal. For example: A modal structure can be expanded by adding (i) worlds, (ii) individuals (to the domains of worlds), or (iii) both. Analogously to results of Tarski and Feferman, it is natural to conjecture that a modal sentence is preserved under these extensions iff, respectively, it is equivalent to a sentence built out of atomic formulas and their negations by using A, ~, possibility, existential quantification and (i) universal quantification, or (ii) necessity, or (iii) neither of these. SAUL A. KRIPKE