GRAEME FORBES

TWO SOLUTIONS TO CHISHOLM'S PARADOX

(Received 28 August, 1983)

(1) In the original version, Chisholm formulated his paradox in the language of Kripke semantics for first-order modal logic, the language of possible worlds and transworld identity (see [3]). Thus Quine's concise statement of the paradox ([18], p. 861): '...you can change anything to anything by easy stages through some connecting series of possible worlds'. But if the paradox does not arise in first-order modal language itself (which could only be because some sentence in its possible-worlds formulation lacks a reverse translation into modal language) this would suggest merely that the concepts of Kripke semantics are in some way incoherent in the part where they exceed the expressive power of modal language, and not that there is a problem with any *logic* used in the interference. However, the paradox *does* arise in modal language, a fact which, granted that we are not to deny any of its premissses, means that we have to look more closely at the logic involved. To fix ideas, here is one instance of the reasoning, in which α is a table, h_1 is the Goodmanian sum of wood from which the table is actually made, and for any h_i, h_{i+1} is a sum of wood which is almost but not wholly coextensive with h_i; finally, h_n is a sum of wood which has no part in common with h_1. Then if 'M–, –' is read '– is constituted of exactly the sum of wood –', we have this first version of Chisholm's Paradox:

$$\begin{array}{l} \Diamond M\alpha h_1 \\ \Box(M\alpha h_1 \rightarrow \Diamond M\alpha h_2) \\ \quad\vdots \qquad\qquad \vdots \\ \Box(M\alpha h_{n-1} \rightarrow \Diamond M\alpha h_n) \\ \hline \Diamond M\alpha h_n \end{array}$$

The paradox arises from the fact that we are inclined to believe *both* that the predicate M has a certain tolerance in it, i.e., crudely, that one and the same table could have been made (without leftovers) from who different sums of

Philosophical Studies **46** (1984) 171–187. 0031–8116/84/0462–0171$01.70

wood if the area of overlap of the two sums is sufficiently great, and *also* that one and the same table could *not* have been made from two sums of wood which do not overlap at all (better formulations are given immediately below).[1] I call this version the first version since it corresponds to Chisholm's own way of stating the problem: if we take any world u possible with respect to the actual world in which α has a certain constitution, then there is a world v possible with respect to u in which α has a slightly different constitution, and since v will also be accessible to the actual world, it is therefore possible for α to have that slightly different constitution; and so on.[2]

In the object language, the inferential move being made here is embodied in the following valid sequent of S5:

$$\text{(C)}\qquad \frac{\begin{array}{l}\Diamond A\\ \Box(A \to \Diamond B)\end{array}}{\Diamond B}$$

Presented thus, it is immediately obvious that one way to resolve Chisholm's Paradox is to reject the S5 principle upon which it turns, that is, in terms of possible worlds, to deny that the accessibility of v from u and u from the actual world implies the accessibility of v from the actual world. Taking appearances at face value, we ought to say that it is merely the assumption of a transitive accessibility relation which is causing the difficulties. So the problem is one of *modal* logic and is to be resolved by abandoning S5 for a modal system in which (*C*) is invalid.

But this is too quick. In S5, the formula '$\Box(A \to \Diamond B)$' is equivalent to the conditional '$\Diamond A \to \Diamond B$', a fact which permits the reasoning in Chisholm's Paradox to be reformulated in the following way:

$$\frac{\begin{array}{l}\Diamond M\alpha h_1\\ \Diamond M\alpha h_1 \to \Diamond M\alpha h_2\\ \quad\vdots \qquad\qquad \vdots\\ \Diamond M\alpha h_{n-1} \to \Diamond M\alpha h_n\end{array}}{\Diamond M\alpha h_n}$$

This formulation, which we call the Sorites formulation, assimilates Chisholm's Paradox to the standard form of a classical Sorites Paradox; and now it is much less clear that the problem arises because of some fallacious modal inference, since there is no modal logic in a standard Sorites; so a solution of Chisholm's Paradox which focusses on the accessibility relation between worlds, a piece of machinery peculiar to modal logic, runs the risk of not directly addressing the heart of the matter. Moreover, it is plausible that the Sorites formulation is the better formulation of the paradox; the intuitive thought which gets the Paradox going is that a sufficiently small degree of change in the original constitution of an object *preserves* possibility for that object, a thought which is naturally expressed as a conditional: *if* such-and-such a constitution is a possibility for α, *then* so is such-and-such. So one who assents to the premisses of Chisholm's Paradox in its first version should also assent to its premisses in the Sorites version, and a solution which works for one version should be accepted only if it works equally well for the other.

(2) Of the two solutions to this paradox which I propose to discuss in detail, only the first satisfies the constraint just suggested. Nevertheless, there are parallels between the solutions in certain respects, and ways in which the second may even seem superior. Let us begin by considering the exact content of the intuition of tolerance which generates the Paradox. One formulation, 'NT' for 'nested tolerance', is evidently the one germane to the first version of the Paradox, while another 'CT' for 'conditional tolerance', is naturally associated with the Sorites formulation. Reading 'S–, –' as '– sufficiently overlaps –', these principles are:

(NT) $\Box(\forall x)(\forall y)(\forall z)(Sxy \rightarrow \Box(Mzx \rightarrow \Diamond Mzy))$

and

(CT) $\Box(\forall x)(\forall y)(\forall z)(Sxy \rightarrow (\Diamond Mzx \rightarrow \Diamond Mzy))$.

(NT) and (CT) are equivalent in quantified S5, but if we introduce a non-transitive accessibility relation, (NT) will not imply (CT), for the same reason as '$\Box(A \rightarrow \Diamond B)$' will not imply '$\Diamond A \rightarrow \Diamond B$'.[3]

Since we have said that the natural formulation of the intuition of tolerance is as a conditional, let us start with the solution which bears on the Sorites formulation of the Paradox. This solution, which is hinted at by Kripke and as been advocated in detail by the present writer, makes use of counterpart theory.[4] The motivation for this approach comes from taking Chisholm's Paradox to be a modal paradox of vagueness. Tolerance arises because of vagueness or fuzziness in the limits of the range of sums of wood which possibly constitute α: there is no sharp distinction between those sums which could, and those which could not, constitute α. Given that there is no fuzziness in the boundaries of particular sums of wood or in the constitution relation, it seems that this vagueness must arise from an underlying vagueness in the concept of possibly being identical to α; however, in standard modal semantics, such vagueness could only be represented by vagueness in α's transworld identity conditions, and a solution of the paradox in which we think of identity as vague would be rather unappealing. But is does make sense to think of *similarity* as being vague, in the sense of admitting *degrees*, and one well-known treatment of ordinary Sorites arguments resolves them by providing a semantics based on infinitely many degrees of truth, a semantics which blocks e.g. the paradox of smallness, generated by the tolerance principle that a man only marginally (say, not perceptibly) taller than a small man is himself small. Since the counterpart relation is fixed by similarity considerations – in the present context, similarity of design and constituting matter – and similarity admits of degrees, the degree-theoretic resolution of non-modal paradoxes may be transcribable, and in fact can be transcribed, into the modal logical context.

Briefly, the details are these. Suppose *a* is small and *b* not perceptibly different in height. Then by tolerance, the conditional 'if *a* is small then *b* is small' is true, and so a paradox can be constructed in which we reason through a chain of conditionals to conclude, of some clearly tall man, that he is actually small. However, the idea that 'small' is a predicate of degree suggests that all we have here is a *reductio* of the application of two-valued logic in a situation where we need instead a logic of infinitely many degrees of truth (represented, say, by the real interval [0, 1]), a need which arises because we have a language containing predicates which can be satisfied to any degree between 0 (complete non-satisfaction) and 1 (complete satisfaction).[5] In this semantics, the degree of truth of a conjunction (disjunction) is the lower (higher) of the degrees of truth of its conjuncts (disjuncts), and the degree of truth of an existential (universal) quantification is the least upper bound (greatest lower bound) of the degrees of truth of all instances (for simplicity, we assume every object has a name). For the conditional, we generalize the classical clause in a slightly different but equally natural way; if the antecedent does not have a higher degree of truth than the consequent, the conditional is wholly true, but if the antecedent is more true than the consequent, the degree of truth of the conditional should be fixed by how *much* more true the antecedent is; for instance, if the gap between antecedent and consequent is close to 1, the conditional should be almost wholly false. Generally,

$$\begin{aligned} \text{Deg}[A \rightarrow B] &= 1 - [\text{Deg}(A) - \text{Deg}(B)] \text{ if } \text{Deg}(A) > \text{Deg}(B) \\ &= 1 \text{ otherwise.} \end{aligned}$$

If we say that a rule is *valid* if its conclusion in any application never takes a degree of truth lower than the greatest lower bound of the degrees of truth of the premisses to which it is applied, then *modus ponens* is invalid: letting $\text{Deg}(A) = 0.8$ and $\text{Deg}(B) = 0.7$, we get 0.9 for '$A \rightarrow B$', 0.8 for the other premiss, but 0.7 for the conclusion. So in a Sorites paradox, we are detaching consequents whose degree of truth is falling steadily towards 0, which, evidently, is something we should not do. Thus if b is actually taller than a, however slightly, his degree of smallness will be slightly less than a's (degrees can differ non-observationally); then by semantic ascent, we see that each application of *modus ponens* in the standard paradox commits the 'fallacy of detachment' just illustrated. Now consider the counterpart-theoretic translation of an arbitrary conditional premiss of the Sorites formulation of Chisholm's Paradox, in which '$Cxyw$' is read as 'x is a counterpart of y at w':

$$(\exists u)(\exists x)(\exists y)(Cx\alpha u \;\&\; Cyh_i u \;\&\; Mxyu) \rightarrow \\ (\exists v)(\exists x)(\exists y)(Cx\alpha v \;\&\; Cyh_{i+1} v \;\&\; Mxyv).$$ [6]

If counterparthood admits of degrees, then the degree to which some x is a counterpart of a at u can be slightly higher than the degree to which any x is a counterpart of α at v, since some x at u is made of more of the (counterpart of) the wood of which α is actually made than is any y at v (we are assuming design is held constant). It is easy to see that the details can be filled in so that we get a conditional with a truer antecedent than consequent, and so Chisholm's Paradox is shown to turn on the fallacy of detachment, just as the paradox of smallness does. It should be emphasized that this solution works by introducing degrees of *de re* possibility, *not* degrees of identity, in order to render 'possibly being identical to α' a complex predicate of degree.

Nathan Salmon has offered a different solution to Chisholm's Paradox, in which he tries to avoid the use of counterpart theory, which he holds to be a mistaken approach to the semantics of modality (see [19], pp. 240–252). Following an idea of Hugh Chandler's (in [21]) Salmon exploits the accessibility relation on worlds in standard Kripke semantics instead. Salmon's accessibility relation admits of three statuses: given two worlds u and v, v may be definitely accessible from u, or definitely inaccessible, or neither, that is, it may be indeterminate whether v is accessible from u. The translation of a conditional premiss of the Sorites version of Chisholm's Paradox, using 'R' for accessibility and 'w^*' for the actual world, will then be as follows:

$$(\exists u)(Ruw^* \,\&\, M\alpha h_i u) \rightarrow (\exists v)(Rvw^* \,\&\, M\alpha h_{i+1} v)$$

But in Salmon's framework, accessibility is not transitive, since u may be definitely accesible from w^* while v is not definitely accessible from w^* although it is accessible from u; hence, if a conditional with true antecedent and intermediate consequent is merely intermediate itself, some premiss of our argument may be only intermediate, so that there would be no reason to expect that argument's conclusion to be true.

(3) Which of the solutions is to be preferred? From here on, I shall be concerned to argue the virtues of the counterpart-theoretic treatment over the vices of the accessibility treatement, but first let us note a point of similarity between the two. Our initial intuition about the premisses of a Sorites argument is that they are all true; if I cannot see any difference in height between Smith and Jones and I agree that Smith is small, then it is hard to see how I could have any grounds for withholding agreement that Jones is small; and even if I later learn that careful measurement has shown Jones to be one millimetre taller than Smith, it is not realistic to insist that I should then be *less* confident about Jones' smallness than I am about Smith's.[7] But Salmon's treatment of the Sorites version of the modal paradox decrees that it will have two conditional premisses which are not true, while on my treatment, it is consistent to hold that every conditional premiss is less than *wholly* true, and necessary to hold this of some. So the intuition that the conditionals are all true is abandoned on both solutions.

But the counterpart-theoretic solution accomodates the intuition in a way that Salmon's solution does not, for his solution does not treat the conditionals uniformly: it makes them all true but for the two where in moving from the antecedent to the consequent we move from worlds with one kind of accessibility status to w^* to worlds with the next best accessibility status. For instance,

$$\Diamond M\alpha h_5 \rightarrow \Diamond M\alpha h_6$$

(say) will have truth-value I because there are definitely accessible worlds where the non-modal part of its antecedent is true but only at best intermediate worlds where the non-modal part of the consequent is true. But there is no intuitive basis for distinguishing two of the conditionals from all the others: intuitively, they all seeem to have the same degree of acceptability, since they all exploit the intuition of tolerance, that if such-and-such is a possibility then so is something slightly different, in the same way. The counterpart-theoretic approach is, by contrast, faithful to this phenomenon, since the natural ascription of degrees of counterparthood makes every premiss almost wholly true; and this also explains why we are inclined to regard them just as true. Similar points apply to (CT) itself. Because a universal quantification has no better truth-value than that of its worst instance, Salmon must say that (CT) has value I, whereas on the counterpart-theoretic approach its degree of truth can be brought arbitrarily close to complete truth by making the conditions for sufficient overlap more and more strict.

These problems with Salmon's accessibility solution have their source where one would expect: in the arbitrary division of worlds into three groups with respect to w^*. Since there are no sharp discontinuities in the range of worlds with respect to the composition of α, a three-valued division of this range is no better than a two-valued one; the same holds for any finite number of values fixed in advance. Indeed, it is hard to see the point of the introduction of the three-way classification of worlds at all, since every principle to which Salmon wishes to object appears to fail in the standard system *B* with non-transitive accessibility. Salmon may protest that it just does seem to him that some premisses of the Sorites version of the Paradox are not true, but it is hard to see how this claim is any better than that of the man who holds that there really is a sharp point at which smallness ends. If we may not stipulate away Sorites paradoxes, we may not stipulate away the Sorites version of Chisholm's Paradox.

Nevertheless, Salmon can respond that even as it stands, his solution fares better with the first version of the Paradox, since it counts all that version's premisses as wholly true and explains the conclusion as resulting from a move permitted only in modal logics with transitive accessibility. Thus to deal with the first version of the Paradox, it is only necessary to abjure such logics in contexts where there is tolerance. The counterpart-theoretic solution, on the other hand, deals with the first version as it does with the Sorites version: each premiss is at best almost wholly true since the '$\Diamond$' carries us to worlds where degree of counterparthood between the relevant objects is less than 1.

In a natural deduction systm it would still be application of *modus ponens* (in conjunction with □-Elimination) which would take us through the premisses to the conclusion, so the fault remains with a principle of ordinary propositional calculus. In sum, therefore, the choice between the counterpart-theoretic solution and Salmon's involves a trade-off: we can retain *modus ponens*, treat the two formulations of the Paradox quite differently, and regard the logical problems involved as pertaining strictly to modal logic, or we can treat the two formulations in the same way and retain an S5-style modal logic, but at the cost of rejecting classical propositional calculus.

Some may think that it must always be preferable to alter our modal logic if the alternative is to change non-modal propositional calculus. But this view simply expresses a failure to take the very great parallel between Chisholm's Paradox and ordinary paradoxes of vagueness seriously, for if the same phenomenon underlies both, there can be no objection to modifying propositional calculus in the modal case, since this is what is standardly done in the non-modal cases; and the particular resolution which uses the apparatus of degrees and rejects *modus ponens* has been widely canvassed and is at least *popular*. From this point of view, then, it is difficult to see a justification for Salmon's asymmetrical treatment of the two formulations of the Paradox, and for his distinguishing (NT) from (CT). To repeat, the intuition about tolerance is that a sufficiently small degree of change in original constitution *preserves* possibility for an object, and this thought has a certain generality about it: it is merely the amount of change which matters, so that *if* one constitution is a possibility, *then* change within the bounds of that amount is also a possibility. Certainly, we may also express the idea in the proposition that *given* that the constitution of an object is thus and so, then it might also have been slightly different, but here we mean to be expressing the *same* thought as before. In Salmon's system, these thoughts come apart; it is therefore no consolation that the second is deemed wholly true, for the manipulations with the accessibility relation have robbed us of the natural way of expressing its content; indeed, if I may anticipate what I shall say about such relations later, it is unclear that what Salmon intends by (NT) is the expression of any thought we can formulate.

(4) These criticisms of Salmon focus on the details of his proposals, but there are also more general points of disagreement between his approach and that of the counterpart theorist. Indeed, the disagreement about details can be dis-

solved by allowing the accessibility relation to be a relation of degree, and arranging that the degrees of accessibility between worlds should mirror in some appropriate way the degrees of counterparthood between worldbound individuals which the counterpart theorist posits. So it is to the more general issues that we have to turn to make a decision about the final merits of the two approaches. These issues concern the apparatus of the two solutions, counterpart theory in the one case and accessibility relations in the other. In this section, I shall argue against some objections to the use of counterpart theory, and in the next, I shall advance some positive considerations about models and representationality which mitigate against the accessibility theorist.

In defense of counterpart theory, it should be emphasized that it is first-order modal sentences, and not any philosopher's extensional equivalents, which are the subject-matter of pretheoretic semantic intuitions.[8] Salmon is only one who argues as if we could appeal to some kind of *intuition* to establish that (1) below means the same as (2) rather than (3):

(1) Possibly, α is F;

(2) For some world w, w is accessible to w^* and α is F at w.

(3) For some world w, some counterpart of α is F at w.

Thus, he writes that counterpart theory is really 'a particularly inflexible brand of essentialism', on the grounds that (1) means that α *itself* is F at some world, so that the counterpart theorist, who confines α to a single world, must regard all the intuitively contingent properties of α as in fact essential to it ([19], p. 236). But in fact, whether or not a theory admits contingency (in matters other than existence) turns only on whether or not it is consistent with the truth of some instance of the *object* language schema 'x exists and is F and possibly exists and is not-F', and by this criterion, counterpart theory admits contingency beyond all question.[9]

To hold that counterpart theory is not exonerated from the charge of rigid essentialism because its rendering of instances of the contingency schema do not mean what the modal instances mean is to think of the extensional sentences as having some meaning given independently, only the sentences of the Kripkean metalanguage having meanings which permit them to be 'matched up' with modal object language sentences. But this conception of the meanings of sentences of these languages invented by philosophers is not very plausible; how, exactly, are these meanings to be characterized? The threat is that a *literal* reading of the existential quantifier over worlds is implied by this view, so that we would have to accept the picture of an array of possible worlds quite literally, or else (preferring a non-literal reading of 'world' would have to reductively identify possible worlds with logical constructions of actual entities; and many find the realism of the former option unattractive, while the latter has recently been shown to be problematic (see [17]). It seems better to think of the meanings of the extensional sentences as being given by those of the modal ones themselves (so far as this is possible); possible worlds semantics of either sort may be thought of as based on the idea that in ordinary English, 'possibly, *A*' and 'there is some possibility that *A*' are equivalent in meaning; the semantics then proceeds by taking the apparent first-order structure in the latter seriously, and in moving from the sentential to the quantificational context it would be up to the theorist himself to decide just how to proceed, given his purposes. But from this starting point, one cannot think of the extensional sentences of either semantics as yielding perspicuous representations of the 'real' meanings of the modal sentences: the translation does not function as, say, Russell thought of his Theory of Descriptions as functioning, as *explanatory* of meaning, even if it does fix the extension of 'valid'. Yet Salmon's criticism makes sense only if we think of the theorist's extensional language in these unlikely ways.[10]

Salmon has another objection to counterpart theory, that it makes certain inferences necessarily truth preserving as a matter of modal logic which, intuitively, are not *logically* valid. So this objection does focus on the object language. The idea is that since the counterpart relation stands for some sort of similarity relation, there must be a two-place predicate of the object language which expresses it. Let '*C*' be this predicate. But then (my version of) counterpart-theoretic semantics validates the inference

$$\frac{\Diamond(\exists x)(Cxa \,\&\, Fx)}{\Diamond Fa}$$

since it interprets the premiss as asserting

(4) $(\exists w)(\exists x)(Exw \,\&\, (\exists y)(Cyaw \,\&\, Cxyw \,\&\, Fxw))$

and the conclusion as asserting that

(5) $(\exists w)(\exists x)(Cxaw \,\&\, Fxw)$.

The trouble is that the counterpart-theoretic semantics ought to be equivalent to the standard semantics when the counterpart relation is a one-one equivalence relation, and to obtain this result we need certain assumptions on the counterpart relation, assumptions which in conjunction with (4) entail (5); two such assumptions are that each object is its own sole counterpart at any world where it has no existing counterpart, and that each object is its own sole counterpart at the world where it exists (so x and y have to be the same in (4)). Granted the need for the equivalence, one cannot just change the semantics, so there appears to be a counterintuitive consequence here.[11]

However, there is a *non-sequitur* in Salmon's step from the introducibility of a two-place relation expressing the similarity relation the crossworld facts about which determine the extension of the counterpart relation in any *intended* model, to the conclusion that the displayed inference is valid. In an *arbitrary* model, the extension of the counterpart relation may be completely untouched by considerations of similarity: all that is required is that it conform to whatever stipulations on the relation are made in the definition of counterpart-theoretic model, for example, the two stipulations just mentioned above. Of course, these stipulations are motivated by the desire to provide an explanation of the transworld heirlines of objects in intended models (transworld identity conditions in the Kripke framework), but still, the intended concept drops out of view when questions of validity arise. Hence in claiming that the displayed inference is valid when the relation 'C' is added to the object language, Salmon is in effect saying not just that "C" 's real meaning is explanatory of counterparthood, but that he proposes to treat the expression as a new *logical constant*. This strategy is made manifest in the translation of the object-language predicate 'C' in (4) by the same three-place metalanguage symbol as expresses the counterpart relation of an arbitrary model, a relation which will have little to do with similarity in general.

Of course, Salmon cannot be denied the right to introduce a new logical constant into modal language, provided he gives it a proper evaluation clause in the required extension of the model theory. Given that the displayed inference is to be valid (or that the translation (4) is to be legitimate) the clause should just be:

(C) $w \models Ctt'$ iff $\langle \mathrm{Ref}(t), \mathrm{Ref}(t'), w\rangle \in C$

where C is the counterpart relation of the model. Nothing less than this guarantees the validity of the inference; in particular, Salmon's requirement that the object-language relation '*C*' express the similarity relation relevant to counterparthood is insufficient. But now there is no force whatsoever in the observation that the validity of the inference is 'counterintuitive'; any peculiarity about the inference is entirely attributable to the peculiarity of the new logical constant, and no-one has suggested that the counterpart theorist is *required* to have such a constant in the modal object language. Were he to introduce one, questions about *intuitions* of validity or invalidity of arguments involving it would not arise; since the constant is not one of ordinary modal discourse, no one *has* any intuitions of the appropriate sort; one must just accept the dictates of the theory.

So we have rebutted both of Salmon's objections to counterpart theory. Recalling our repsonse to the first, however, it seems that we have still to provide a reason to *prefer* a counterpart-theoretic resolution of Chisholm's Paradox, since the import of our argument there was that the introduction of a counterpart relation between worldbound individuals was no less acceptable, from the point of view of intuitive semantic evidence, than any of the Kripkean technicalia, in particular, the accessibility relation. We turn now to advancing considerations to show that the counterpart-theoretic solution is the superior one.

(5) There are two strong objections to the accessibility solution. One is that it is internally incoherent, in that it implies the rejection of a principle which is needed to motivate the search for *any* solution to the Paradox, and the other is that only an S5-style system is adequate for representing the logic of broadly logical or 'metaphysical' necessity. Let us take up the question of coherence first.

According to Salmon, the conclusion of Chisholm's Paradox is false not because there is no possible world in which α is made from h_n, since he holds that there are such worlds; rather, the conclusion is false because all such worlds are definitely inaccessible from w^*. This means that while it is false to assert '$\Diamond M\alpha h_n$', we could obtain a true assertion by iterating 'possibly' a sufficient number of times. However, it is also clear that we can truly assert '$\Diamond(\exists x)(Mxh_n)$', since there is nothing impossible about some table or other being constructed from h_n. Furthermore, it is clear that for any world u at which '$M\alpha h_n$' is true, there is an accessible world v at which a sentence of the form 'Mth_n' is true, where t is not a name of α, and which is otherwise indistinguishable from u. That is, if t is replaced with 'α' in every sentence true at v then the resulting sentences together with all other sentences true at v yield a complete description of u (recall we are assuming every possible object has a name). So u and v are two worlds which differ merely with respect to the identity of a single individual, yet differ they must, since v is accessible from w^* while u is not.

The problem is that it is hard to see why someone comfortable with the distinction between u and v should regard the conclusion of Chisholm's Paradox as false in the first place, since the argument for the falsity of the conclusion relies on a certain principle about the concept of identity which the distinction between u and v would appear to flout. Consider an arbitrary world w in which some table other than α is made from h_n according to the design of α in w^*. If the conclusion of the Paradox is actually true (so that there is really no paradox), this table is indistinguishable in every intrinsic respect from α as it is in a world realizing the truth of the conclusion, which is a *reductio* of the hypothesis that the conclusion is true if we accept that numerical distinctions between entities must be grounded in differences between them in intrinsic respects. Someone who accepts this principle can avoid the *reductio* only by insisting that there are no such worlds as w, but since h_1 and h_n do not overlap, this is tantamount to denying that it is possible for α to be made from h_1 while an other table is constructed from h_n, and is therefore an implausible response; it may even be falsified by the actual world.[12]

Returning to the allegedly distinct worlds u and v, we now ask whether it is consistent to admit that these worlds are distinct while holding that the Paradox's conclusion is false since there could not be such a distinction between tables as its truth would require. If the distinction between the tables is illusory because it cannot grounded, why is the distinction between u and v not equally illusory? Salmon cannot say that this latter distinction is grounded in the facts about accessibility just mentioned, since *that* difference is grounded by the presence of α in u with its impossible (relative to w^*) constitution, and it is the existence of such a u which is in question. So Salmon must say that he rejects the conclusion of Chisholm's Paradox on the basis of a restricted version of the principle about identity, such as that it is necessary that x is identical to y if x could have had an intrinsic nature which y could also have had,[13] a principle consistent with ungrounded differences holding across worlds not both of which are accessible to the reference world. But making such a restriction appears quite *ad hoc*. The principle on which rejection of the conclusion of the Paradox turns is intended to express a truth about the concept of identity *simpliciter*, not the concept of transworld identity between worlds accessible to a given world. So in the alleged failure of transworld identity between the tables in *u* and *v*, we have a flouting of what one who rejects the conclusion of Chisholm's Paradox should regard as a conceptual truth; it is irrelevant that the particular instance does not involve only accessible worlds. Hence the internal incoherence of this accessibility solution to Chisholm's Paradox follows.

There is also a more general reason to reject any kind of accessibility solution. In giving a model of certain modal facts, one can distinguish between what Kaplan calls 'artifacts' of the model on the one hand and its representational features on the other (see [12]). If a feature is merely artifactual, then any two models of certain modal facts which are isomorphic but for that feature should be equally acceptable representations of those facts; hence the accessibility theorist must show that the accessibility relation is no mere artifact of the model, since he does not regard *u* and *v* above as equivalent for the purposes of modelling modal facts. As an example of a representational feature of models, according to our modal concepts, consider the distinction between its being possible for a particular object *x* to be *F* and its being possible for some object or other to be *F*; this distinction is captured in models by a transworld relation holding between objects in the domains of worlds, an identity relation in Kripke models and a counterpart relation in others, and the role this relation plays in the evaluation of formulae expressing that distinction. So the relation is representational. But to what phenomenon can the accessibility theorist point to establish the representationality of the accessibility relation in a way that does not presuppose the correctness of his views about the questions at issue here?

The only phenomena which come to mind are apparent intuitions of difference in content between various judgements with iterated modal operators. For instance, it might be said that at the intuitive level, 'it is possible that it is possible that *A*' makes what appears to be weaker assertion than 'it is possible that *A*': perhaps we can agree that there is a way things could have been such that if they had been thus, it would have been possible that *A*, without committing ourselves to the proposition that it *is* possible that *A*. Then this distinction in content would be representable in models with a non-transitive accessibility relation. But it is doubtful that the distinction can be made out. It is easy to hear one statement as weaker than the other if, for instance, we allow ourselves to equivocate on 'it is possible that': obviously, it may be logically possible that is is physically possible that A, without its actually being physically possible that A. However, when such equivocation is guarded against, and when we are careful to mean 'possible in the broadly logical sense' whenever we use 'it is possible that', it is much less clear that iteration of operators has a weakening effect.

We can say something stronger than this: when 'it is possible that' has this meaning, there is in fact a good reason why iteration should be redundant. If we consider substantial philosophical theses whose formulations employ broadly logical possibility and necessity, such as the theses that the members of a set are essential to it or that if it exists, an organism belongs of necessity to the biological kind to which it actually belongs, we see that there is a conceptual character to such claims: establishing them involves investigating the notions of set and set membership, and of kind and subsumption under a kind, and the interconnections of these concepts with the idea of what it is to be an individual thing of the given sort. What the broadly logical necessities are is therefore fundamentally an *a priori* matter, to do with the content of our concepts, even though with the addition of *a posteriori* information, necessary *a posteriori* truths can be inferred. If this is correct, the idea of contingent possibility or necessity to which the accessibility theorist is committed hardly makes sense: surely no-one will want to say that a merely possibly possible world would have been possible if our concepts had been different, or if we had had the concepts required to understand the propositions true at that world, which as a matter of contingent fact we do not. So it is my position that the representationality of the accessibility relation in the context of broadly logical possibility is at best extremely dubious. Returning to Chisholm's Paradox, then, I conclude that the counterpart-theoretic resolution is the correct one.[14]

NOTES

[1] This second thought is argued for in Section 5 below. The claim that M is tolerant is defended in my paper [8], in which the counterpart-theoretic solution of the Paradox to be defended here is set out in detail. But the present paper can be read independently.

[2] In [4] Chisholm writes (p. 149): '... there will be a world possible in respect to *that* world, and therefore also in respect to this one, in which x and y will have exchanged still other small parts'. Nathan Salmon drew my attention to the use of nested modalities in Chisholm's own exposition of the problem.

[3] The following is a model of (NT) in which (CT) is false. Let $W = \{w_i\}$, $i \in N$, and let all sentences of the form Sh_ih_{i+1} be true at every world. For every pair of worlds (w_i, w_{i+1}) let $M\alpha h_i$ be true at the first member, and $M\alpha h_{i+1}$ be true at the second, and let these sentences be false at all other worlds. Finally, set Rw_iw_j iff either (i) $w_i = w_j$, or (ii) $w_j = w_{i+1}$, or (iii) Rw_jw_i, and let w^* be w_0.

[4] See [14], Note 18 (p. 51) and Note 57 (p. 115) for Kripke's remarks. Counterpart theory is due to David Lewis; see [15].

[5] Here and elsewhere I follow [10].

[6] For a short sketch of this version of counterpart theory, see [7].

[7] In holding that 'if Smith is tall then Jones is tall' should not be rejected, I am following the discussion of observational predicates in [20]. For further discussion, see [8].

[8] I take this point from Allen Hazen's [11], where he uses it to rebut criticisms of counterpart theory by Kripke and Plantinga in which the object language/metalanguage distinction is not observed.

[9] In Lewis' original version, contingent existence is not admitted. In [7], this limitation is overcome.

[10] The views on which these remarks are based are propounded by Colin McGinn in his [16], and in my own [9].

[11] See [19, p. 235]. Correspondence with Salmon has improved my understanding and formulation of his objection. When I say that counterpart theory should be equivalent to the standard S5 semantics in the special case when the counterpart relation is a 1–1 equivalence relation, I mean that any such counterpart theoretic model should be isomorphic to some counterpart model itself obtained from a standard model S by replacing each object x in the domain D of S by the objects $\langle x, w\rangle$, one for each w such that in S, $x \in \text{dom}(w)$, putting $\langle x, u\rangle$ in $\text{dom}(v)$ iff $x \in \text{dom}(v)$ in S and $u = v$, keeping relations the same under this replacement, and setting $C(\langle x, w\rangle, \langle y, u\rangle, v)$ iff either $x = y$, $u = v$ and $\langle y, u\rangle \in \text{dom}(v)$, or $\langle x, w\rangle = \langle y, u\rangle$ and $-(\langle y, u\rangle \in \text{dom}(v))$.

[12] The argument in the text here can be found (with a little interpretation) in Note 56 to [14], and is examined at length in Chapter 7 of [19] (in [13], the Note is rendered almost incomprehensible by misprints). I arrived independently at a related argument for the necessity of origin in which the importance of principles about identity is emphasized; see my [6]. Of course, these principles have their analogue for the counterpart theorist. Apart from such principles, I can think of no other reason to reject the conclusion of Chisholm's Paradox. It should be emphasized that there is no general objection to the hypothesis of two worlds which are the same but for the identity of a single individual; for example, the Castor/Pollux example in [1] is quite unproblematic, since here the transworld differences reduce to intraworld intertemporal differences. An objectionable distinction would be one of the sort drawn by Chisholm, who holds that if a human brain were divided and the resulting halves placed in different bodies so that two persons resulted, the original person would be identical to one of these two, even though the latter are indistinguishable in every respect we would regard as relevant to the question of identity with the original person. Here is a surely unintelligible hypothesis of identity, unintelligible because ungrounded in any intrinsic relation between the persons said to be identical. See Chisholm's [5]. My view is that Salmon's position is unintelligible in the way Chisholm's is: Chisholm posits an identity where none can hold, Salmon denies one where it must hold.

[13] The particular principle which Salmon is willing to endorse is the following ([19], p. 249, Note 28): 'necessarily, if a table x is the only table originally constructed from a certain hunk of matter according to a certain plan, then necessarily, any table that is the only table originally constructed from that hunk of matter according to that plan is the table x and no other'. This principle identifies intrinsically indistinguishable tables across u and v if v is accessible from u, but not otherwise. Thus certain possibly possible objects which are such that it is contingently impossible for them to exist may be intrin-

sically indistinguishable from objects which possibly exist but, *a fortiori*, are distinct from them. So we speak here of an identity relation that extends across worlds within the range of modal operators arbitrarily iterated, and the claim in the text is that that relation must satisfy the principle that identities and non-identities be grounded; and this truth about identity is the only reason to accept the principle which Salmon endorses.

14 I am grateful to Nathan Salmon for some helpful comments on earlier versions of this paper.

BIBLIOGRAPHY

[1] Adams, R.: 1978, 'Primitive thisness and primitive identity', The Journal of Philosophy 76, pp. 5–26.
[2] Chandler, H.: 1976, 'Plantinga and the contingently possible', Analysis 35, pp. 106–9.
[3] Chisholm, R.: 1968, 'Identity through possible worlds: some questions', Noûs 1, pp. 1–8.
[4] Chisholm, R.: 1970, 'Reply to Strawson's comments', in Kiefer, H. E. and Munitz, M. K., Language, Belief and Metaphysics (SUNY Press).
[5] Chisholm, R.: 1976, Person and Object (Open Court).
[6] Forbes, G.: 1980, 'Origin and identity', Philosophical Studies 37, pp. 353–262.
[7] Forbes, G.: 1982, 'Canonical counterpart theory', Analysis 42, pp. 33–37.
[8] Forbes, G.: 1983, 'Thisness and vagueness', Synthese 54, pp. 235–259.
[9] Forbes, G.: 1983, 'Physicalism, instrumentalism and the semantics of modal logic', The Journal of Philosophical Logic 12, pp. 271–298.
[10] Goguen, J.: 1969, 'The logic of inexact concepts', Synthese 19, pp. 325–73.
[11] Hazen, A.: 1979, 'Counterpart-theoretic semantics for modal logic', The Journal of Philosophy 76, pp. 318–38.
[12] Kaplan, D.: 1975, 'How to Russell a Frege–Church', The Journal of Philosophy 72, pp. 716–29.
[13] Kripke, S.: 1972, 'Naming and necessity', in Davidson, D. and Harman, G. (eds.): Semantics of Natural Language (Reidel), pp. 253–355.
[14] Kripke, S.: 1980, Naming and Necessity (Harvard).
[15] Lewis, D.: 1968, 'Counterpart theory and quantified modal logic', The Journal of Philosophy 65, pp. 113–126.
[16] McGinn, C.: 1981, 'Modal reality', in Healey, R. (ed.): Reduction, Time and Metaphysics (Cambridge University Press), pp. 143–87.
[17] McMichael, A.: 1983, 'A problem for actualism about possible worlds', The Philosophical Review 92, pp. 49–66.
[18] Quine, W. V. O.: 1976, 'Worlds away', The Journal of Philosophy 73, pp. 859–63.
[19] Salmon, N.: 1981, Reference and Essence (Princeton University Press).
[20] Wright, C.: 1975, 'On the coherence of vague predicates', Synthese 30, pp. 325–65.

Department of Philosophy,
Tulane University,
New Orelans, LA 70118,
U.S.A.