

# *Canonical Counterpart Theory*

GRAEME FORBES<sup>1</sup>

IN a recent article in *Analysis*, Graeme Hunter and William Seager (1981) attempt to rescue counterpart theory (CT) from some objections of Hazen 1979. They see these objections as arising from ‘uncritical use of the translation scheme originally proposed by Lewis’, and intend to meet them by refraining from use of that scheme. But they do not offer a new scheme; they say ‘...it is no more necessary to have one to capture the sense of modal idiom than it is to capture the sense of quantificational idiom...Appeal to truth value is the single most important criterion of correct translation’ (Hunter and Seager 1981:72). Thus, where the scheme of Lewis (1968) translates a truth by a falsehood or conversely, Hunter and Seager simply produce a sentence which they claim to be a better translation, without articulating any structural principles they employ to arrive at their candidate.

A friend of CT should not be happy with this procedure. Let  $\mathcal{L}_m$  be first-order modal language and  $\mathcal{L}_c$  be the language of counterpart theory. Sentences of  $\mathcal{L}_c$  are intended to interpret or elucidate the meanings of  $\mathcal{L}_m$ -sentences in the strong sense that they should be the outputs of an adequate theory of meaning, which we can formulate as a model theory, for  $\mathcal{L}_m$ , written in  $\mathcal{L}_c$  plus set theory. The interest of CT resides in its potential to provide an alternative approach to the model theory of modal logic, and from the clauses of any such alternative model theory we should be able to read off a translation scheme for  $\mathcal{L}_m$  into  $\mathcal{L}_c$ , just as we can in the standard case. If CT’s motivation is sound, which Hunter and Seager do not dispute, yet there is no translation scheme and so no counterpart-theoretic model theory, one might suspect some defect in the original language  $\mathcal{L}_m$ . And in fact, Hunter and Seager find certain  $\mathcal{L}_m$ -sentences multiply ambiguous with respect to  $\mathcal{L}_c$ ; but if a counterpart-theoretic model theory is possible, to say

<sup>1</sup> For this reprinting of the paper I have made minor verbal improvements and also resolved a crucial ambiguity in clauses (7) and (9).

that an  $\mathcal{L}_m$ -sentence has more than one permissible  $\mathcal{L}_c$ -translation is just to say that we have not yet made up our minds what the model theory is to be.

To resolve these issues, we have to provide a model theory, or equivalently, a translation scheme into the language of a first-order theory of counterparts from which a model theory can be derived. The scheme must be free of the technical drawbacks of Lewis's, so as to be invulnerable to Hazen's objections. Such a scheme is provided below, prefaced with motivating remarks. In devising it, quantified S5 in the canonical formulation due to Kripke (1963) and Fine (1978) has been kept in mind. The main points about this formulation are these:

- (i) there is no accessibility relation on the set  $W$  of worlds;
- (ii)  $\Box$  expresses *strong* necessity, that is, the truth of  $\Box A$  at any world requires the truth of  $A$  at every world, not just at worlds in which the denotations of any terms in  $A$  all exist;
- (iii) a predicate can be satisfied at a world  $w$  by objects which do not exist in  $w$  – thus, Jones is identical to himself even with respect to worlds in which he does not exist;
- (iv)  $\mathcal{L}_m$  contains a logically constant predicate *exists*, whose extension at each world  $w$  in any model is the inner domain  $D_w$  of  $w$ ;
- (v) quantifiers are *actualist* (i.e. a quantified variable ranges over only the *inner* domain of the world with respect to which the scope formula is being evaluated).

$\mathcal{L}_c$  is, for convenience, a two-sorted language which contains the non-modal connectives of  $\mathcal{L}_m$ , say negation and conjunction, and also predicates and constants corresponding to those of  $\mathcal{L}_m$  except that each is boosted by one place (but for '='). In deciding upon a counterpart predicate we take note of the distinction marked under (iii) above. Thus, counterparts of an object  $x$  must be available to satisfy certain predicates with respect to worlds at which  $x$  has no existing counterparts. One way of meeting this requirement is to stipulate that for each  $x$  in the set  $D$  of

all possible objects in a model (the union of all the  $D_w$ ), if  $x$  has no counterpart at  $w$  which is in, that is, which exists at,  $w$ , then  $x$  is its own sole counterpart at  $w$ . We also stipulate that each  $x$  in  $D$  belongs to exactly one  $D_w$ , and at this  $w$ ,  $x$  is its own, sole, counterpart. Further, these are the only two ways in which any  $x$  can be a counterpart of itself at a world  $w$ . Finally, if one of  $x$ 's counterparts at  $w$  is in  $w$ , they all are. From the way these stipulations are stated, we can see that it is natural to have a three-place counterpart predicate  $Cv_1v_2w$ , which is to be read  $v_1$  is a counterpart of  $v_2$  at  $w$ . So  $C(x, Jones, w)$  says  $x$  is a counterpart of Jones at  $w$ , which leaves it open whether  $x$  is in  $w$  or not. The stipulations themselves are axioms of CT, or alternatively, clauses of counterpart-theoretic model theory.

With these preliminaries, let us proceed to the translation scheme Trans. We stipulate that the translation of any  $\mathcal{L}_m$ -formula  $\phi$  into a formula of  $\mathcal{L}_c$  is the *relativization* of  $\phi$  to the actual-world constant ' $w^*$ '. That is,

$$\forall \phi \in \text{Form}(\mathcal{L}_m), \text{Trans}(\phi) = \text{Rel}(\phi, 'w^*').$$

For each formula  $\phi$  of  $\mathcal{L}_m$ , we compute  $\text{Trans}(\phi)$  by means of the following recursive definition of  $\text{Rel}(\phi, w)$  for an arbitrary term  $w$ . In the schematic clauses,  $\text{Rel}$  is a functor which for particular values of the schematic letters  $v, A$ , etc., yields terms for  $\mathcal{L}_c$ -formulae; corner quotes on both sides of the clauses are fulfilling their usual function of selective quotation; thus, e.g., ' $\neg \text{Rel}(A, w)$ ' expands into ' $\neg \wedge \text{Rel}(A, w)$ '.

- (1) If  $Ft_1, \dots, t_n$  is atomic,  $\text{Rel}(Ft_1, \dots, t_n, w) = Ft_1, \dots, t_n, w$  except when  $F$  is '=';  $\text{Rel}(\ulcorner t_1 = t_2 \urcorner, w) = \ulcorner t_1 = t_2 \urcorner$ ;
- (2)  $\text{Rel}(\ulcorner \neg A \urcorner, w) = \ulcorner \neg \text{Rel}(A, w) \urcorner$ ;
- (3)  $\text{Rel}(\ulcorner A \wedge B \urcorner, w) = \ulcorner \text{Rel}(A, w) \wedge \text{Rel}(B, w) \urcorner$ ;
- (4)  $\text{Rel}(\ulcorner (\forall v)\phi \urcorner, w) = \ulcorner (\forall v)(\text{Exists}(v, w) \rightarrow \text{Rel}(\phi, w)) \urcorner$ ;
- (5)  $\text{Rel}(\ulcorner (\exists v)\phi \urcorner, w) = \ulcorner (\exists v)(\text{Exists}(v, w) \wedge \text{Rel}(\phi, w)) \urcorner$ .

For a modal formula  $\diamond A$  or  $\square A$  there are two cases. If  $t$  is an occurrence of a constant or free variable in  $\diamond A$  or  $\square A$  that is not within the scope of any modal operator in  $A$ , say that  $t$  is *immediately* within the scope of the  $\diamond$  or  $\square$  prefixed to  $A$  (thus in  $\diamond(Fx \wedge \square Fb)$ ,  $x$  but not  $b$  is immediately within the scope of  $\diamond$ ). Then if in  $\diamond A$  there are no term-occurrences immediately within the scope of the  $\diamond$ ,

$$(6) \text{Rel}(\ulcorner \diamond A \urcorner, w) = \ulcorner (\exists w_j)(\text{Rel}(A, w_j)) \urcorner$$

But if  $t_1, \dots, t_n$  are occurrences of constants or free variables immediately within the scope of the displayed  $\diamond$  in  $\diamond A$ ,

$$(7) \text{Rel}(\ulcorner \diamond A \urcorner, w) = \ulcorner (\exists w_j)(\exists v_1) \dots (\exists v_n)(Cv_1 t_1 w_j \wedge \dots \wedge Cv_n t_n w_j \wedge \text{Rel}(A(t_i/v_i), w_j)) \urcorner.$$

The  $v_i$  and world variable are peculiar to  $\mathcal{L}_c$  and are chosen to avoid clash of variables, and  $A(t_i/v_i)$  is the result of replacing each relevant  $t_i$  in  $A$  with  $v_i$ .

There are two analogous clauses for  $\square$ :

$$(8) \text{Rel}(\ulcorner \square A \urcorner, w) = \ulcorner (\forall w_j)(\text{Rel}(A, w_j)) \urcorner;$$

$$(9) \text{Rel}(\ulcorner \square A \urcorner, w) = \ulcorner (\forall w_j)(\forall v_1) \dots (\forall v_n)((Cv_1 t_1 w_j \wedge \dots \wedge Cv_n t_n w_j) \rightarrow \text{Rel}(A(t_i/v_i), w_j)) \urcorner.$$

Some comments. Clauses (4) and (5) reflect point (v) about canonical S5, that quantifiers there are actualist; but it would be easy to accommodate possibilist quantifiers. Clauses (7) and (9) reflect point (ii), that  $\square$  is strong, although we could accommodate a weak  $\square$ . In addition, these clauses exhibit an insensitivity to iterated modalities that is lacking in Lewis's original scheme. For example, in S5 we want  $\diamond \diamond Fa$  to be equivalent to  $\diamond Fa$  even if the counterpart relation is not transitive. (6) and (7) give  $\text{Trans}(\diamond \diamond Fa)$  as

$$(10) (\exists w)(\exists w')(\exists y)(Cyaw' \wedge Fyw')$$

but the clause  $T2_j$  of Lewis (1968), simplified to a two-sorted language, results in a rendering of  $\diamond\diamond Fa$  as

$$(11) (\exists w)(\exists x)(In(x, w) \wedge Cxa \wedge (\exists w')(\exists y)(In(y, w') \wedge Cyx \wedge Fyw')).$$

The  $\mathcal{L}_c$ -cognate of (11) is not equivalent to (10), since it will speak of a counterpart of  $x$  at  $w$ , and this object, when the counterpart relation is not transitive, may not be a counterpart of  $a$  at  $w'$ .

What of the notorious  $\mathcal{L}_m$ -sentence  $(\forall x)\Box(\exists y)(x = y)$ , invalid in  $S5$  with contingent existence but translated into a theorem of CT by Lewis's scheme? Our scheme translates it as

$$(12) (\forall x)(Exists(x, w^*) \rightarrow (\forall w)(\forall z)(Czxw \rightarrow (\exists y)(Exists(y, w) \wedge z = y)))$$

which is false for precisely the right reason:  $z$  can be a counterpart of  $x$  at  $w$  without existing in  $w$ . It is the introduction of the three-place counterpart relation and the inner/outer domain distinction which facilitates this result.

Next, what becomes of the inference from  $\Box Rab$  to  $\Box(\exists x)Rax$ , where  $R$  is atomic? In their discussion of this inference, Hunter and Seager do not explicitly state the crucial point: assuming actualist quantifiers and contingent existence, the inference is correct only if  $\Box$  is strong and the Falsehood Principle (that a non-logical atomic sentence is false at a world where the things named in it do not all exist) is in force. Our translation scheme for strong  $\Box$  produces, respectively, (13) and (14):

$$(13) (\forall w)(\forall z)(\forall t)((Czaw \wedge Ctbw) \rightarrow Rztw)$$

$$(14) (\forall w)(\forall z)(Czaw \rightarrow (\exists x)(Exists(x, w) \wedge Rzxw)).$$

(14) is not a consequence of (13) in the first-order theory of counterparts so far developed, but if we include the Falsehood Principle, perhaps by adding an axiom schema restricted to atomic substitutions, (14) will be a consequence of (13) in the

resulting theory. On the other hand, if we alter the translation scheme to make  $\Box$  weak (this occasions further complications),  $\text{Trans}(\Box(\exists x)Rax)$  will not be a consequence of  $\text{Trans}(\Box Rab)$  in any reasonable CT. These are the desired results.  $\Box Rab$  was a sentence which Hunter and Seager's methodology led them to pronounce quadruply ambiguous (1981:74) but in fact their four readings correspond to four different sentences of  $\mathcal{L}_m$  with strong  $\Box$ , each of which has a single correct translation into  $\mathcal{L}_c$ , as given by the scheme above.

Finally, what becomes of the necessity of identity? Although not discussed by Hunter and Seager, it was in difficulties with this formula that Hazen found the greatest inadequacy of (Lewis 1968) and the main motivation for his own complex proposals about the correct way of doing counterpart-theoretic semantics; and nothing we have said so far meets this part of Hazen's position. The necessity of identity is translated into a theorem of CT only if each object has exactly one counterpart at each world. This would have to be added to the axioms of CT as a special stipulation, while the validity of the modal thesis itself is delivered by an intrinsic feature of standard S5 models, the feature that the transworld heirlines of a model are given by the real crossworld identities in it; from the usual point of view, if there are no such identities (with respect to inner domains) then every property of an object is essential to it. What this suggests, however, is that the advantage of counterpart theory may lie just in its suitability for contexts where transworld heirlines are not "identity-like" and so cannot be given by any identity relation.

Hazen himself provides a clue to what such contexts might be like, when he points out (*op. cit.*, p. 330) that a counterpart theory which simply permits branching heirlines has strange-seeming consequences if the *Actually* operator is introduced into  $\mathcal{L}_m$ . For instance, the following will be true:

$$(15) \ \diamond(\exists x)(\text{Actually}(\exists y)(x = y \wedge \text{Actually}(Fx) \wedge \text{Actually}(\neg Fx)))$$

although it seems to verge on contradiction. However on some views, there are

contexts where the meaningful assertion that  $P$  is in respect of truth neither better nor worse off than the assertion that not- $P$ , examples where we have a *borderline case* of  $P$ 's obtaining. A thought prompted by this is that counterpart theory comes into its own when the possibility of being a borderline case of a counterpart of a given object arises, or better, where there is the possibility of degrees of counterparthood. If this is the only kind of case we need countenance in which the transworld heirline relation is not an equivalence relation, the special case for which standard modal model theory is appropriate may be characterized as that in which it is always the case that a given object either definitely is or definitely is not a counterpart of another object. But there can be many other cases. Chisholm (1967) has pointed out that by making a sequence of small changes in an object which preserve identity we can accumulate a large change which does not preserve identity, which throws the coherence of transworld identity into doubt. But a counterpart relation which admits of degrees need not be transitive in any relevant sense, and can be employed to construct a consistent model of the situation Chisholm has in mind. In such a model, on an appropriate interpretation, (15) is of course not wholly true, but nor is it wholly false; rather, it has an intermediate, non-classical degree of truth related to the degrees to which the relevant counterpart relations hold. This is no more puzzling than the fact that on the envisaged approach to vagueness and borderline cases, employing essentially the Lukasiewicz infinite-valued logic, instances of  $P \wedge \neg P$  can take intermediate truth values; for in response to the question 'Is that person an adult?' asked of some precocious adolescent, we may perfectly coherently reply, 'Well, he is and he isn't.'

## BIBLIOGRAPHY

- Chisholm, Roderick. 1968. 'Identity Through Possible Worlds: Some Questions.' *Noûs* 1:1-8.
- Fine, Kit. 1978. 'Model Theory for Modal Logic Part I: The De Re/De Dicto Distinction.' *The Journal of Philosophical Logic* 7:125-56.
- Hazen, Allen. 1979. 'Counterpart-Theoretic Semantics for Modal Logic.' *The Journal of Philosophy* 76:319-338.
- Hunter, Graeme, and William Seager. 1981. 'The discreet charm of counterpart theory.' *Analysis* 41 (2):73-76.
- Kripke, Saul. 1963. 'Semantical Considerations on Modal Logic.' In *Reference and Modality*, edited by L. Linsky. Oxford and New York: Oxford University Press.
- Lewis, David. 1968. 'Counterpart Theory and Quantified Modal Logic.' *The Journal of Philosophy* 65:113-126.