Logique & Analyse **147-148** (1994), 255-263

COUNTERPART THEORY AS A SEMANTICS FOR MODAL LOGIC*

Lin WOOLLASTON

Abstract

Counterpart theory is supposed to provide a semantics for intensional languages. In counterpart theory *de re* modality is represented in terms of an object's counterparts in other worlds. Unlike our standard possible worlds semantics, an object and its counterparts are not identified with one another. The relation between an object and its counterparts is, in general, neither transitive nor symmetric. A non-transitive and non-symmetric counterpart relation means that an object need not have exactly one counterpart in any world. Allowing an object to have no counterpart in a given world results in basic modal principles, such as K and $\Box(\alpha \wedge \beta) \supset \Box\beta$, failing in counterpart theory. I also show that supposing an individual to have more than one counterpart in a given world results in the invalidity in our modal language of either universal instantiation or $\Box(\alpha \vee \sim \alpha)$.

Introduction

One way in which David Lewis motivates his metaphysics is to suggest that it provides a semantics for intensional languages (Lewis 1986; 19-20). The metaphysics are formalized by the postulates of counterpart theory. The semantics, which incorporate these postulates, are then said to give the correct analysis of modal operators. Lewis's semantics incorporate the idea that possible individuals inhabit only one world with the idea that intensional locutions concerning these possible individuals are represented in terms of their counterparts at other worlds. Lewis rejects the idea that we can take the aggregate of an individual and its counterparts to constitute a transworld individual in the usual sense of standard modal semantics. Counterpart theory differs from these standard semantics in that the relation between an individual and its counterparts is not an equivalence relation.

* I am especially grateful to Professor M.J. Cresswell and Dr. Edwin Mares for their generous help, support and encouragement.

Instead the relation between an object and its counterparts is said to be one of similarity and thus is, in general, neither transitive nor symmetric.

In the literature, however, the semantical issues involved in counterpart theory have often been overshadowed by the metaphysical concerns associated with arguments for and against transworld identity. In what follows I examine counterpart theory to see whether it does capture the semantics for intensional locutions. From the semantics that we require for our intensional language, we can then ascertain the consequences for counterpart theory.

To assess the semantic adequacy of a modal logic one must consider the class of frames which characterize it. According to Lewis (1986; 20), counterpart theory does give us the correct semantic interpretation because the metaphysics support the postulation of certain frames and, since these frames exist, they give correct interpretations of the modal operators. These are the grounds upon which Lewis (1968) bases his rejection of such controversial modal principles as the Barcan Formula.

On the other hand, Lewis (1986; 40) himself acknowledges that "[t]he right [semantic] interpretation, for us, is the one that specifies truth conditions under which we are indeed truthful and do indeed rely on one another's truthfulness." In the following formal analysis of the semantics of counterpart theory I focus on modal principles which are almost universally accepted as valid. These are principles of regular and normal modal logics - K, $\Box(\alpha \wedge \beta) \supset \Box\beta$ and universal instantiation (UI) - which a formal semantics should validate if it is to be a contender for a semantics of our natural intensional languages. I show that counterpart theory does not validate these principles.

Counterpart Theory

The basis of the logical system for counterpart theory involves the introduction of primitive predicates and postulates to the lower predicate calculus. Lewis (1968; 113) uses the following primitive predicates,[1]

(1) Wx x is a possible world
(2) xIy x is in possible world y
(3) Ax x is actual
(4) xCy y is a counterpart of x

[1] My notation varies slightly from that of Lewis (1968). In particular, note that for the counterpart predicate Lewis understands Cxy to mean that x is a counterpart of y.

Lewis's postulates encapsulate the principles of the semantics of counterpart theory. Most especially we note that nothing is in two worlds. We also note that anything in a world is a counterpart of itself. Thus the counterpart relation is reflexive. The following discussion meets the requirement that nothing is a counterpart of anything else in its own world although it need not presuppose this postulate. Lewis (1986; 214) remarks that while the postulate that nothing is a counterpart of anything else in its own world is a feature of some counterpart relations, such a restriction on the counterpart relation constitutes giving up some of the built-in flexibility of counterpart theory. Counterpart theory also involves the extension of the domain of quantification to contain every possible world and everything in every possible world.

Lewis (1968; 118) provides a translation scheme whereby every well-formed formula α of quantified modal logic can be translated into the logic of counterpart theory. This is shown in (5) below.[2]

(5) α^w (α holds in world w) is defined as follows,

a. α^w is α where α is atomic

b. $(\sim\alpha)^w$ is $\sim\alpha^w$

c. $(\alpha \wedge \beta)^w$ is $(\alpha^w \wedge \beta^w)$

d. $(\alpha \vee \beta)^w$ is $(\alpha^w \vee \beta^w)$

e. $(\alpha \supset \beta)^w$ is $(\alpha^w \supset \beta^w)$

f. $(\alpha \equiv \beta)^w$ is $(\alpha^w \equiv \beta^w)$

g. $(\forall x\alpha)^w$ is $\forall x(xIw \supset \alpha^w)$

h. $(\exists x\alpha)^w$ is $\exists x(xIw \wedge \alpha^w)$

i. Where $x_1, \ldots, x_n$ are the variables free in α and $y_1, \ldots, y_n$ are the variables which uniformly replace them, $(\Box\alpha x_1 \ldots x_n)^w$ is $\forall w'$ $\forall y_1 \ldots \forall y_n ((Ww' \wedge y_1 Iw' \wedge x_1 C y_1 \wedge \ldots \wedge y_n Iw' \wedge x_n C y_n) \supset \alpha^{w'} y_1 \ldots y_n)$

j. Where $x_1, \ldots, x_n$ are the variables free in α and $y_1, \ldots, y_n$ are the variables which uniformly replace them, $(\Diamond\alpha x_1 \ldots x_n)^w$ is $\exists w'$ $\exists y_1 \ldots \exists y_n (Ww' \wedge y_1 Iw' \wedge x_1 C y_1 \wedge \ldots \wedge y_n Iw' \wedge x_n C y_n \wedge \alpha^{w'} y_1 \ldots y_n)$

2 In addition to this translation scheme Lewis privileges the actual world in a way which is not relevant here and which is not consonant with his later work.

Why Counterpart Theory is not a Semantics for a Normal (or even Regular) Modal Logic

In this section I show the failure in counterpart theory of basic modal principles such as K, $\Box(\alpha \wedge \beta) \supset \Box\beta$ and the substitution of tautologous equivalents. These results show that counterpart theory is not a semantics for a normal modal logic. More fundamentally, since we cannot obtain $\Box(\alpha \wedge \beta) \supset \Box\beta$ from the valid $(\alpha \wedge \beta) \supset \beta$, counterpart theory is not a semantics for a regular logic.[3]

Counterpart theory allows that there are worlds in which an object has no counterparts. And the translation scheme for $\Box$ stipulates that those worlds where some x free in α does not have a counterpart y are irrelevant to the analysis of $\Box\alpha$. So counterpart theory will invalidate certain principles of modal logic where the antecedent and the consequent are in the scope of distinct $\Box$'*s* and the antecedent contains more (or different) well-formed subformulae than the consequent. To show that counterpart theory is neither normal nor regular it suffices to show the invalidity of $\Box(\alpha \wedge \beta) \supset \Box\beta$. Take as an example of $\Box(\alpha \wedge \beta) \supset \Box\beta$, $(\Box(\phi x \wedge \psi y) \supset \Box\psi y)^w$, which translates into counterpart theory as (6) below.

(6) $\forall w' \forall x' \forall y'((Ww' \wedge x'Iw' \wedge xCx' \wedge y'Iw' \wedge yCy') \supset (\phi x' \wedge \psi y')) \supset$
$\forall w' \forall y'((Ww' \wedge y'Iw' \wedge yCy') \supset \psi y')$

The invalidity of (6) is proved as follows.

(7) $(\Box(\phi x \wedge \psi y) \supset \Box\psi y)^w$ is invalid according to counterpart theory.

Proof: The falsifying model consists of two worlds w_1 and w_2, with a_1 and b_1 the only individuals in w_1 and b_2 the only individual in w_2. Let b_2 be the counterpart of b_1, but let a_1 have no counterpart at all in w_2. Then, where w, w', x, y and y' are assigned, respectively, w_1, w_2, a_1, b_1 and b_2, we have:

Ww, *Ww'*
xIw, *yIw*, *y'Iw'*
xCx, *yCy*, *y'Cy'*, *yCy'*

[3] Hazen (1979; 326-327) alludes to these problems when he notes that counterpart theory does not validate the inference from $\Box\phi\, ab$ to $\Box\exists x \phi\, ax$ (my notation with a, b as individual constants). Hazen (1979; 327) remarks that in counterpart theory we have to abandon the pattern in modal reasoning whereby "a conclusion, validly derived from premises that are themselves asserted to be necessarily true, is necessary."

ϕx, ψy, $\psi y'$

$\forall w'\forall x'\forall y'((Ww'\wedge x'Iw'\wedge xCx'\wedge y'Iw'\wedge yCy')\supset(\phi x'\wedge \psi y'))$ [ie $(\Box\ (\phi x\wedge \psi y))^{w}$] is true because both ϕx and ψy hold in w and because a_1 has no other counterpart than itself (in particular a_1 does not have a counterpart in w_2) and b_2 has no counterpart in w_1 other than itself.

$\forall w'\forall y'((Ww'\wedge y'Iw'\wedge yCy')\supset \psi y')$ [ie $(\Box\psi y)^{w}$] is false because yIw', ycy' yet $\psi y'$ does not hold in w'.

Thus $\Box(\phi x\wedge \psi y)\supset\Box\psi y$ is invalid in counterpart theory. ■

The model given in (7) can also be used to falsify $\Box((\phi x\supset\phi x)\wedge\psi y)\supset\Box\psi y$. Thus we cannot substitute $((\phi x\supset\phi x)\wedge\psi y)$ into the antecedent of $\Box\psi y\supset\Box\psi y$.

Multiple Counterparts and Universal Quantification

Counterpart theory allows that an individual may have more than one counterpart in a given world. In consequence, Lewis (1968; 124) notes that $(x=y)\supset\Box(x=y)$ is not a theorem of counterpart theory. However the necessity of identity is not the only principle at stake. The following is an instance of UI.

(8) $\forall y\sim\Box\phi xy\supset\sim\Box\phi xx$

The counterpart theoretic translation of (8) is given as follows.

(9) $\forall y(yIw\supset\sim\forall w'\forall x'\forall y'((Ww'\wedge x'Iw'\wedge xCx'\wedge y'Iw'\wedge yCy')\supset\phi x'y'))$
$\supset\sim\forall w'\forall x'((Ww'\wedge x'Iw'\wedge xCx')\supset\phi x'x')$

The invalidity of (9) is shown in (10) below.[4]

[4] I thank Max Cresswell for pointing out the invalidity of (9) to me.

(10) $(\forall y \sim \Box\phi\, xy \supset \sim \Box\phi\, xx)^{w}$ is invalid according to counterpart theory.

Proof: The falsifying model consists of two worlds w_1 and w_2, with a_1 the only individual in w_1 and a_2 and b_2 the only individuals in w_2. Let a_2 and b_2 both be counterparts of a_1. Then, where w, w', x, y, x', y' are assigned, respectively, w_1, w_2, a_1, a_1, a_2 and b_2, we have:

Ww, Ww'
xIw, $x'Iw'$, $y'Iw'$
xCx, $x'Cx'$, $y'Cy'$, xCx', xCy', yCx', yCy'
Define ϕ so that $\phi\, xx$, $\phi\, x'\, x'$, $\phi\, y'\, y'$, $\sim \phi\, x'\, y'$

Then $(Ww' \wedge x'Iw' \wedge xCx' \wedge y'Iw' \wedge yCy') \supset \phi\, x'y'$ is false. Thus $\sim \forall w'\, \forall x'\, \forall y'\, ((Ww' \wedge x'\, Iw' \wedge xCx' \wedge y'\, Iw' \wedge yCy') \supset \phi\, x'\, y')$ is true. Since a_1 is the only possible assignment to y of an individual in w_1 we have that $\forall y(yIw \supset \sim \forall w'\, \forall x'\, \forall y'\, ((Ww' \wedge x'\, Iw' \wedge xCx' \wedge y'\, Iw' \wedge yCy') \supset \phi\, x'y'))$ [ie $(\forall y \sim \Box\phi xy)^{w}$] is true.

$\forall w' \forall x' ((Ww' \wedge x'Iw' \wedge xCx') \supset \phi\, x'x')$ is true because $\phi\, xx$ holds in w and both $\phi\, x'x'$ and $\phi\, y'y'$ hold in w'. So $\sim \forall w'\, \forall x'\, ((Ww' \wedge x'Iw' \wedge xCx') \supset \phi\, x'x')$ is false. So (9) is invalid. ■

There is, however, something odd about (9) treated as an instance of UI. It is the correct translation of (8) according to Lewis's principles, but it is not itself an instance of UI in the language of counterpart theory. That is, it is not an instance of the scheme $\forall x\alpha \supset \alpha[y/x]$. Thus, it might seem a good idea to reformulate the principles of translation so as to translate (8), and modal formulae like it, as proper instances of UI in the language of counterpart theory. If we do so, it would seem that we should translate (8) as follows.

(11) $$\forall y(yIw \supset \sim \forall w'\, \forall x'\, \forall y'\, ((Ww' \wedge x'\, Iw' \wedge xCx' \wedge y'\, Iw' \wedge yCy') \supset \phi\, x'\, y'))$$
$$\supset (xIw \supset \sim \forall w'\, \forall x'\, \forall y'\, ((Ww' \wedge x'\, Iw' \wedge xCx' \wedge y'\, Iw' \wedge xCy') \supset \phi\, x'\, y'))$$

(11) is an instance of UI and is in fact valid. But it does not satisfy Lewis's translation scheme for $\Box$. So the problem is that the relevant versions of UI in the language of counterpart theory and in modal logic do not correspond. The consequent of (9) is shown below as (12) as opposed to the consequent of (11) which is shown below as (13).

(12) $\sim \forall w' \forall x' \big((Ww' \wedge x' Iw' \wedge xCx') \supset \phi x' x' \big)$

(13) $\big(xIw \supset \sim \forall w' \forall x' \forall y' \big((Ww' \wedge x' Iw' \wedge xCx' \wedge y' Iw' \wedge yCy') \supset \phi x' y' \big) \big)$

Lewis's translation scheme for $\Box$ is given in (14) below.[5]

(14) Where $x_1, \ldots, x_n$ are the variables free in α and $y_1, \ldots, y_n$ are the variables which uniformly replace them, $(\Box \alpha x_1 \ldots x_n)^w$ is $\forall w' \forall y_1 \ldots \forall y_n \big((Ww' \wedge y_1 Iw' \wedge x_1 Cy_1 \wedge \ldots \wedge y_n Iw' \wedge x_n Cy_n) \supset \alpha^{w'} y_1 \ldots y_n \big)$

Using this translation scheme, (12) is the translation of $(\sim \Box \phi\, xx)^w$ (13) can be obtained from $(\sim \Box \phi\, xx)^w$ by a translation scheme for $\Box$ such as the following.

(15) Where $x_1, \ldots, x_n$ are all the occurrences of free variables in α and $y_1, \ldots, y_n$ are the variables replacing them, $(\Box \alpha x_1 \ldots x_n)^w$ is $\forall w' \forall y_1 \ldots \forall y_n \big((Ww' \wedge y_1 Iw' \wedge x_1 Cy_1 \wedge \ldots \wedge y_n Iw' \wedge x_n Cy_n) \supset \alpha^{w'} y_1 \ldots y_n \big)$

But even if we adopt (15) as the translation scheme for $\Box$ problems appear. If (15) were the translation scheme for $\Box$ then $\Box(\alpha \vee \sim\alpha)$ would be invalid in counterpart theory.[6] This is shown in (16) below.

(16) $\Box(\phi\, x \vee \sim \phi\, x)$ is invalid according to counterpart theory with translation scheme (15) replacing (14).

Proof: The falsifying model consists of two worlds w_1 and w_2, with a_1 the only individual in w_1 and a_2 and b_2 the only individuals in w_2. Let a_2 and b_2 both be counterparts of a_1. Then where w, w', x, x', and y' are assigned, respectively, w_1, w_2, a_1, a_2 and b_2 we have:

Ww, Ww'
xIw, $x'Iw'$, $y'Iw'$
xCx, $x'Cx'$, $y'Cy'$, xCx', xCy'

[5] Although Lewis in not explicit about this, the translation scheme for $\Box$ is presumably to be taken in such a way that the same variable is treated in the same way wherever it occurs in the well-formed formula α. Lewis (1968; 124) remarks that the translation of the converse Barcan Formula is a theorem. And the translation of $\Box \forall x(\phi\, x \vee \sim \phi\, x) \supset \forall x \Box (\phi\, x \vee \sim \phi\, x)$ requires uniform replacement of free variables inside the $\Box$ operator if it is going to be valid.

[6] It would also result in a difference between 'x is a self-admirer' and 'x admires x' if *self-admirer* is a one-place predicate.

$\phi x, \sim \phi x', \phi y'$

$(\Box(\phi x \vee \sim \phi x))^{w}$ is $\forall w' \forall x' \forall y' ((Ww' \wedge x' Iw' \wedge xCx' \wedge y' Iw' \wedge xCy') \supset (\phi x' \vee \sim \phi y'))$. This is false because Ww', $x' Iw'$, xCx', $y' Iw'$, xCy' yet $\sim (\phi x' \vee \sim \phi y')$ ■

By using Lewis's translation scheme (14) we obtain the following translation of $\Box(\phi x \vee \sim \phi x)$.

(17) $\forall w' \forall x' ((Ww' \wedge x' Iw' \wedge xCx') \supset (\phi x' \vee \sim \phi x'))$

(17) is valid in counterpart theory. Its negation is the counterpart-theoretic translation of $\sim \Box(\phi x' \vee \sim \phi x')$. (18) below is the counterpart-theoretic translation of $\forall y \sim \Box(\phi x \vee \sim \phi y) \supset \sim \Box(\phi x \vee \sim \phi x)$.

(18) $\forall y(yIw \supset \sim \forall w' \forall x' \forall y' ((Ww' \wedge x' Iw' \wedge xCx' \wedge y' Iw' \wedge yCy') \supset (\phi x' \vee \phi y'))) \supset \sim \forall w' \forall x' ((Ww' \wedge x' Iw' \wedge xCx') \supset (\phi x' \vee \sim \phi x'))$

(18) is invalid. So UI fails even for an intensional language with just a single one-place predicate.

One goal of counterpart theory is to provide an interpretation of our standard quantified logic which represents modality (Lewis 1968; 113). The two most obvious counterpart-theoretic candidates for representing modality are those structures produced by the translation schemes given in (14) and (15). Instead of thinking of them as rival translation schemes think of them as rival ways of representing modality in an extensional logic. Either each free variable is assigned a counterpart or each free occurrence of a variable is assigned a counterpart.

Lewis may want to argue that the quantification hidden in the formulae of modal logic means that what appear to be instances of UI in modal logic are not in fact so. But even though (9) is not an instance of UI, under the interpretation given by (14) it represents an instance of UI. Its invalidity shows that uniform replacement of free variables does not adequately represent having more than one counterpart in another world.

On the other hand, while (15) seems to capture the idea that one thing can have more than one counterpart in another world, the replacement of different occurrences of the same variable with different variables means that $\Box(\alpha \vee \sim \alpha)$ is invalid. That x has more than one counterpart in another world w' should mean that $(\sim \Box x = x)^{w}$ because x has a counterpart at w' with whom it is possibly non-identical. And $\Box x = x$ is invalidated by (15), although valid according to Lewis's (14). Thus Lewis may want to accept (15) and the invalidity of $\Box(\phi x \vee \sim \phi x)$ and reject the necessity of the principle of excluded middle. Either way, $\exists y \Box(\phi x \vee \sim \phi x)$ is invalid since its translation by (14) and its translation by (15) each correspond to the following.

(19) $\exists y(yIw \wedge \forall w' \forall x' \forall y'((Ww' \wedge x'Iw' \wedge xCx' \wedge y'Iw' \wedge yCy') \supset (\phi x' \vee \sim \phi y')))$

So whether we endorse the analysis of $\Box$ given by (14) or the analysis given by (15), as long as an object can have multiple counterparts (19) is invalid.

Victoria University of Wellington

REFERENCES

Hazen, Allen. (1979) Counterpart-theoretic semantics for modal logic. *Journal of Philosophy*. 76: 319-338.
Lewis, David K. (1968) Counterpart theory and quantified modal logic. *Journal of Philosophy*. 65: 113-126.
Lewis, David K. (1986) *On the Plurality of Worlds*. Oxford: Basil Blackwell.