13 Counterpart Theory

13.1 Philosophical Preliminaries

As originally presented in 1968, David Lewis's counterpart theory represents a way to translate the langauge of quantified modal logic into a first-order logical theory (one with classical quantifier rules) without any intensional operators, but in which, instead, we quantify over worlds in the object language, and introduce additional predicates for such things as inclusion within worlds, and the *counterpart* relation for objects in different worlds.

Lewis regards possible worlds as genuine entities, and the things in them as well. He accepts a theory known as *modal realism* according to which possible objects, although not actual, are taken to be genuinely real things in the full sense. Considered as a formal theory, one need not accept Lewis's metaphysics of modality to accept CT; however, one must maintain that it is somehow intelligible to quantify over worlds and non-actual objects.

(Of course, we do that in the *metalanguage* when we do Garson-style semantics for modal logic; however, it is arguable that the commitment to possible worlds in accepting a theory like fS5 is at least much more oblique. Again considered as a formal system, one might accept fS5 and think of modality as a *sui generis* notion. Garson's models are then perhaps heuristics or constructs that can be used to define a notion of validity for a logic with modal operators that is extensionally right, without showing the true nature of necessity or possibility.)

Lewis's counterpart relation C is paramount his way of thinking of *de re* modality. In standard (Kripkean) objectual semantics for modal logics, whether or not it is possible for me to have a property I don't have, or whether it is possible for me not to have some property I do have, depends on what properties *I* have in other possible worlds. This presupposes the intelligibility of *trans-world identity*, the same thing (me) existing at multiple possible worlds. Lewis instead understands de re modal questions in terms of the properties had by my *counterparts* in other worlds, people who closely resemble me in some relevant sense. It is possible for me to have worn a different color shirt today because in some other world there is someone qualitatively exactly like me, except wearing a different color shirt (sitting in a room qualitatively like this one, teaching students very similar to you, etc.)

13.2 Syntax of CT

Counterpart theory does not make use of individual constants. (This requires slightly adjusting the inference rules so that "arbitrary" constants are not needed for (Q∀In); instead *free* variables are used. Some of you may already be familiar with the use of free variables in proofs.)

Counterpart theory (CT) makes use of the following predicates along with the intended meaning listed:

Wx	x is a world
Ixy	x is in (world) y
Ax	x is actual (exists in the actual world)
Cxy	x is a counterpart of y

Lewis also adopts the following definition:

@ *abbreviates* $\imath$x ∀y(Iyx ↔ Ay)

Here, @ is the actual world, or the world containing all and only actual things. Lewis defines this with a description; but it seems clear that he accepts the Russellian method for eliminating descriptions in context. (He should have given us conventions for resolving scope ambiguities for @, but in virtue of (P7) below, such ambiguities won't matter much.

Lewis might have instead taken @ as the theory's own individual constant and defined "Ax" instead as "Ix@".

13.3 Axioms of CT

System CT is built on pure first-order logic with identity (roughly, Garson's QL, modulo the lack of individual constants), and adds the following particular axioms:

(P1) ∀x ∀y(Ixy → Wy)
(P2) ∀x ∀y ∀z(Ixy & Ixz → y = z)
(P3) ∀x ∀y(Cxy → ∃z Ixz)
(P4) ∀x ∀y(Cxy → ∃z Iyz)
(P5) ∀x ∀y ∀z(Ixy & Izy & Cxz → x = z)
(P6) ∀x ∀y(Ixy → Cxx)
(P7) ∃x(Wx & ∀y(Iyx ↔ Ay))
(P8) ∃x Ax

(P1) says that things are only in worlds. (P2) says that nothing is in more than one world. This is sometimes called the principle of *world-bound individuals*, and is in effect Lewis's rejection of transworld identity, strictly speaking.

(P3) and (P4) together say that all counterparts are in worlds. (P5) and (P6) together say anything in any world has itself, and *only* itself, as its counterpart in that world.

(P7) says that there is a world that contains all and only actual things; it is this axiom that allows us to introduce "@" by description (without fear that the existential clause of the contextual definition will be false). (P8) says that something is actual.

Lewis considers but rejects certain additional principles.

$\nvdash_{\text{CT}}$ $\forall x \, \forall y (\text{C}xy \rightarrow \text{C}yx)$
(The counterpart relation need not be symmetric.)

$\nvdash_{\text{CT}}$ $\forall x \, \forall y \, \forall y (\text{C}xy \, \& \, \text{C}yz \rightarrow \text{C}xz)$
(The counterpart relation need not be transitive.)

$\nvdash_{\text{CT}}$ $\forall x \, \forall y \, \forall z \, \forall w (\text{C}xy \, \& \, \text{C}xz \, \& \, \text{I}yw \, \& \, \text{I}zw \rightarrow y = z)$
(Something needn't be counterpart to only one thing in a world.)

$\nvdash_{\text{CT}}$ $\forall x \, \forall y \, \forall z (\text{C}yx \, \& \, \text{C}zx \, \& \, \text{I}yw \, \& \, \text{I}zw \rightarrow y = z)$
(Something needn't have only one counterpart in a world.)

$\nvdash_{\text{CT}}$ $\forall x \, \forall w \, \forall w' (\text{W}w \, \& \, \text{W}w' \, \& \, \text{I}xw \rightarrow \exists y(\text{I}yw' \, \& \, \text{C}yx))$
(Something needn't have a counterpart in every other world.)

$\nvdash_{\text{CT}}$ $\forall x \, \forall w \, \forall w' (\text{W}w \, \& \, \text{W}w' \, \& \, \text{I}xw \rightarrow \exists y(\text{I}yw' \, \& \, \text{C}xy))$
(Something needn't be the counterpart of something in every other world.)

13.4 Translating from QML to CT

Lewis provides a means for translating sentences stated in the language of quantified modal logic (one that adds only "□", or it with "◇" to a standard first-order predicate language), such as Garson's qK or fK or their extensions, into the language of CT.

At the core of a translation is a function, which we may write $|A|^w$, where A is formula and w is a variable for worlds. We can define this recursively as follows:

1. If A is an atomic sentence or "⊥", then $|A|^w$ is A itself.

2. If A takes the form $(B \rightarrow C)$, then $|A|^w$ is $(|B|^w \rightarrow |C|^w)$.

(It then follows from their definitions that if A takes the form $(B \, \& \, C)$, then $|A|^w$ is $(|B|^w \, \& \, |C|^w)$, and so on for $\sim$, $\vee$ and $\leftrightarrow$.)

3. If A takes the form $\forall x \, Bx$, then $|A|^w$ is $\forall x(\text{I}xw \rightarrow |Bx|^w)$.

(It follows that if A takes the form $\exists x \, Bx$ then $|A|^w$ is $\exists x(\text{I}xw \, \& \, |Bx|^w)$.)

4. If A takes the form $\Box B$, and the free variables of B are $x_1, \ldots, x_n$, then $|A|^w$ is $\forall w' \forall y_1 \ldots \forall y_n(\text{W}w' \, \& \, \text{I}y_1 w' \, \& \, \ldots \, \& \, \text{I}y_n w' \, \& \, \text{C}y_1 x_1 \, \& \, \ldots \, \& \, \text{C}y_n x_n \rightarrow |[\ldots [B]^{y_1}/x_1 \ldots]^{y_n}/x_n|^{w'})$.

(It follows that if A takes the form $\Diamond B$, and the free variables of B are $x_1, \ldots, x_n$, then $|A|^w$ is $\exists w' \exists y_1 \ldots \exists y_n(\text{W}w' \, \& \, \text{I}y_1 w' \, \& \, \ldots \, \& \, \text{I}y_n w' \, \& \, \text{C}y_1 x_1 \, \& \, \ldots \, \& \, \text{C}y_n x_n \, \& \, |[\ldots [B]^{y_1}/x_1 \ldots]^{y_n}/x_n|^{w'})$.)

- Hence, if B is a sentence without free variables, then $|\Box B|^w$ is $\forall w'(\text{W}w' \rightarrow |B|^{w'})$.

- If Bx has one free variable, x, then $|\Box Bx|^w$ is $\forall w' \forall y(\text{W}w' \, \& \, \text{I}yw' \, \& \, \text{C}yx \rightarrow |By|^{w'})$.

$|A|^w$ is basically the translation of A relativized to world w. The complete translation of a sentence A of QML in Lewis's CT is A relativized to the actual world $|A|^@$, or more fully $\exists w(\forall x(\text{I}xw \leftrightarrow \text{A}x) \, \& \, |A|^w)$. (Don't confuse the schematic "A", italics, with the predicate "A" of Lewis's langauge, no italics.)

Examples:

- The translation of "$\forall x \, \text{Fx}$" is "$\forall x(\text{I}x@ \rightarrow \text{Fx})$".

- The translation of "$\Diamond \exists x \, \text{Rxx}$" is "$\exists w(\text{W}w \, \& \, \exists x(\text{I}xw \, \& \, \text{Rxx}))$".

- The translation of "$\exists x \, \Diamond \, \text{Fx}$" is "$\exists x(\text{I}x@ \, \& \, \exists w \, \exists y(\text{W}w \, \& \, \text{I}yw \, \& \, \text{C}yx \, \& \, \text{Fy}))$."

HOMEWORK: Translate the following from QML to CT:

1. $\Box \, \forall x \, \text{Fx}$

2. $\forall x \, \Box \, \text{Fx}$

3. $\forall x \, \Diamond \, \exists y \, \text{Ryx}$

4. $\Box \, \forall x(\text{Fx} \rightarrow \Diamond \, \text{Gx})$

Translation in the other direction (from CT to QML), however, is not always possible.

13.5   Comparison with other systems

- While the rules for the quantifiers are classical, and have a univocal domain, ranging over all possibilia, quantification is often restricted to actual things, or all those in some world, so in practice, the theory has some of the features of a free logic. However, it has no genuine terms, and *a fortiori*, none that don't refer to an existent thing.

- The *de re/de dicto* distinction is captured by treating all individually referring expressions according to Russell's theory of descriptions and applying the contextual definitions accordingly.

- In some ways the modality espoused is close to that of S5, since it in effect defines necessity in terms of truth in *all* possible worlds whatever.

  In line with this, for every A which is a *closed* sentence, the translations of the appropriate instances of the following are theorems:
  (M) $\Box A \to A$
  (B) $A \to \Box \Diamond A$
  (4) $\Box A \to \Box \Box A$
  (5) $\Diamond A \to \Box \Diamond A$
  However, these are restricted to closed sentences. Since Lewis's system is without constants, and proofs often proceed using free variables, having the above is not quite as strong as having Garson's analogous principles. In Garson's system, for example, (Q∀In) allows you to go from $\Box Fa \to \Box \Box Fa$ to $\forall x(\Box Fx \to \Box \Box Fx)$. Lewis's system does not work thsi way, and does not have the translation of $\forall x(\Box Fx \to \Box \Box Fx)$ as a theorem schema. In Lewis's system, whether a *de re* modal claim is necessarily necessary depends on the counterparts of one's counterparts; since these are not always one's own counterparts, the "quantified into" version of (4) does not hold.

  You could perhaps state this by saying that the rules of S5 apply to *de dicto* modalities for Lewis, but not to *de re* modalities. One could in effect regain S5 by taking the counterpart relation to be an equivalence relation.

- An analogue of Garson's (RC) cannot easily be formulated in Lewis's system due to the lack of constants. However, Lewis does not accept the translation of $\forall x \, \forall y(x = y \to \Box \, x = y)$, since it would require that no two things in the actual world can have the same counterpart, which (as we've seen), he denies. (Similar considerations apply to $\forall x \, \forall y(x \neq y \to \Box \, x \neq y)$.

- We translation of the Converse Barcan Formula:

  $$\text{(CBF)} \quad \Box \, \forall x \, Ax \to \forall x \, \Box \, Ax$$

  is a theorem of CT, but not the translation of the Barcan Formula:

  $$\text{(BF)} \quad \forall x \, \Box \, Ax \to \Box \, \forall x \, Ax$$

  The reason the latter fails is that there may be objects in other possible worlds that are not the counterparts of any objects in this world. (As we saw, Lewis rejects the idea that every object in a given world must be the counterpart of anything in a given other world.)

  Most difficulties with (CBF), moreover, are diffused by the nature of the system as a whole.

13.6   Possible Expansions/ Modifications

As Lewis points out, counterpart theory could also be made to proxy for logics with other kinds of intensional or modal operators (deontic operators, epistemic modal operators) by defining additional relations between possible worlds in CT to act as limited accessibility relations between worlds, and give a translation scheme for sentences using such an operator in terms of it.

For example, if "Rxy" meant that world x represented a morally permissible world to world y, then something of the form $\mathbf{O} \, A$ could be defined in CT as $\forall w(\text{W}w \, \& \, \text{R}w@ \to |A|^w)$ and $\mathbf{P} A$ as $\exists w(\text{W}w \, \& \, \text{R}w@ \, \& \, |A|^w)$, etc. Lewis gives more detailed methodology for this in section V of the paper. It would even be fairly easy to mix and match different kinds of such operators together in the same context just by using different relations.

The basic approach of eschewing modal operators in favor of quantifying over worlds in the object language, along with introducing relations between worlds, and between things and worlds, can be adapted and modified and made to work even with very different views about the right way to formalize the details. For example, we might have such a system that allows for transworld identity or overlap, but adds an index to each predicate to relativize the predicate to a given world, or something similar. Or one might formulate a counterpart theory in which the counterpart relation is held to be transitive or symmetric, etc., and yield different results.