The Logic of Counterpart Theory with Actuality

Adam Rigoni and Richmond H. Thomason
Philosophy Department, University of Michigan

Journal of Philosophical Logic, forthcoming.
Final publication available at www.springerlink.com.

Abstract

It has been claimed that counterpart theory cannot support a theory of actuality without rendering obviously invalid formulas valid or obviously valid formulas invalid. We argue that these claims are not based on logical flaws of counterpart theory itself, but point to the lack of appropriate devices in first-order logic for "remembering" the values of variables. We formulate a mildly dynamic version of first-order logic with appropriate memory devices and show how to base a version of counterpart theory with actuality on this. This theory is, in special cases, equivalent to modal first-order logic with actuality, and apparently does not suffer from the logical flaws that have been mentioned in the literature.

1. Introduction

Since shortly after its inception in [Lewis, 1968], and continuing to the present day, counterpart theory has tended to draw fire. We divide the various criticisms directed at it into three categories. The first and earliest criticisms construe the formulas of counterpart theory as making direct metaphysical claims, which they then dispute. Allen Hazen, in [Hazen, 1976] rightly dismisses these criticisms as somewhat naive. Other criticisms treat counterpart theory as a formulation of first-order modal logic and compare it with more standard formulations that are based on Kripke frames and postulate individuals that are already individuated across worlds.[1] Here there is a legitimate area of debate. Some formulas that counterpart theory renders satisfiable are implausible on logical grounds. A number of examples are cited in [Hazen, 1976]; also see [Cresswell, 2004]. But this debate seems inconclusive. From the beginning, first-order modal logic has proved to be intuitively challenging from a logical standpoint, and—as usual in such cases—has produced a variety of alternative logics, making it difficult to draw definitive conclusions. Yet counterpart theory has some logical advantages: because of its flexibility, it can be a useful tool in investigating combinations of modality with first-order quantification.[2]

Third, some brief remarks of Hazen's about the interaction between actuality and counterpart theory [Hazen, 1976, p. 330] inspired several proposals for revising Lewis' translation

[1]See [Garson, 1984, Braüner and Ghilardi, 2006] for general information concerning logical approaches to first-order modal logic.

[2]See, for instance, [Corsi, 2002].

from modal to first-order logic in order to accommodate an actuality operator.[3] The realization that attempts to carry out such a revision apparently are flawed led to an extensive review of the difficulties in [Fara and Williamson, 2005]. The authors of this paper conclude:

> there is no coherent way to extend Lewis' scheme for translation from the language of quantified modal logic to the language of counterpart theory, if quantified modal logic is regarded, as it should be, as containing an actuality operator.
> [Fara and Williamson, 2005, p. 453]

We think that this conclusion is premature, that it presupposes a somewhat superficial assessment of the challenge posed by combining actuality with counterpart theory, and that, in engaging a problem that is primarily logical, it displays a pernicious insensitivity to logical methodology. We begin by expanding on the methodological points, and then describe a logical project that, we claim, leads to a reasonable accommodation of actuality in something that is very like Lewis' framework, although it does require a nontrivial extension of the underlying logic. Some features of the resulting logic may be controversial, but as far as we can see these features can be attributed to counterpart theory itself, rather than to any special limitations of the counterpart approach that are uniquely incompatible with an actuality operator. Therefore, although metaphysical objections to counterpart theory may remain, objections on purely logical grounds seem to fail.

2. Methodological remarks and a proposal

Logic has become a branch of mathematics, so it inherits the methodology of mathematics. In mathematics, general questions are settled positively, by providing a rigorous proof covering all possible cases, or negatively, by providing a counterexample. Consider the problem of whether there is a general method for trisecting an angle with ruler and compass. A history of failed attempts to provide such a method goes back to antiquity. Despite this history of failures, the mathematical question was considered to be open until 1837. In fact, it became one of the most important open questions in geometry. The accumulation of failed attempts did not in any way settle the question, and the impossibility of a general method was only established when Pierre Wantzel devised an algebraic representation of ruler and compass constructions. Wantzel's proof introduces something fundamentally new—the algebraic representation—an idea that is entirely missing from the record of failed attempts.

In effect, Fara and Williamson supply a list of failed attempts to combine actuality and counterpart theory, and conclude from this that no adequate formulation of this extension of counterpart theory exists. If the problem is mathematical, this is not appropriate methodology. On the other hand, if the problem cannot be formulated mathematically, because the notion of an "adequate formulation" is hopelessly imprecise, then of course there would be no way to find a provably correct way of adding actuality to counterpart theory. But for the same reason, proving that such a formulation is impossible would be hopeless. In fact, in

[3] See [Forbes, 1982, Ramachandran, 1989].

this case it is hard to see the point of concluding or conjecturing that there is no coherent way of reconciling counterpart theory with actuality.

Is the problem hopelessly imprecise? We believe it is not. The difficulty here is to make the notion of a coherent extension of counterpart theory precise enough so that an impossibility theorem can be proved or a coherent extension provided. The propositional case of the problem is easy to formalize, using generally accepted theories of the logic of actuality such as that of [Hodes, 1984]. The first-order case is more challenging, due to the fact that some logicians feel that first-order counterpart theory without actuality is in some sense incoherent. But even this case may be tractable, since the question is *not* whether first-order counterpart theory is incoherent, but whether the addition of actuality makes first-order counterpart theory incoherent in new and significantly different ways. Making such a differential notion of coherence precise may not be as difficult as trying to settle the vexing question of under what conditions a first-order modal logic is coherent.

In this paper, we work with a fairly weak notion of differential coherence: that a version of counterpart theory with actuality should be equivalent to standard first-order modal logic with actuality when the counterpart relation produces one and only one counterpart of each individual in each world. It is an open question whether there are appropriate ways to strengthen this notion, and whether counterpart theory can support an actuality operator that respects these strengthened notions.

The problem of counterpart theory with actuality is a special case of a familiar logical problem: how to extend a given logic by enriching its language. Typically, the set of intended models of the base logic is well defined. The central problem is then to characterize the models of the extended logic in a way that naturally and conservatively extends the original models and that does justice to whatever intuitions are available. Sometimes the problem is trivial; this is what happens when the new constructs are definable in the base logic. But the history of modern logic provides many instances where the interpretation of the base logic needs to be generalized in some fundamental way to accommodate the extension.

Adding first-order quantifiers to boolean propositional logic is a classical case of such a nontrivial extension. Models of propositional logic are simply assignments of truth-values to atomic formulas. This view of models is unable to deal with first-order models in which some individuals have no names. Tarski's solution to the problem was to generalize the notion of a model to make assignments of truth-values relative to variable assignments (or, as they are often called, *sequences*). As usual, this idea has many uses beyond the original one: for instance, it is crucial for algebraic logic and dynamic logic.

Adding an actuality operator to a normal modal logic (or a nowness operator to a normal tense logic) is a more germane case. An actuality operator $[@]$ cannot be a normal modality, since $A \to [@]A$ is valid but $\Box[A \to [@]A]$ is not. Worse, the model theory of normal modal logics automatically validates the Necessity Rule, according to which $\Box A$ is valid whenever A is. One might be tempted to think that this shows that actuality cannot be added to normal propositional modal logics; but that would be a hasty conclusion. As in the previous case, the model theory of modal logic needs to be generalized to enable this extension. The truth of formulas in a model needs to be relativized not to one, but to two worlds. This device, often called "double indexing," enables the semantic evaluation of a formula to "remember" the base world from which the evaluation started and relative to which actuality is to be interpreted.[4]

This train of thought suggests that an initial obstacle in adding actuality to counterpart theory—and perhaps the only obstacle—is to find a way to incorporate double-indexing in counterpart theory. But this can't be done by relativizing satisfaction to two worlds rather than one, as in modal logic. Counterpart theory does not relativize satisfaction to worlds, but adopts the quantificational apparatus of first-order logic, places worlds in the domain of the first-order quantifiers, and relativizes satisfaction to variable assignments.

First-order logic is analogous to modal logic in validating an analog of the necessity rule: $\forall x\, A$ is valid whenever A is; therefore, we should expect a problem to arise in adding actuality to counterpart theory that is similar to the one we find in normal modal logic. The natural analog in counterpart theory to modal double-indexing would be revised variable assignments that associate two individuals, rather than one, to individual variables. The hope would be that this change to the base logic would smooth the way for the addition of actuality to counterpart theory, just as the addition of double-indexing enables the addition of actuality to propositional modal logic. This, in essence, is our project.

In this paper we show that counterpart theory can be modified to deal with modal statements involving actuality, provided that the first-order basis of the theory is amended as described above. We show that this modification is invisible in the usual language of first-order logic; it affects the logic only in the presence of operators like actuality, that are two-dimensional.

We prove that the quantifier-free part of two-dimensional propositional modal logics that with actuality is equivalent to the corresponding two-dimensional counterpart theory with actuality. We go on to show that first-order two-dimensional modal logic is equivalent to a restricted fragment of the two-dimensional counterpart theory. These results, together with case-by-case examination of specific examples, provide some confidence in the adequacy of the counterpart theory.

In actualizing counterpart theory, we proceed as follows: (i) We prove that there is a version of counterpart theory (worldless counterpart theory) making no explicit reference to worlds, that is equivalent to Lewis' standard counterpart theory as formulated in [Lewis, 1968]. This is inessential to our ultimate goal, but is a valuable simplifying step. (ii) We prove that there is a natural translation from propositional modal logic into counterpart theory such that a modal formula is valid if and only if its translation is. (iii) We show how to properly introduce actuality into counterpart theory. In contrast to previous approaches (e.g., [Forbes, 1982, Hazen, 1979, Hazen, 1976, Ramachandran, 1989]), we pave the way for an actuality operator not by extensive alterations to Lewis' translation scheme but rather by modifying the satisfaction relation. These modifications parallel those of traditional two-dimensional modal logics such as that found in [Hodes, 1984], where satisfaction becomes a relation between a model, a formula, and not one but two worlds. (iv) Finally, we prove the

[4] The problem and its solution is well explained in [Kaplan, 1978].

equivalence results mentioned above.

3. Formulating counterpart theory

3.1. Motivating a simplification of counterpart theory

David Lewis' presentation of counterpart theory in [Lewis, 1968] uses a first-order theory with two designated two-place predicates $\mathsf{C}xy$ ("x is a counterpart of y") and $\mathsf{I}xy$ ("x is in the world y"), and two one-place predicates $\mathsf{A}x$ ("x is actual") and $\mathsf{W}x$ ("x is a world"). If this theory is used to formalize a modal subject matter, it will of course contain other constants as well. For definiteness, we will assume that the primitive logical operators of the theory are negation $\neg$, the conditional $\rightarrow$, and the universal quantifier $\forall$. We can assume (when useful) that the language has no individual constants, since such constants can be replaced by predicates.

Lewis provides eight postulates as a basis for counterpart theory. To these postulates, we add a further postulate (P9), saying that worlds (and only worlds) are unindividuated entities—they have no counterparts. This postulate may not be required for logical or metaphysical purposes, but it is certainly natural, and it provides a simplified ontology that makes our technical work easier.

(P1) $\forall x \forall y[\mathsf{I}xy \rightarrow \mathsf{W}y]$
(P2) $\forall x \forall y \forall z[[\mathsf{I}xy \wedge \mathsf{I}xz] \rightarrow y = z]$
(P3) $\forall x \forall y[\mathsf{C}xy \rightarrow \exists z\, \mathsf{I}xz]$
(P4) $\forall x \forall y[\mathsf{C}xy \rightarrow \exists z\, \mathsf{I}yz]$
(P5) $\forall x \forall y \forall z[[\mathsf{I}xy \wedge \mathsf{I}zy \wedge \mathsf{C}xz] \rightarrow x = z]$
(P6) $\forall x \forall y[\mathsf{I}xy \rightarrow \mathsf{C}xx]$
(P7) $\exists x[\mathsf{W}x \wedge \forall y[\mathsf{I}yx \leftrightarrow \mathsf{A}y]]$
(P8) $\exists x \mathsf{A}x$
(P9) $\forall x[\mathsf{W}x \leftrightarrow \neg \exists y\, \mathsf{C}yx]$

Let $\mathcal{L}^C$ be the language of such a theory, and M be a model of the theory on the domain D. M, an ordinary first-order model, will assign a set M_{C} of pairs of D to C, a set M_{I} of pairs of D to I, and a subset M_{A} of D to A.

Although worlds are a crucial motivating part of counterpart theory, they are dispensable for technical purposes, because information about worlds is implicitly available in the counterpart relation itself: a world can be represented by an arbitrarily chosen individual belonging to it, giving rise to a *worldless* version of counterpart theory. The models of the modified theory are much simpler and easier to work with.

Worldless counterpart theory is not fully equivalent to ordinary counterpart theory: the chief difference, of course, is that the worldless theory restricts its ontology to entities that are individuated by the counterpart relation. But, as we will show, these differences in the apparatus used to define modality do not affect the part of counterpart theory that corresponds to familiar modal logics.

Lewis' account of modality involves quantification over both worlds and counterparts; for instance,

$$\Box P\,xy$$

is expanded to

$$\forall w[\mathsf{W}w \rightarrow \forall x'\forall y'[[\mathsf{C}\,x'x \wedge \mathsf{C}\,y'y \wedge \mathsf{I}\,x'w \wedge \mathsf{I}\,y'w] \rightarrow P\,x'y']].$$

The latter formula is equivalent to

$$\forall x'\forall y'[[\exists w[\mathsf{W}w \wedge \mathsf{I}\,x'w \wedge \mathsf{I}\,y'w] \wedge \mathsf{C}\,x'x \wedge \mathsf{C}\,y'y] \rightarrow P\,x'y'].$$

Explicit quantification over worlds can be eliminated here by replacing $\exists w[\mathsf{W}w \wedge \mathsf{I}\,x'w \wedge \mathsf{I}\,y'w]$ in this formula by a direct relation of cohabitation between x' and y':

$$\forall x'\forall y'[[\mathsf{Coh}\,x'y' \wedge \mathsf{C}\,x'x \wedge \mathsf{C}\,y'y] \rightarrow P\,x'y'].$$

More generally, quantification over worlds can be imitated by quantifying over counterparts that are designated as representatives of the unique world that they inhabit.

A general language for *worldless counterpart theory* has three designated predicates: a two-place predicate R, a two-place predicate C, and a one-place predicate A. R denotes a relation between a counterpart and a fixed coinhabitant that serves to represent the counterpart's world. C denotes the counterpart relation. The predicate A denotes the property of inhabiting the actual world.

Rather than dealing with this general language, we will consider a specialized sublanguage for worldless counterpart theory. We will show that the sublanguage is equivalent to a sublanguage $\mathcal{L}^{C}_{Coh}$ of ordinary counterpart theory satisfying postulates (P1)–(P9). Since this language is adequate for formulating counterpart-based theories of necessity, this justifies using worldless counterpart theory in our investigation.

We don't attach any metaphysical significance to the fact that explicit quantification over worlds is not needed in the counterpart theory of modality. Anyone who takes counterpart theory or modal logic seriously is likely to think in terms of possible worlds, and—as in ordinary modal logic—explicit quantification over worlds can easily be added if it is desired. The easiest way to extend the simplified logic in this way would be to use a two-sorted first-order logic, with one sort for counterparts and another for worlds.

3.2. Technical presentation of worldless counterpart theory

The cohabitation-based worldless counterpart language $\mathcal{L}^{WLC}_{Coh}$ retains the counterpart predicates C and A, having these as well as a cohabitation predicate Coh as its only dedicated predicate constants.

Definition 1. Cohabitation-based worldless counterpart language $\mathcal{L}^{WLC}_{Coh}$.

Formulas of $\mathcal{L}^{WLC}_{Coh}$ are defined by the following induction.

(1.1) $P\,x_1 \ldots x_n$ is a formula of $\mathcal{L}^{WLC}_{Coh}$ if P is an n-place predicate of $\mathcal{L}^{C}$ other than W and I and $x_1, \ldots, x_n$ are variables.

(1.2) $\mathsf{Coh}\,xy$ is a formula of $\mathcal{L}^{WLC}_{Coh}$, where x and y are variables.

(2) If A and B are formulas of $\mathcal{L}^{WLC}_{Coh}$, so are $\neg A$ and $A \rightarrow B$.

(3) If A is a formula of $\mathcal{L}^{WLC}_{Coh}$ and x is a variable, $\forall x\, A$ is also a formula of $\mathcal{L}^{WLC}_{Coh}$.

Definition 2. Worldless counterpart frame.

A *counterpart frame* $\mathcal{F}$ for worldless counterpart theory is an ordered quadruple $\langle \mathrm{D}, \mathrm{r}, \mathrm{C}, @\rangle$ satisfying the following four conditions.

(1) D is a nonempty set.

(2) r is a function from D to D such that for all $\mathrm{d} \in \mathrm{D}$, $\mathrm{r}(\mathrm{r}(\mathrm{d})) = \mathrm{r}(\mathrm{d})$.

(3) C is a subset of D^2 such that for all $\mathrm{d}, \mathrm{e} \in \mathrm{D}$, if $\mathrm{r}(\mathrm{d}) = \mathrm{r}(\mathrm{e})$ then $\langle \mathrm{d}, \mathrm{e}\rangle \in \mathrm{C}$ if and only if $\mathrm{d} = \mathrm{e}$.

(4) $@ \in \mathrm{D}$ and $\mathrm{r}(@) = @$.

D is the domain of the frame; r is a function that for each worldbound individual picks the individual that represents the world they both inhabit.

A model M of $\mathcal{L}^{WLC}_{Coh}$ on a worldless counterpart frame $\langle \mathrm{D}, \mathrm{r}, \mathrm{C}, @\rangle$ assigns appropriate values M_P to constants P other than C, A and Coh. The values that M gives to C, A, and A are determined by D: $\mathrm{M}_{\mathsf{C}} = \mathrm{C}$, $\mathrm{M}_{\mathsf{A}} = \{\mathrm{d} \,/\, \mathrm{r}(\mathrm{d}) = @\}$, and $\mathrm{M}_{\mathsf{Coh}} = \{\langle \mathrm{d}, \mathrm{e}\rangle \,/\, \mathrm{r}(\mathrm{d}) = \mathrm{r}(\mathrm{e})\}$.

Remark 1. As an immediate consequence of Definition 2, if $\mathcal{F} = \langle \mathrm{D}, \mathrm{r}, \mathrm{C}, @\rangle$ is a worldless counterpart frame, then C is reflexive.

Definition 3. Habitation.

Let $\langle \mathrm{D}, \mathrm{r}, \mathrm{C}, @\rangle$ be a worldless counterpart frame. We will say that $\mathrm{d} \in \mathrm{D}$ inhabits $\mathrm{w} \in \mathrm{D}$ if $\mathrm{r}(\mathrm{d}) = \mathrm{w}$.

We now show that worldless counterpart theory is equivalent to Lewis' counterpart theory on a sublanguage adequate for translating modal logic with necessity. We define the appropriate sublanguage of Lewis' theory and show that its models and the models of worldless counterpart theory are interchangeable.

Definition 4. Cohabitation-based sublanguage $\mathcal{L}^C_{Coh}$ of $\mathcal{L}^C$.

A language $\mathcal{L}^C$ for Lewis-style counterpart theory will contain the special predicates C, I, A, and W. Assume that $\mathcal{L}^C$ has no individual constants—this loses no generality. The *cohabitation-based sublanguage* $\mathcal{L}^C_{Coh}$ of $\mathcal{L}^C$ does away with the predicates W and I of $\mathcal{L}^C$. but retains the counterpart relation C, the cohabitation relation $\exists x'[\mathsf{W}\,x \wedge \mathsf{I}\,xx' \wedge \mathsf{I}\,yx']$, and the actuality predicate A.

(1.1) $P\,x_1 \ldots x_n$ is a formula of $\mathcal{L}^C_{Coh}$ if P is an n-place predicate of $\mathcal{L}^C$ other than W and I, and $x_1, \ldots, x_n$ are variables.

(1.2) $\exists x'[\mathsf{W}x' \wedge \mathsf{I}\,xx' \wedge \mathsf{I}\,yx']$ is a formula of $\mathcal{L}^C_{Coh}$ if x and y are variables, where x' is the first variable differing from both x and y.

(2) If A and B are formulas of $\mathcal{L}^C_{Coh}$, so are $\neg A$ and $A \to B$.

(3) If A is a formula of $\mathcal{L}^C_{Coh}$ and x is a variable, $\forall x\, A$ is also a formula of $\mathcal{L}^C_{Coh}$.

There is a straightforward translation from $\mathcal{L}^C_{Coh}$ to $\mathcal{L}^{WLC}_{Coh}$.

Definition 5. $\tau(A)$.

The translation $\tau(A)$ of a formula A of $\mathcal{L}^C_{Coh}$ into $\mathcal{L}^{WLC}_{Coh}$ is defined as follows.

(1.1) $\tau(P\,x_1 \ldots x_n) = P\,x_1 \ldots x_n$ for all basic formulas of $\mathcal{L}^C_{Coh}$, where P is any predicate other than W and I.

(1.2) $\tau(\exists x'[\mathsf{W}x' \wedge \mathsf{I}\,xx' \wedge \mathsf{I}\,yx']) = \mathsf{Coh}(x, y)$.

(2) $\tau(\neg A) = \neg\tau(A)$, $\tau(A \to B) = \tau(A) \to \tau(B)$.

(3) $\tau(\forall x A) = \forall x\, \tau(A)$.

Theorem 1. Any model M of a language $\mathcal{L}^C$ for Lewis' counterpart theory that satisfies postulates (P1)–(P9) is equivalent over the sublanguage $\mathcal{L}^C_{Coh}$ to a corresponding model of worldless counterpart theory. That is, there is a model M′ of the worldless language $\mathcal{L}^{WLC}_{Coh}$ such that for all formulas A of $\mathcal{L}^C_{Coh}$, $\mathrm{M} \models_{\mathrm{f}} A$ iff $\mathrm{M}' \models_{\mathrm{f}} \tau(A)$, for all variable assignments f which assign only nonworld values.

Proof. Let M be a model of (P1)–(P9) on a domain D. (P9) ensures that $\mathrm{D} = \mathrm{D}_1 \cup \mathrm{D}_2$ and that D_1 and D_2 are disjoint, where D_1 is M_{W} (the extension of W in M) and D_2 is the set of elements satisfying $\exists y \mathsf{C} yx$. Postulates (P6)–(P8) ensure that D_2 is nonempty. Postulates (P1)–(P3) ensure that there is a function η from D_2 to D_1 such that $\eta(\mathrm{d}) = \mathrm{e}$ iff $\langle \mathrm{d}, \mathrm{e}\rangle \in \mathrm{M}_{\mathsf{I}}$. Therefore there is a function θ from D_2 to D_2 such that $\theta(\mathrm{d}) = \theta(\mathrm{e})$ iff $\eta(\mathrm{d}) = \eta(\mathrm{e})$ and $\theta(\theta(\mathrm{d})) = \theta(\mathrm{d})$. Postulates (P4) and (P6) ensure that M_{C} is a subset of D_2^2. Postulate (P5) ensures that if $\theta(\mathrm{d}) = \theta(\mathrm{e})$ and $\langle \mathrm{d}, \mathrm{e}\rangle \in \mathrm{M}_{\mathsf{C}}$ then $\mathrm{d} = \mathrm{e}$, and (P5) and (P6) ensure that $\langle \mathrm{d}, \mathrm{d}\rangle \in \mathrm{M}_{\mathsf{C}}$, for all $\mathrm{d} \in \mathrm{D}_2$. (P7) ensures that there is a $\mathrm{d}_0 \in \mathrm{D}_1$ such that for all $\mathrm{e} \in \mathrm{D}_2$, $\eta(\mathrm{e}) = \mathrm{d}_0$ iff $\mathrm{e} \in \mathrm{M}_{\mathsf{A}}$. By (P8), there is an $\mathrm{e}' \in \mathrm{D}_2$ such that $\eta(\mathrm{e}') = \mathrm{d}_0$. Let $\theta(\mathrm{e}')$ be e_0. Then for all $\mathrm{e} \in \mathrm{D}_2$, $\theta(\mathrm{e}) = \mathrm{e}_0$ iff $\mathrm{e} \in \mathrm{D}_{\mathsf{A}}$.

Let D be D_2, r be θ, C be C_{M}, and @ be e_0. The remarks above show that D is a worldless counterpart frame.

Define a worldless model M' on $\mathcal{F}$ so that M' assigns to each predicate of $\mathcal{L}^{WLC}_{Coh}$ other than W and I the restriction of that predicate to the individuated elements of the domain of M. That is, for all such predicates P, $P_{\mathrm{M}'}$ is the restriction of M_P to D_2. In particular, then, $M'_{\mathsf{C}} = M_{\mathsf{C}}$ and $M'_{\mathsf{A}} = M_{\mathsf{A}}$. Finally, $\langle \mathrm{d}, \mathrm{e}\rangle \in M'_{\mathsf{Coh}}$ iff $\theta(\mathrm{d}) = \theta(\mathrm{e})$.

It is straightforward to show by induction on the complexity of formulas that for all formulas A of $\mathcal{L}^{WLC}_{Coh}$, and for all assignments f of values in D_2 to the variables of $\mathcal{L}^{WLC}_{Coh}$, $\mathrm{M} \models_{\mathrm{f}} A$ iff $\mathrm{M}' \models_{\mathrm{f}} \tau(A)$.

In the other direction, we show that models for the cohabitation fragment of worldless counterpart theory can be converted to equivalent models of Lewis' counterpart theory.

Theorem 2. Any model of worldless counterpart theory is equivalent on the cohabitation sublanguage to a corresponding model of Lewis' counterpart theory satisfying (P9). That is, for each model M of $\mathcal{L}^{WLC}_{Coh}$ on a worldless frame $\langle \mathrm{D}, \mathrm{r}, \mathrm{C}, @\rangle$, there is a model M' of the corresponding language $\mathcal{L}^{C}$ for Lewis' counterpart theory such that M' satisfies (P9) and for all formulas A of $\mathcal{L}^{C}_{Coh}$, $\mathrm{M}' \models_{\mathrm{f}} A$ iff $\mathrm{M} \models_{\mathrm{f}} \tau(A)$, for all variable assignments f on M'.

Proof. Let M be a model of $\mathcal{L}^{WLC}_{Coh}$ on the frame $\mathcal{F} = \langle \mathrm{D}, \mathrm{r}, \mathrm{C}, @\rangle$. Using objects foreign to D, create an isomorphic copy W of r(D); let α be the isomorphism from r(D) to W. Let D′ be $\mathrm{D} \cup \mathrm{W}$, let M'_{W} be W, let M'_{I} be $\{\langle \mathrm{d}, \mathrm{w}\rangle \ / \ \mathrm{r(d)} = \mathrm{e}, \text{ and } \mathrm{w} = \alpha(\mathrm{e})\}$, and let M'_{C} be M_{C}. For predicates P of $\mathcal{L}^{WLC}_{Coh}$ other than Coh, let M'_P be M_P.

It is straightforward to show that M′ satisfies all of the postulates (P1)–(P9). For instance, suppose that $\langle \mathrm{d}, \mathrm{w}\rangle \in \mathrm{M}'_{\mathsf{I}}$, $\langle \mathrm{d}', \mathrm{w}\rangle \in \mathrm{M}'_{\mathsf{I}}$, and $\langle \mathrm{d}', \mathrm{d}\rangle \in \mathrm{M}'_{\mathsf{C}}$. Now, M'_{I} is defined so that $\mathrm{w} \in \mathrm{W}$. Since α is an isomorphism, there is a unique member of r(D), say e, such that $\alpha(\mathrm{e}) = \mathrm{w}$. Then $\mathrm{r(d)} = \mathrm{e}$ and $\mathrm{r(d')} = \mathrm{e}$. But $\langle \mathrm{d}', \mathrm{d}\rangle \in \mathrm{M}'_{\mathsf{C}}$, so$\langle \mathrm{d}', \mathrm{d}\rangle \in \mathrm{M}_{\mathsf{C}}$. So, by Condition (3) of Definition 2, $\mathrm{d} = \mathrm{d}'$. Therefore, M′ satisfies Postulate (5).

Furthermore, by an induction on complexity of formulas of $\mathcal{L}^{C}_{Coh}$ we can verify that $\mathrm{M}' \models_{\mathrm{f}} A$ iff $\mathrm{M} \models_{\mathrm{f}} \tau(A)$, for all formulas A of $\mathcal{L}^{C}_{Coh}$, where f is a variable assignment on M.

These results justify using the simpler worldless counterpart theory in the subsequent investigations. The results will transfer to a part of Lewis' counterpart theory that is adequate for characterizing modal logic. From here on, when we talk about "counterpart theory," we will mean worldless counterpart theory.

4. Counterpart theory and modal logic

Both counterpart theory and ordinary modal logic with the Kripke interpretation treat necessity as a sort of universal quantifier. We now show that this similarity runs fairly deep: the two approaches are equivalent for propositional modal logic.

4.1. Modal counterpart theory as a normal modal logic

The modal language $\mathcal{L}^{WLC}_{Coh\Box}$ is the result of adding a primitive necessity operator $\Box$ to $\mathcal{L}^{WLC}_{Coh}$, and restricting the atomic formulas to those having the form Pw, where w is a designated free variable.

Definition 6. Propositional modal language $\mathcal{L}^{WLC}_{Coh\Box}$.
Let w be a fixed designated individual variable of $\mathcal{L}^{WLC}_{Coh}$. The *propositional modal sublanguage* $\mathcal{L}^{WLC}_{Coh\Box}$ of $\mathcal{L}^{WLC}_{Coh}$ is the quantifier-free sublanguage defined as follows.

(1) Pw is a formula of $\mathcal{L}^{WLC}_{Coh\Box}$ if P is a 1-place predicate of $\mathcal{L}^{WLC}_{Coh\Box}$ other than C and Coh.

(2) If A and B are formulas of $\mathcal{L}^{WLC}_{Coh\Box}$, so are $\neg A$, $A \to B$, and $\Box A$.

The designated variable w serves in the proof of Theorem 3 as a world designator. We can assume, if we like, that $\mathrm{f}(w) = \mathrm{r}(\mathrm{f}(w))$.

The operator $\Box$ can be defined in terms of first-order quantification, C, and Coh, using Lewis' equivalence:

(L) $\Box A \;\leftrightarrow\; \forall y_1 \ldots \forall y_n[\mathsf{C}\, y_1 x_1 \wedge \ldots \mathsf{C}\, y_n x_n \wedge \mathsf{Coh}\, y_1 \ldots y_n] \;\rightarrow\; A^{y_1}\!/x_1 \ldots {}^{y_n}\!/x_n$,
where $x_1, \ldots, x_n$ are all the variables occurring free in A and $y_1, \ldots, y_n$ are n different variables not occurring in A.

Here, $A^{y}\!/x$ is the result of replacing every free occurrence of x in A with an occurrence of y. In case A contains no free variables, the equivalence (L) is simply $\Box A \leftrightarrow A$. The generalized cohabitation predicate Coh that is used in (L) is defined by the following induction.

Definition 7. $\mathsf{Coh}\, t_1 \ldots t_n$.
$\mathsf{Coh}\, st = \exists x[\mathsf{R}\, sx \wedge \mathsf{R}\, tx]$.
$\mathsf{Coh}\, t_1 \ldots t_n t = \mathsf{Coh}\, t_1 \ldots t_n \wedge \mathsf{Coh}\, t_n t$.

Note that $\mathrm{M} \models_{\mathrm{f}} \mathsf{Coh}\, x_1 \ldots x_n$ iff there is a $\mathrm{d} \in \mathrm{D}$ such that $\mathrm{r}(f(x_i)) = \mathrm{d}$ for all i such that $1 \leqslant i \leqslant n$.

Satisfaction in a model, relative to a variable assignment f, is defined as usual in first-order logic for formulas other than $\Box A$. The satisfaction clause for $\Box A$ that is determined by (L) reads as follows:

$\mathrm{M} \models_{\mathrm{f}} \Box A$ iff
$\mathrm{M} \models_{\mathrm{f}} \forall y_1 \ldots \forall y_n[\mathsf{C}\, y_1 x_1 \wedge \ldots \mathsf{C}\, y_n x_n \wedge \mathsf{Coh}\, y_1 \ldots y_n] \rightarrow A^{y_1}\!/x_1 \ldots {}^{y_n}\!/x_n$,
where $x_1, \ldots, x_n$ are all the variables occurring free in A and $y_1, \ldots, y_n$ are n different variables not occurring in A.

(In case A contains just one free variable x, $\mathrm{M} \models_{\mathrm{f}} \Box A$ iff $\mathrm{M} \models_{\mathrm{f}} \forall y[\mathsf{C}\, yx \rightarrow A^{y}\!/x]$, where y is a variable not occurring in A. In case A contains no free variables, $\mathrm{M} \models_{\mathrm{f}} \Box A$ iff $\mathrm{M} \models_{\mathrm{f}} A$.)

We begin by showing that counterpart theory gives rise to a propositional modal logic with a quite standard possible-worlds interpretation. We can relate formulas of counterpart theory with one designated free variable in a natural way to propositional modal logic, and the logic induced by the Lewis scheme corresponds to a standard modal logic in which the necessity operator can be interpreted in the usual way, using relations over a set of possible worlds.

Definition 8. Counterpart model M of $\mathcal{L}^{WLC}_{Coh\Box}$.
Let $\mathcal{F} = \langle \mathrm{D}, \mathrm{r}, \mathrm{C}, @ \rangle$ be a frame for worldless counterpart theory. A counterpart model M of $\mathcal{L}^{WLC}_{Coh\Box}$ on $\mathcal{F}$ is an ordinary first-order model of the modal language $\mathcal{L}^{WLC}_{Coh\Box}$ on the domain D, satisfying the Lewis scheme (L).

Our idea is to think of variable assignments as worlds. To keep the size of the resulting set of worlds under control, we restrict ourselves to finitary assignments.

Definition 9. Finitary variable assignment.

A finitary variable assignment (or sequence) on a frame $\mathcal{F}$ is an eventually constant function f from the individual variables of $\mathcal{L}^{WLC}_{Coh\Box}$ to D. That is, we assume a fixed ordering of the variables and that the set of these variables is denumerable, so that the set of variables has the form $\{x_1, x_2, \ldots\}$. Then for each finitary variable assignment f, there is an n such that for all $m, m' > n$, $\mathrm{f}(x_m) = \mathrm{f}(x_{m'})$.

Definition 10. Local variable assignment.

A variable assignment f for the language $\mathcal{L}^{WLC}_{Coh\Box}$ on a frame $\mathcal{F}\langle \mathrm{D}, \mathrm{r}, \mathrm{C}, @\rangle$ is *local* (on world w) iff for all variables x and y, $\mathrm{r}(\mathrm{f}(x)) = \mathrm{r}(\mathrm{f}(y)) = \mathrm{w}$.

In other words, a variable assignment is local if all its values inhabit the same world.

Definition 11. World of a local variable assignment.

The world of a local variable assignment is the world that all of its values inhabit.

Definition 12. $\sigma(A)$.

Let $\mathcal{L}_\Box$ be the standard language of propositional modal logic. The atomic formulas of $\mathcal{L}_\Box$ are variable-free, and $\Box A$ is a formula of $\mathcal{L}_\Box$ if A is. We now define a translation σ from $\mathcal{L}^{WLC}_{Coh\Box}$ to $\mathcal{L}_\Box$ by removing the variable w. That is, the translation $\sigma(A)$ of a formula A of $\mathcal{L}^{WLC}_{Coh}$ into $\mathcal{L}_\Box$ is defined as follows:

(1) $\sigma(Pw) = P$.

(2) $\sigma(\neg A) = \neg\sigma(A)$, $\sigma(A \to B) = \sigma(A) \to \sigma(B)$.

(3) $\sigma(\Box A) = \Box\sigma(A)$.

Definition 13. Modal frame and model corresponding to a counterpart model.

Let M be a counterpart model of $\mathcal{L}^{WLC}_{Coh\Box}$ on the frame $\mathcal{F} = \langle \mathrm{D}_\mathcal{F}, \mathrm{r}_\mathcal{F}, \mathrm{C}, @\rangle$. The corresponding modal (Kripke) frame $\mathcal{F}'$ is the pair $\langle \mathrm{W}, \mathrm{R}\rangle$, where W is the set of finitary, local variable assignments on D and (where $\mathrm{f}, \mathrm{g} \in \mathrm{W}$) $\langle \mathrm{f}, \mathrm{g}\rangle \in \mathrm{R}$ iff for all variables x, $\langle \mathrm{g}(x), \mathrm{f}(x)\rangle \in \mathrm{C}$. And the corresponding modal model M$'$ is defined by letting $\mathrm{f} \in \mathrm{M}'_P$ iff $\mathrm{f}(w) \in \mathrm{M}_P$.

In other words, f bears the relation R to g iff g assigns every variable x a counterpart of the value that f gives to x. (Since g is local, all these counterparts must inhabit the same world.)

Definition 14. Modal satisfaction.

Let M$'$ be a modal model of $\mathcal{L}_\Box$, and f, g be local, finitary variable assignments. The modal satisfaction relation $\mathrm{M} \models'_\mathrm{f} A$ is defined as usual for formulas A of $\mathcal{L}_\Box$. In particular,

$\mathrm{M}' \models'_\mathrm{f} P$ iff $\mathrm{f} \in \mathrm{M}'_P$. And the clause for $\Box A$ reads as follows:

$\mathrm{M}' \models'_\mathrm{f} \Box A$ iff $\mathrm{M}' \models'_\mathrm{g} A$ for all g such that $\langle \mathrm{f}, \mathrm{g}\rangle \in \mathrm{R}$.

Theorem 3. Satisfaction according to Lewis' rule in $\mathcal{L}^{WLC}_{Coh\Box}$ and modal satisfaction according to Definition 14 coincide. That is, for all counterpart models M of $\mathcal{L}^{WLC}_{Coh\Box}$, $\mathrm{M} \models_{\mathrm{f}} A$ iff $\mathrm{M}' \models'_{\mathrm{f}} \sigma(A)$, where M′ is the modal model corresponding to M according to Definition 13.

Proof. We induce on the complexity of formulas of $\mathcal{L}^{WLC}_{Coh\Box}$. The only nontrivial case is the one for formulas of the form $\Box A$. Suppose first that $\mathrm{M} \models_{\mathrm{f}} \Box A$, where M is a model on a worldless counterpart frame $\mathcal{F}$. Using the satisfaction condition for (L), this holds iff for all $\mathrm{d} \in \mathrm{D}$, if $\langle \mathrm{d}, \mathrm{f}(w)\rangle \in \mathrm{C}$ then $\mathrm{M} \models_{\mathrm{f}[\mathrm{d}/w]} A$. Since A contains no free variables other than w, this iff for all g such that $\langle \mathrm{f}, \mathrm{g}\rangle \in \mathrm{R}$, $\mathrm{M} \models_{\mathrm{g}} A$. By the hypothesis of induction, this iff $\mathrm{M}' \models'_{\mathrm{g}} \sigma(A)$ for all such g. This iff $\mathrm{M}' \models'_{\mathrm{f}} \Box\sigma(A)$, i.e. iff $\mathrm{M}' \models'_{\mathrm{f}} \sigma(\Box A)$.
In the other direction, note that the above argument reverses.

Theorem 3 guarantees that Condition (L) produces a normal modal propositional logic, with local variable assignments playing the role of possible worlds. Since the relation R given by Definition 13 is reflexive, the construction of Theorem 3 yields a frame validating the modal **T** axiom, $\Box A \to A$. This theorem is restricted to the propositional case, but later, in Theorem 10, we will provide a generalization involving the same idea.

4.2. Quantifiers in first-order modal logic and in counterpart theory

Standard approaches to the modal logic of quantification begin with domains of objects that are individuated across worlds. Models literally employ the idea of "the same" individual in different worlds. It is natural on these approaches to think of modalities as properties of propositions. 'Five is necessarily prime', for instance, is true because the proposition that five is prime is necessary. This necessity involves, of course, the number five and the property of being prime, but involves them indirectly; the necessity applies directly to the proposition that results from combining the individual and the property.

Counterpart theory is less uncritical about trans-world sameness. It uses world-bound individuals and the counterpart relation to account for necessity. There is no appeal to sameness of an individual across worlds in interpreting formulas such as $\exists x \Box P x$. Necessity is explained by condition (L), so that the formula is true if and only if there is an individual whose counterparts all satisfy P. No natural notion of a proposition is directly involved in this account of modality—modal formulas are complex conditions involving properties of world-bound individuals and the counterpart relation, and these conditions do not factor naturally into a propositional element and a modal operator.

We have expressed these differences informally, and the comparisons involve some notions (especially, that of individuation) which are problematic. But they are reflected in formal characteristics of the models of these two approaches.

In counterpart theory, a one-place predicate P is assigned a set of individuals; atomic formulas such as Px are interpreted locally—in a specific world—because a world is implicitly determined by a value for x. Therefore the spectrum of values of Px, as x is allowed to vary, is global—it reflects the behavior of P in all worlds. In first-order modal logics, the interpretation of every formula, including Px, is relativized to a world: as x is allowed to vary, the values of Px represent the truth-values of Px in a single world. But the values of formulas and many of their components can differ across worlds, and by varying the world, one recovers a spectrum of values. Individual variables occupy a special—and in some versions of modal quantification theory an exceptional—place in the theory. The value of x—an individuated object—is global because x is assigned an object that occupies many (in fact, all) worlds. Although you can talk about the value of x in a world, this way of putting it is somewhat misleading, since this value inhabits all worlds, and is the same in all of them.

In counterpart theory, the formula $\forall x \Box Px$ says that every counterpart of anything satisfies P; in worldless counterpart theory, this amounts to saying that everything satisfies P. In modal quantification theory, $\forall x \Box Px$ says that for every individual, that same individual satisfies P in every world.

Perhaps the most striking difference between the two theories of modality is that in counterpart theory $\forall x\, Px \rightarrow \Box \forall x\, Px$ is valid, whereas in modal quantification theory it is invalid. This difference creates a strong suspicion that there is no simple, natural mapping between the two theories that preserves validity, and it reinforces the idea, already present in the motivation of the theories, that the ontologies of the two approaches are fundamentally different.

One approach to relating the two approaches was suggested by Lewis:[5] treat world-bound individuals as pairs consisting of a world and an individuated individual. This idea helps to clarify the differences between the two approaches to quantification and modality. Models of modal quantification theory involve frames of the sort $\langle \mathrm{W}, \mathrm{R}, \mathrm{D}, \mathrm{w}_0 \rangle$, where W is the set of worlds, R is a binary relation over W, D is the set of individuals, and w_0 is a member of W representing the actual world. Such a frame corresponds in a natural way to a (worldless) counterpart frame $\mathcal{F} = \langle \mathrm{D}, \mathrm{r}, \mathrm{C}, \mathrm{w}_0 \rangle$, where $\mathrm{D} = \mathrm{W} \times \mathrm{D}$, $\mathrm{r}(\langle \mathrm{w}, \mathrm{d} \rangle) = \langle \mathrm{w}, \mathrm{d}_0 \rangle$ (where d_0 is an arbitrary member of D), and $\langle \mathrm{w}, \mathrm{d} \rangle \, \mathrm{C} \, \langle \mathrm{w}', \mathrm{d}' \rangle$ iff $\mathrm{d} = \mathrm{d}'$. In other words, the counterpart relation applies to world-object pairs that involve the same object.

Counterpart relations defined in this way over world-individual pairs have certain special properties: for instance, a pair must have one and only one counterpart per world. This provides support for Lewis' claim that counterpart theory is more flexible and general than quantified modal logic.[6]

However, this way of relating models of modal quantification theory to models of counterpart theory does not induce a natural correspondence between formulas. Modality in normal modal logic involves an (implicit) quantifier over worlds. Modality in counterpart theory involves a quantifier over world-individual pairs. This is reflected in the fact that formulas like $\Diamond \forall x\, P\, x \rightarrow \Box \forall x\, P\, x$ are valid in counterpart theory. This formula, of course, is invalid in modal quantification theory.

[5] [Lewis, 1968][p. 115].

[6] This point only applies to the simplest versions of quantified modal logic. It would not apply, for instance, to a version that quantifies over a (restricted) set of partial functions from worlds to individuals.

Such facts mean that it will not be straightforward to establish object-language level correspondences between the two approaches.

5. Introducing actuality into first-order logic

5.1. Actuality in ordinary modal logic

It can be instructive to extend a propositional modal logic $\mathcal{L}_\Box$ by adding an "actuality operator" $[@]$, obtaining a language $\mathcal{L}_{\Box @}$. The idea is that interpretations are anchored in an actual world, and that $[@]A$ means that A holds in that world. Thus, for instance, $A \leftrightarrow [@]A$ is valid, but $\Box[A \leftrightarrow [@]A]$ is not. This extension of modal logic can be interpreted using "two-dimensional" interpretations; see [Hodes, 1984]. These two-dimensional models employ *double indexing* techniques: formulas are satisfied with respect not to a single world but with respect to an *index*, a pair $\langle \mathrm{w}, \mathrm{w}' \rangle$ of worlds. The crucial clauses of the satisfaction definition are:

(1) $\mathrm{M} \models_{\mathrm{w},\mathrm{w}'} \Box A$ iff for all u such that $\mathrm{w}\, R\, \mathrm{u}$, $\mathrm{M} \models_{\mathrm{u},\mathrm{w}'} A$.
(2) $\mathrm{M} \models_{\mathrm{w},\mathrm{w}'} [@]A$ iff $\mathrm{M} \models_{\mathrm{w}',\mathrm{w}'} A$.

Two-dimensional validity is defined as follows.

Definition 15. Validity for formulas of $\mathcal{L}_{\Box @}$.
A formula of A of $\mathcal{L}_{\Box @}$ is valid if and only if for every model M, $\mathrm{M} \models_{\mathrm{w},\mathrm{w}} A$ for all worlds w of M.

By identifying the two arguments of the satisfaction relation, Definition 15 confines attention in determining validity to *diagonal indices*. It is this restriction that makes $A \leftrightarrow [@]A$ valid, even though $\Box[A \leftrightarrow [@]A]$ is invalid.

The satisfaction definition for one-dimensional modal logic is a recursion in which a parameter representing a possible world is maintained in the evaluation of a formula. In determining whether $\mathrm{M} \models_{\mathrm{w}} A$, an original value is given to the parameter 'w' at the outset of the recursion. This value remains constant when boolean formulas are tested: for instance, whether $\mathrm{M} \models_{\mathrm{w}} \neg A$ depends on whether $\mathrm{M} \models_{\mathrm{w}} A$. It is varied for modal formulas: whether $\mathrm{M} \models_{\mathrm{w}} \Box A$ depends on whether, for each world u such that $\mathrm{w}\, R\, \mathrm{u}$, $\mathrm{M} \models_{\mathrm{u}} A$.

In two-dimensional modal logic, the first parameter plays the role of the one-dimensional parameter: it keeps track of the world in which a formula is evaluated as the recursion unfolds. The second parameter serves to remember a value that was visited earlier in the recursion—in fact, the initial value. In testing for validity, the first value for each recursion will be the world designated as actual for purposes of the evaluation. Since the value of the second parameter never changes in the course of the recursion, it will serve to store this world.

We are now in a position to begin the project, outlined above in Section 2, of altering first-order logic in a way that parallels the change in propositional modal logic from one to two satisfaction parameters.

5.2. Two-dimensional first-order logic

The question is this: what generalization of the satisfaction relation would be needed to support an actuality operator, assuming that modality is characterized in terms of Condition (L)?

5.2.1. A special case

Consider the following special case of Lewis' condition (L):

$$\Box P\,x \leftrightarrow \forall y[\mathsf{C}\,yx \rightarrow P\,y].$$

This produces the following satisfaction condition for $\Box P\,x$:

$$\mathrm{M} \models_{\mathrm{f}} \Box P\,x \text{ iff } \mathrm{M} \models_{\mathrm{f}} \forall y[\mathsf{C}\,yx \rightarrow P\,y].$$

This condition refers the question of whether $\Box P$ holds of $\mathrm{f}(x)$ to the question of whether P holds of all counterparts of $\mathrm{f}(x)$.

Let $\mathcal{L}_x^{\mathrm{FOL}}$ be a sublanguage of first-order logic in which the only variable to occur free in formulas is x. The atomic formulas of the language have the form $P\,x$, where P is a one-place predicate letter. Formulas are closed under boolean operations, and if A is a formula then so is $\forall y[\mathsf{C}\,yx, A]$. This formula is equivalent to $\forall y[\mathsf{C}\,yx \rightarrow A^{y}\!/x]$ but is treated as an independent construction, not as a universally quantified conditional.

For this fragment of first-order logic, we can relativize satisfaction to two individuals, in exactly the same way that two-dimensional modal logic relativizes satisfaction to two worlds, obtaining a semantics for modality and actuality. The satisfaction clauses are as follows.

(1) $\mathrm{M} \models_{\mathrm{d,e}} P\,x$ iff $\mathrm{d} \in \mathrm{M}_P$.
(2) Boolean connectives are interpreted as usual.
(3) $\mathrm{M} \models_{\mathrm{d,e}} \forall y[\mathsf{C}\,yx, A]$ iff $\mathrm{M} \models_{\mathrm{d',e}} A$ for all $\mathrm{d'} \in \mathrm{D}$ such that $\langle \mathrm{d'}, \mathrm{d}\rangle \in \mathrm{M}_{\mathsf{C}}$.
(4) $\mathrm{M} \models_{\mathrm{d,e}} [@]A$ iff $\mathrm{M} \models_{\mathrm{e,e}} A$.

A formula A of $\mathcal{L}_x^{\mathrm{FOL}}$ is valid if and only if $\mathrm{M} \models_{\mathrm{d,d}} A$ for all models M of $\mathcal{L}_x^{\mathrm{FOL}}$, for all $\mathrm{d} \in \mathrm{D}$, where D is the domain of the model M.

This semantics parallels Hodes' interpretation of standard modal logic with actuality, and is equivalent to it, assuming that the modal language is quantifier-free. That is, we can map formulas of modal actuality logic into $\mathcal{L}_x^{\mathrm{FOL}}$ in such a way that formulas containing a single free variable, x, are valid in the Hodes logic if and only if their translations are valid in $\mathcal{L}_x^{\mathrm{FOL}}$.

5.2.2. The general case

Full first-order logic, however, has infinitely many individual variables. If we wish to base counterpart theory with actuality on this logic, as Lewis does, we have to consider how to introduce a mechanism for remembering variable assignments that is sufficiently general to recall the appropriate world on demand when the supply of individual variables is infinite.

No doubt there are many ways to do this. For our purposes, we want a simple mechanism as close as possible to Hodes' two-dimensional method for interpreting modal logic with actuality. This suggests using "two-dimensional variable assignments," which associate pairs of individuals with variables. Equivalently, we use two one-dimensional variable assignments to interpret formulas. The first assignment serves to interpret first-order quantifiers; the second assignment remembers previous values.

Quantifiers serve two purposes in counterpart theory: (i) characterizing necessity and (ii) supporting generalizations about individuals, as in any first-order theory. In two-dimensional modal logic, the first purpose requires a record of previous values, but the second does not. To capture this difference in two-dimensional first-order logic, we introduce two sorts of variables. We start, as usual, with an infinite set of individual variables, which we call "plain variables." With each plain variable, we then associate infinitely many "companion variables." That is, where x is a plain variable the companions $x^c_1, x^c_2, \ldots$ are associated with x. Companion variables behave in most respects like ordinary variables, but quantification with a companion variable remembers the previous value of the variable, while quantification with a plain variable does not. (See Clause (3.2) of Definition 18, below.)

Definition 16. The language $\mathcal{L}^{\mathrm{FOL}}_{\mathrm{2D}}$.
$\mathcal{L}^{\mathrm{FOL}}_{\mathrm{2D}}$ is an ordinary first-order language, extended with infinitely many "companion variables" for each "plain variable" of the language. Where x is a plain variable, 'x^c' denotes a companion of x.

When we say that quantifying with a plain variable x does not remember previous values of x, this means that the clause for $\mathrm{M} \models_{\mathrm{f,g}} \forall x\, Px$ will change not only the value of f, but the value of g. For technical reasons that will only become more clear in the proofs of Theorems 8 and 10, we make the change to g as general as possible: $\mathrm{M} \models_{\mathrm{f,g}} \forall x\, Px$ iff for all $\mathrm{d} \in \mathrm{D}$, $\mathrm{M} \models_{\mathrm{f}[\mathrm{d}/x],\mathrm{g}[\mathcal{C}(\mathrm{d})/x]} A$, where $\mathcal{C}$ is a function from D to D. In other words, g assigns x an arbitrary function of d. The following definition imposes further conditions on the function $\mathcal{C}$ in counterpart frames.

Definition 17. Two-dimensional first-order frame, two-dimensional counterpart frame.
A two-dimensional first-order frame is a pair $\langle \mathrm{D}, \mathcal{C}\rangle$, where D is a nonempty set (the domain of the frame) and $\mathcal{C}$ is a function from D to D. A two-dimensional counterpart frame is a 6-tuple $\langle \mathrm{D}, \mathcal{C}, \mathrm{r}, \mathrm{C}, @, \mathrm{d}'\rangle$, where $\langle \mathrm{D}, \mathrm{r}, \mathrm{C}, @\rangle$ is a worldless counterpart frame, $\mathcal{C}$ is a function from D to D, and $\mathrm{r}(\mathrm{d}') = \mathrm{d}'$ (i.e., d' is a world of the frame). The following conditions apply:

(1) for all $\mathrm{d} \in \mathrm{D}$, $\mathrm{r}(\mathcal{C}(\mathrm{d})) = \mathrm{d}'$,

(2) if there is an e such that $\langle \mathrm{e}, \mathrm{d}\rangle \in \mathrm{C}$ and $\mathrm{r}(\mathrm{e}) = \mathrm{d}'$, then $\langle \mathcal{C}(\mathrm{d}), \mathrm{d}\rangle \in \mathrm{C}$. (In other words, all values of $\mathcal{C}$ must inhabit the world d', and if there is a counterpart of d inhabiting d' then $\mathcal{C}(\mathrm{d})$ is such a counterpart.)

Remark 2. We distinguish d', the world that anchors the actuality operator, from Lewis' actual world @. But it is tempting and very natural to identify the two.

Definition 18. Satisfaction in two-dimensional first-order logic.

Let M be a first-order model of $\mathcal{L}_{2D}^{\mathrm{FOL}}$ on the domain D.

(1) $\mathrm{M} \models_{\mathrm{f,g}} P\,x_1 \ldots x_n$ iff $\langle \mathrm{f}(x_1), \ldots, \mathrm{f}(x_n) \rangle \in \mathrm{M}_P$.

(2) Boolean connectives are interpreted as usual.

(3.1) Where x is a (plain) individual variable, $\mathrm{M} \models_{\mathrm{f,g}} \forall x\, A$ iff for all $\mathrm{d} \in \mathrm{D}$, $\mathrm{M} \models_{\mathrm{f[d}/x\mathrm{],g[}\mathfrak{c}\mathrm{(d)}/x]} A$.

(3.2) Where x^c is a (companion) variable, $\mathrm{M} \models_{\mathrm{f,g}} \forall x^c\, A$ iff for all $\mathrm{d} \in \mathrm{D}$, $\mathrm{M} \models_{\mathrm{f[d}/x^c\mathrm{],g}} A$.

As usual in two-dimensional settings, for purposes of checking validity, satisfaction in $\mathcal{L}_{2D}^{\mathrm{FOL}}$ begins with special pairs of assignments, ones that are designated as initial. An initial assignment pair is not only diagonal, but is local, and the values of g are correlated to the values of f by the function $\mathfrak{c}$.

Definition 19. Local, uniform, initial variable assignment pair.

A variable assignment pair $\langle \mathrm{f}, \mathrm{g} \rangle$ for $\mathcal{L}_{2D}^{\mathrm{FOL}}$ is *local* on w and w′ iff f is local on w and g is local on w′. It is *uniform* if for all plain variables x and companions x^c for x, $\mathrm{f}(x^c) = \mathrm{f}(x)$. $\langle \mathrm{f}, \mathrm{g} \rangle$ is *initial* iff (i) $\mathrm{f} = \mathrm{g}$, (ii) $\langle \mathrm{f}, \mathrm{g} \rangle$ is uniform, and (iii) $\langle \mathrm{f}, \mathrm{g} \rangle$ is local on some w and w′.

Definition 20. Validity for formulas of $\mathcal{L}_{2D}^{\mathrm{FOL}}$.

A formula A of $\mathcal{L}_{2D}^{\mathrm{FOL}}$ is valid iff for all models M and initial variable assignment pairs $\langle \mathrm{f}, \mathrm{g} \rangle$, $\mathrm{M} \models_{\mathrm{f,g}} A$.

If we ignore companion variables, these changes make no difference: validity in $\mathcal{L}_{2D}^{\mathrm{FOL}}$ is the same as first-order validity. We record this fact as a theorem.

Theorem 4. Let A be a formula of $\mathcal{L}_{2D}^{\mathrm{FOL}}$. Then A is valid in $\mathcal{L}_{2D}^{\mathrm{FOL}}$ iff A is satisfied in first-order logic by every normal variable assignment.

Proof. Where $\mathrm{M} \models_{\mathrm{f}}$ is the usual relation of first-order satisfaction in M relative to a variable assignment, it is easily shown by induction on the complexity of A that, for all assignment pairs $\langle \mathrm{f}, \mathrm{g} \rangle$, $\mathrm{M} \models_{\mathrm{f,g}} A$ iff $\mathrm{M} \models_{\mathrm{f}} A$ in ordinary first-order logic. The theorem follows immediately. It also follows that if A contains no companion variables, A is valid in $\mathcal{L}_{2D}^{\mathrm{FOL}}$ iff A is valid in ordinary first-order logic.

Theorem 4 shows that $\mathcal{L}_{2D}^{\mathrm{FOL}}$ is not a particularly interesting generalization if we confine ourselves to the language of first-order logic. This is analogous to the fact that validity in two-dimensional modal logic is uninteresting without an actuality operator or some other operator that depends on the second satisfaction parameter.

We now extend $\mathcal{L}_{2D}^{\mathrm{FOL}}$ to include an actuality operator, obtaining the language $\mathcal{L}_{2D@}^{\mathrm{FOL}}$. Actuality takes the form of a modal operator in $\mathcal{L}_{2D@}^{\mathrm{FOL}}$, as it does in modal logic. Therefore, in addition to the usual formation rules of first-order logic, $\mathcal{L}_{2D@}^{\mathrm{FOL}}$ has the following rule.

If A is a formula of $\mathcal{L}_{2D@}^{\mathrm{FOL}}$, so is $[@]\,A$.

Our satisfaction rule for actuality is analogous to the modal rule: it forgets the first parameter, replacing it with values from the second parameter.

$$\mathrm{M} \models_{\mathrm{f,g}} [@]A \text{ iff } \mathrm{M} \models_{\mathrm{g,g}} A.$$

The interaction of the first-order actuality operator with two-dimensionality is illustrated by the formula $Px \to \forall x^c [@] Px^c$. This formula holds at the initial assignment pair $\langle \mathrm{f}, \mathrm{f} \rangle$ in a model M, i.e., $\models_{\mathrm{f,f}} Px \to \forall x^c [@] Px^c$, iff

(i) if $\mathrm{f}(x) \in \mathrm{M}_P$ then for all $\mathrm{d} \in \mathrm{D}$, $\mathrm{M} \models_{\mathrm{f[d}/x^c\mathrm{],f}} [@] Px^c$.

And (using the satisfaction clause for $[@]$) (i) iff

(ii) if $\mathrm{f}(x) \in \mathrm{M}_P$ then for all $\mathrm{d} \in \mathrm{D}$, $\mathrm{M} \models_{\mathrm{f,f}} Px^c$.

Since the reference to d has been eliminated from the satisfaction clause, (ii) boils down to

(iii) if $\mathrm{f}(x) \in \mathrm{M}_P$ then $\mathrm{M} \models_{\mathrm{f,f}} Px^c$.

Now, $\mathrm{M} \models_{\mathrm{f,f}} Px^c$ iff $\mathrm{f}(x^c) \in \mathrm{M}_P$. But then (iii) follows from the fact that $\langle \mathrm{f}, \mathrm{f} \rangle$ is initial.

The following remark generalizes this example slightly.

Remark 3. Let A have no free occurrences of x^c, and let A' be A^{x^c}/x. Then $A \leftrightarrow \forall x^c [@] A'$ is valid in $\mathcal{L}^{\mathrm{FOL}}_{2\mathrm{D}@}$.

Proof. Let $\langle \mathrm{f}, \mathrm{f} \rangle$ be initial. $\mathrm{M} \models_{\mathrm{f,g}} \forall x^c [@] A'$ iff $\mathrm{M} \models_{\mathrm{f[d}/x^c\mathrm{],f}} [@] A'$ for all $\mathrm{d} \in \mathrm{D}$. But (using the satisfaction clause for $[@]$) this iff $\mathrm{M} \models_{\mathrm{f,f}} A'$. Using properties of substitution in first-order-logic, the fact that A has no free occurrences of x^c, and the fact that $\langle \mathrm{f}, \mathrm{f} \rangle$ is initial, this iff $\mathrm{M} \models_{\mathrm{f,g}} A$.

The following remark follows directly from this, using simple properties of first-order logic.

Remark 4. Let A have no free variables other than the distinct variables $x_1, \ldots, x_n$, let the companion variables $x_1^c, \ldots, x_n^c$ not occur in A, and let A' be $A^{x_1^c}/x_1 \ldots {}^{x_n^c}/x_n$. Then

$$A \to \forall x_1^c \ldots \forall x_n^c [[\mathsf{C}\, x_1^c x_1 \wedge \ldots \wedge \mathsf{C}\, x_n^c x_n \wedge \mathsf{Coh}\, x_1 \ldots x_n] \to [@] A']$$

is valid in $\mathcal{L}^{\mathrm{FOL}}_{2\mathrm{D}@}$.

The validity cited in Remark 4 is analogous to the validity of $A \to \Box [@] A$ in modal logic with actuality, and is an important component of the case for the adequacy of this treatment of actuality in counterpart theory.

The above remarks show that two-dimensional first-order logic with actuality produces some distinctive validities that seem to go beyond first-order logic. But it seems implausible to attach any metaphysical significance to the mechanisms that create these differences. The domain of the quantifiers is unchanged, and the only novelty is the ability to recover previous values of variables. If anything, the differences between $\mathcal{L}^{\mathrm{FOL}}_{2\mathrm{D}@}$ and ordinary first-order logic seem more epistemological than metaphysical, being a matter of memory and so of epistemology rather than, say, of ontology.

Counterpart theory based on two-dimensional first-order logic with actuality is our candidate for a plausible extension of counterpart theory that accommodates an actuality operator.

6. Justifying two-dimensional counterpart theory as a logic of actuality

We have proposed a satisfaction definition for counterpart theory with actuality that was designed to parallel the two-dimensional definition for the corresponding possible-worlds based account. It would be desirable, so far as this is possible, to support the adequacy of this theory with general results. But, as we explained in Section 1, we can't do this by showing somehow that the theory has no implausibilities, because some features of counterpart theory *without* actuality are controversial and—at least to some—implausible. The best we can do is to show that the addition of an actuality operator adds no new implausibilities to ones that counterpart theory may already have. To do that, we propose to eliminate the differences between plain counterpart theory and first-order modal logic, and to ask whether adding an actuality operator to this specialized version of counterpart theory introduces divergences from modal logic with actuality. We begin this exercise with the quantifier-free case.

6.1. The quantifier-free case

We use the translation σ that was used in Section 4.1 to relate counterpart theory to propositional modal logic, extending the results of that section to languages with actuality.

Definition 21. The language $\mathcal{L}^{WLC}_{Coh\Box@}$.
$\mathcal{L}^{WLC}_{Coh\Box@}$ is the result of basing $\mathcal{L}^{WLC}_{Coh\Box}$ (see Definition 6) on the extension $\mathcal{L}^{\mathrm{FOL}}_{2\mathrm{D}@}$ of first-order-logic with companion variables and an actuality operator.

To relate $\mathcal{L}^{WLC}_{Coh\Box@}$ to the modal language with actuality $\mathcal{L}_{\Box@}$ that was discussed in Section 5.1, we extend the translation function σ (taking formulas of $\mathcal{L}^{WLC}_{Coh\Box}$ to formulas of $\mathcal{L}_{\Box}$) that was defined in Definition 12, adding a clause to cover actuality.

Definition 22. $\sigma(A)$ for $\mathcal{L}_{\Box@}$.
(4) $\sigma([@]A) = [@]\sigma(A)$

We now proceed to prove that the validities of the modal language and their counterpart theoretical equivalents are the same.

The two-dimensional modal frame and model corresponding to a two-dimensional counterpart model M of $\mathcal{L}^{WLC}_{Coh\Box@}$ is defined as in Definition 13.

Theorem 5. Let M′ be the two-dimensional modal model corresponding to the two-dimensional counterpart model M of $\mathcal{L}^{WLC}_{Coh\Box@}$. Then for all formulas A of $\mathcal{L}^{WLC}_{Coh\Box@}$ and local variable assignments f and g on $\mathcal{L}^{WLC}_{Coh\Box@}$, $\mathrm{M} \models_{\mathrm{f,g}} A$ iff $\mathrm{M}' \models'_{\mathrm{f,g}} \sigma(A)$.

The proof of this theorem is an induction that doesn't differ in important respects from the proof of Theorem 3. The additional case, for formulas $[@]A$, is straightforward: $\mathrm{M} \models_{\mathrm{f,g}} [@]A$ iff $\mathrm{M} \models_{\mathrm{g,g}} A$. By the hypothesis of induction, this iff $\mathrm{M}' \models'_{\mathrm{g,g}} \sigma(A)$, and this iff $\mathrm{M}' \models'_{\mathrm{f,g}} [@]\sigma(A)$, i.e. iff $\mathrm{M}' \models'_{\mathrm{f,g}} \sigma([@]A)$.

In the other direction, we start with a two-dimensional modal model and construct an equivalent counterpart model.

Definition 23. Two-dimensional counterpart frame corresponding to a two-dimensional modal model.

A two-dimensional counterpart frame $\mathcal{F} = \langle \mathrm{D}, \mathcal{C}, \mathrm{r}, \mathrm{C}, @, \mathrm{d}' \rangle$ corresponds to a two-dimensional modal model M′ of $\mathcal{L}_{\Box@}$ on a modal frame $\langle \mathrm{W}, \mathrm{R} \rangle$ with initial world w′ iff $\mathrm{D} = \mathrm{W}$, r is the identity function, $\langle u, v \rangle \in \mathrm{C}$ iff $\langle v, u \rangle \in \mathrm{R}$, and $\mathrm{d}' = @$. And the corresponding two-dimensional model M has $\mathrm{w} \in \mathrm{M}_P$ iff $\mathrm{w} \in \mathrm{M}'_P$.

Theorem 6. If $\mathrm{f}(w) = \mathrm{w}$ and $\mathrm{g}(w) = \mathrm{w}'$ then $\mathrm{M} \models_{\mathrm{f,g}} A$ iff $\mathrm{M}' \models'_{\mathrm{w,w'}} \sigma(A)$, where M and M′ are as in Definition 23.

Except for actuality formulas, the proof of this theorem is the same as the one that establishes the equivalence of modal and corresponding first-order models. For instance, see [Blackburn *et al.*, 2001, Section 2.4] for details concerning this sort of result. The generalization to two dimensions and case of the induction for actuality formulas are straightforward.

From the preceding two theorems, we have the equivalence of validity for two-dimensional modal logic and counterpart theory in the propositional case.

Theorem 7. $\sigma(A)$ is valid in $\mathcal{L}_{\Box@}$ iff A is valid in $\mathcal{L}^{WLC}_{Coh\forall@\mathrm{w}}$.

This result provides some positive support for the idea that two-dimensional counterpart theory does not suffer from anomalies that can be ascribed to its treatment of the actuality operator. To strengthen this support, we need to address stronger fragments of the two logics, supporting quantifiers over individuals.

6.2. Comparing counterpart theory to first-order modal logic

We now consider the much more complex case where first-order quantifiers are allowed. We begin by characterizing the modal logic.

6.2.1. The first-order modal logic $\mathcal{L}_{\Box@\forall}$

$\mathcal{L}_{\Box@\forall}$ is obtained from $\mathcal{L}_{\Box@}$ by allowing (plain) variables in atomic formulas and adding a clause for universally quantified formulas: if A is a formula of $\mathcal{L}_{\Box@\forall}$ so is $\forall x\, A$, where x is any plain variable. Identities are counted as atomic formulas.

Definition 24. $\mathcal{L}_{\Box@\forall}$ frames and models.

A $\mathcal{L}_{\Box@\forall}$ frame is a tuple $\langle \mathrm{W}, \mathrm{R}, \mathrm{D}\rangle$, where $\langle \mathrm{W}, \mathrm{R}\rangle$ is a modal frame and D is a nonempty set. A $\mathcal{L}_{\Box@\forall}$ model on a frame $\langle \mathrm{W}, \mathrm{R}, \mathrm{D}\rangle$ assigns each n-place predicate P of $\mathcal{L}_{\Box@\forall}$ a subset of $\mathrm{D}^n \times \mathrm{W}$.

Definition 25. $\mathcal{L}_{\Box@\forall}$ satisfaction.

The only new case is that of formulas $\forall x\, A$:

$\mathrm{M} \models_{\mathrm{f,w,w'}} \forall x\, A$ iff for all $\mathrm{d} \in D$, $\mathrm{M} \models_{\mathrm{f[d/}x\mathrm{],w,w'}} \forall x\, A$.

6.2.2. Adjusting counterpart theory for the comparison

A plausible and fair comparison between counterpart theory and first-order modal logic requires several adjustments, all of them quite independent from considerations having to do with actuality.

First, we need to address a fundamental difference in the way quantifiers are interpreted in the two theories. This difference is illustrated by the formula $\forall x\, Px \to \Box\forall x\, Px$, which was mentioned above in Section 4.2.

This formula is valid in counterpart theory, but is blatantly invalid in possible-worlds based first-order modal logic. In counterpart theory, the quantifier ranges over all counterparts, so that its truth means that not only everything in the actual world satisfies P, but also everything in every other world. The interconnection between truth about counterparts and necessity in counterpart theory produces the validity. There is no natural direct correspondence, then, between modal $\forall x A$ and counterpart theoretical $\forall x A$.

To adjust for this difference, we need to use relativized quantifiers to translate modal first-order quantifiers in counterpart theory. The relativized quantifiers are restricted to individuals that belong to the world of evaluation. Although standard counterpart theory does not appeal to a "world of evaluation," the special variable w in the translated formulas provides what is needed here. Relativized quantifiers in the two-dimensional first-order language $\mathcal{L}^{\mathrm{FOL}}_{\mathrm{2D@}}$ have the form $\forall x[\mathsf{Coh}\, xw \to A]$, where x is a plain variable.

The following definition introduces the sublanguage $\mathcal{L}^{WLC}_{Coh\forall@}$ of two-dimensional counterpart theory of actuality. $\mathcal{L}^{WLC}_{Coh\forall@}$ is analogous to the language $\mathcal{L}^{WLC}_{Coh}$ discussed in Section 3.2.

Definition 26. The language $\mathcal{L}^{WLC}_{Coh\forall @}$.

Atomic formulas of $\mathcal{L}^{WLC}_{Coh\forall @}$ are the correlates of atomic formulas of first-order modal logic. That is, they contain no free variables and (except for identities) have w as an extra variable. The formulas of $\mathcal{L}^{WLC}_{Coh}$ are closed under boolean operations, relativized quantification, counterpart necessity, and actuality.

(1) Where x and y are plain variables, $\mathsf{C}\,xy$ and $\mathsf{Coh}\,xy$ are atomic formulas of $\mathcal{L}_{\Box @\forall}$.

(2) If $Px_1 \ldots x_n$ is an atomic formula of $\mathcal{L}_{\Box @\forall}$, then $Px_1 \ldots x_n w$ is an atomic formula of $\mathcal{L}^{WLC}_{Coh}$.

(3) $x{=}y$ is an atomic formula of $\mathcal{L}^{WLC}_{Coh}$.

(4) If A is a formula of $\mathcal{L}^{WLC}_{Coh\forall @}$, so is $[@]\,A$.

(5) If A is a formula of $\mathcal{L}^{WLC}_{Coh\forall @}$, so is $\forall x[\mathsf{Coh}\,xw \to A]$.

(6) If A is a formula of $\mathcal{L}^{WLC}_{Coh\forall @}$, $x_1, \ldots, x_n$ are the free variables in A, and for $1 \leqslant i \leqslant n$, x^c_i is the first companion of $\mathrm{A}x_i$ not occurring in A, then $\forall x^c_1 \ldots \forall x^c_n[[\mathsf{C}\,x^c_1 x_1 \ldots \mathsf{C}\,x^c_n x_n \wedge \mathsf{Coh}\,x^c_1 \ldots x^c_n] \to A$ is a formula of $\mathcal{L}^{WLC}_{Coh\forall @}$.

The satisfaction clause given in Definition 18 for $\forall w^c[\mathsf{C}\,w^c w \to Pw^c]$ refers to the satisfaction conditions of $\mathsf{C}\,w^c w \to Pw^c$, which is not a formula of $\mathcal{L}_{\Box @\forall}$, since it contains a free companion variable. The following remark provides a way around this problem.

Remark 5. $\mathrm{M} \models_{\mathrm{f,g}} \forall x^c_1 \ldots \forall x^c_n[[\mathsf{C}\,x^c_1 x_1 \ldots \mathsf{C}\,x^c_n x_n \wedge \mathsf{Coh}\,x^C_1 \ldots x^C_n] \to A]$ iff for all $\mathrm{d}_1, \ldots, \mathrm{d}_n \in \mathrm{D}$, $\mathrm{M} \models_{\mathrm{f}[\mathrm{d}/y_1 \ldots \mathrm{d}/y_n],\mathrm{g}} A^{y_1}/x^c_1 \ldots {}^{y_n}/x^c_n$, where $y_1, \ldots, y_n$ are plain variables not occurring in A.

6.2.3. Equivalence of $\mathcal{L}^{WLC}_{Coh\forall @}$ and first-order modal logic

We map formulas of $\mathcal{L}_{\Box @\forall}$ into the language $\mathcal{L}^{WLC}_{Coh\forall @}$ by adding an extra argument w to atomic formulas. As in Definition 12, w is a designated (plain) variable of $\mathcal{L}^{WLC}_{Coh\forall @}$. Boolean operators and the actuality operator are treated homomorphically, first-order quantifiers are translated with a restricted quantifier, and the Lewis scheme with companion variables is used to translate necessity formulas.

The extra argument allows us to accommodate modal formulas without free variables. If, for instance, the translation A' of A were to have no free variables, then $A' \leftrightarrow \Box A'$ would be counterpart-valid.

Definition 27. The translation $\pi(A)$.

For purposes of the translation, we suppose that the language of $\mathcal{L}^{\mathrm{FOL}}_{2\mathrm{D@}}$ contains all the variables of the first-order language, as well as a designated plain variable w that does not belong to the modal language, and a suite of companion variables for all of the plain variables.

The translation function π from formulas of $\mathcal{L}_{\Box @ \forall}$ to formulas of $\mathcal{L}^{\mathrm{FOL}}_{2\mathrm{D@}}$ is defined as follows:

(1) $\pi(Px_1 \dots x_n) = Px_1 \dots x_n w$.
(2) π is homomorphic for identities, boolean formulas, and actuality.
(3) $\pi(\forall x\, A) = \forall x[\mathsf{Coh}\, xw \to \pi(A)]$.
(4) $\pi(\Box A) = \forall x^c_1 \dots \forall x^c_n \forall w^c[[\mathsf{C}\, x^c_1 x_1 \wedge \dots \wedge \mathsf{C}\, x^c_n x_n \wedge \mathsf{C}\, w^c w \wedge \mathsf{Coh}\, x^c_1 \dots x^c_n w^c] \to \pi(A)^{x^c_1}/x_1 \dots {}^{x^c_n}/x_n {}^{w^c}/w]$, where $x_1, \dots, x_n$ are all the variables occurring free in A and $x^c_1, \dots, x^c_n, w^c$ do not occur in $\pi(A)$.
(5) $\pi([@]A) = [@]\pi(A)$.

Counterpart theory is much less constrained about individuation than first-order modal logic. To obtain a fair comparison, we confine our attention to *simple* models of counterpart theory, in which the counterpart relation imitates the individuation policy of first-order modal logic.

Definition 28. Simple model of counterpart theory.

Let M be a model on a two-dimensional counterpart frame $\mathcal{F} = \langle \mathrm{D}, \mathcal{C}, \mathrm{r}, \mathrm{C}, @, \mathrm{d}' \rangle$. M is a *simple model* iff (i) each individual in M has exactly one counterpart in each world, and in fact there is a function $\textsc{Ctrprt}$ from individuals and "worlds" in r(D) to individuals such that $\textsc{Ctrprt}(\mathrm{d}, \mathrm{w})$ is the unique counterpart of d inhabiting w:

For all $\mathrm{d}, \mathrm{e} \in \mathrm{M}$, $\mathrm{C}(\mathrm{d}, \mathrm{e})$ and $\mathrm{r}(\mathrm{e}) = \mathrm{w}$ iff $\textsc{Ctrprt}(\mathrm{d}, \mathrm{w}) = \mathrm{e}$.

Also, (ii) $\mathcal{C}(\mathrm{d}) = \textsc{Ctrprt}(\mathrm{d}, \mathrm{w}')$ and (iii) for all $n+1$-place predicates P of $\mathcal{L}^{\mathrm{FOL}}_{2\mathrm{D@}}$ other than the reserved predicates C and Coh, if $\langle \mathrm{d}_1, \dots, \mathrm{d}_n, \mathrm{w} \rangle \in M_P$ then $\mathrm{r}(\mathrm{w}) = \mathrm{w}$ and for all i, $1 \leqslant i \leqslant n$, $\mathrm{r}(\mathrm{d}_i) = \mathrm{w}$.

Definition 29. Validity in simple models of $\mathcal{L}^{\mathrm{FOL}}_{2\mathrm{D@}}$, validity in $\mathcal{L}^{WLC}_{Coh\forall @}$.

A formula A of $\mathcal{L}^{\mathrm{FOL}}_{2\mathrm{D@}}$ is valid in on simple models iff $\mathrm{M} \models_{\mathrm{f},\mathrm{g}} A$ for every simple model M and every initial assignment pair $\langle \mathrm{f}, \mathrm{g} \rangle$ on the frame of M. And a formula A of $\mathcal{L}^{WLC}_{Coh\forall @}$ is valid iff A is valid on simple models.

Simple models of counterpart theory validate the following two formulas, characteristic of first-order modal logic:

$$\forall x \forall y[x = y \to \Box x = y],$$
$$\forall x \forall y[\Diamond x = y \to x = y].$$

We turn now to the relation between the translation π and satisfaction in models, beginning by showing how to construct an equivalent counterpart model from a modal model.

Definition 30. Two-dimensional counterpart model corresponding a two-dimensional first-order modal model.

Let M' be a two-dimensional model of $\mathcal{L}_{\Box@\forall}$ on a Kripke frame $\langle \mathrm{W}, \mathrm{R}, \mathrm{D}\rangle$, where R is reflexive, and let $\mathrm{w}' \in \mathrm{W}$. The counterpart model for $\mathcal{L}^{\mathrm{FOL}}_{2\mathrm{D}@}$ *corresponding to* M' and w' is the model M defined as follows. The frame of M is $\langle \mathcal{D}, \mathcal{C}, \mathrm{r}, \mathrm{C}, @, \mathrm{d}'\rangle$, where:

(1) $\mathcal{D} = \mathrm{W} \times \mathrm{D}$;
(2) $\mathrm{r}(\langle \mathrm{w}, \mathrm{d}\rangle) = \langle \mathrm{w}, \mathrm{d}_0\rangle$, where d_0 is an arbitrary fixed member of D;
(3) $\mathcal{C}(\langle \mathrm{w}, \mathrm{d}\rangle) = \langle \mathrm{w}', \mathrm{d}\rangle$;
(4) $\langle\langle \mathrm{w}', \mathrm{d}'\rangle, \langle \mathrm{w}, \mathrm{d}\rangle\rangle \in \mathrm{C}$ iff $\langle \mathrm{w}, \mathrm{w}'\rangle \in \mathrm{R}$ and $\mathrm{d} = \mathrm{d}'$;
(5) $@ = \mathrm{d}' = \langle \mathrm{w}', \mathrm{d}_0\rangle$.

Finally, where P is an n-place predicate of $\mathcal{L}_{\Box@\forall}$, $\langle\langle \mathrm{u}_1, \mathrm{d}_1\rangle, \ldots, \langle \mathrm{u}_n, \mathrm{d}_n\rangle, \langle \mathrm{w}, \mathrm{d}\rangle\rangle \in \mathrm{M}_P$ iff $\mathrm{u}_1 = \ldots = \mathrm{u}_n = \mathrm{w}$, $\mathrm{d} = \mathrm{d}_0$ and $\langle \mathrm{d}_1, \ldots, \mathrm{d}_n, \mathrm{w}\rangle \in \mathrm{M}'_P$.

Remark 6. Let M' and M be as in Definition 30. Then M is simple with function CTRPRT, where $\textsc{Ctrprt}(\langle \mathrm{w}, \mathrm{d}\rangle, \langle \mathrm{u}, \mathrm{d}_0\rangle) = \langle \mathrm{u}, \mathrm{d}\rangle$.

Definition 31. Variable assignment pair $\langle \mathrm{f}, \mathrm{g}\rangle$ on M corresponding to parameters h, w, w', on a two-dimensional modal model M'.

Let M' and M be as in Definition 30, let h be a variable assignment for the modal language $\mathcal{L}_{\Box@\forall}$, and let $\mathrm{w}, \mathrm{w}' \in \mathrm{W}$. A variable assignment pair $\langle \mathrm{f}, \mathrm{g}\rangle$ on M *corresponds to* $\mathrm{h}, \mathrm{w}, \mathrm{w}'$, i.e., $\langle \mathrm{h}, \mathrm{w}, \mathrm{w}', \mathrm{f}, \mathrm{g}\rangle \in \Pi$, iff:

(1) $\mathrm{f}(x) = \langle \mathrm{w}, \mathrm{h}(x)\rangle$ and $\mathrm{g}(x) = \langle \mathrm{w}', \mathrm{h}(x)\rangle$, for all plain variables x;
(2) $\mathrm{f}(w) = \langle \mathrm{w}, \mathrm{d}_0\rangle$ and $\mathrm{g}(w) = \langle \mathrm{w}', \mathrm{d}_0\rangle$.

These definitions are motivated by the intuitive correspondence between statements about an individual d in a world w in the modal model and statements about the pair $\langle \mathrm{w}, \mathrm{d}\rangle$ in the corresponding counterpart model. Our first result shows that the correspondence we have defined preserves truth: a formula is satisfied over reflexive frames in the modal logic if and only if its translation is satisfied in the corresponding counterpart model.

Theorem 8. Let M' be a two-dimensional model of $\mathcal{L}_{\Box@\forall}$ on a Kripke frame $\langle \mathrm{W}, \mathrm{R}, \mathrm{D}\rangle$, where R is reflexive, and M be the corresponding two-dimensional counterpart model. Then, where A is a formula of $\mathcal{L}_{\Box@\forall}$ and $\langle \mathrm{h}, \mathrm{w}, \mathrm{w}', \mathrm{f}, \mathrm{g}\rangle \in \Pi$, $\mathrm{M}' \models_{\mathrm{h},\mathrm{w},\mathrm{w}'} A$ iff $\mathrm{M} \models_{\mathrm{f},\mathrm{g}} \pi(A)$.

Proof. All the conditions on counterpart frames from Definition 2 are met automatically except for Condition (3), which requires that if $\mathrm{r}(\mathrm{d}) = \mathrm{r}(\mathrm{e})$ then $\langle \mathrm{d}, \mathrm{e}\rangle \in \mathrm{C}$ iff $\mathrm{d} = \mathrm{e}$. This condition, amounting to $\langle \mathrm{d}, \mathrm{d}\rangle \in \mathrm{C}$ for all $\mathrm{d} \in \mathcal{D}$, follows from the reflexivity of R.

The result is proved by induction on the complexity of A. We omit the cases for boolean formulas, which, as usual, are trivial. To simplify the presentation, we confine ourselves (except in the case of identities $x = y$) to formulas of $\mathcal{L}_{\Box@\forall}$ containing a single free variable, x. The arguments for the general cases don't differ in any important respects.

If A is an atomic formula Px, $\pi(A)$ is Pxw. Now, $\mathrm{M}' \models_{\mathrm{h,w,w'}} Px$ iff $\langle \mathrm{h}(x), \mathrm{w}\rangle \in \mathrm{M}'_P$. And this iff $\langle\langle \mathrm{w}, \mathrm{f}(x)\rangle, \langle \mathrm{w}, \mathrm{d}_0\rangle\rangle \in \mathrm{M}_P$. And this iff $\mathrm{M} \models_{\mathrm{f,g}} Pxw$, where $\mathrm{f}(w) = \langle \mathrm{w}, \mathrm{d}_0\rangle$, $\mathrm{f}(x) = \langle \mathrm{w}, \mathrm{h}(x)\rangle$.

If A is $x = y$, $\mathrm{M}' \models_{\mathrm{h,w,w'}} A$ iff $\mathrm{h}(x) = \mathrm{h}(y)$. But this iff $\langle \mathrm{w}, \mathrm{h}(x)\rangle = \langle \mathrm{w}, \mathrm{h}(y)\rangle$, this iff $\mathrm{f}(x) = \mathrm{f}(y)$, and this iff $\mathrm{M} \models_{\mathrm{f,g}} A$.

If A is $\Box B$, where x is the only free variable occurring in A, then $\pi(A)$ is $\forall x^x \forall x^c[[\mathsf{C}\, x^c x \wedge \mathsf{C}\, w^c w \wedge \mathsf{Coh} x^x w^c] \to \pi(B)^{x^c}/x^{w^c}/w]$.
Now, $\mathrm{M}' \models_{\mathrm{h,w,w'}} A$ iff

(i) $\mathrm{M}' \models_{\mathrm{h,u,w'}} B$ for all $\mathrm{u} \in \mathrm{W}$ such that $\langle \mathrm{w}, \mathrm{u}\rangle \in \mathrm{R}$.

By the hypothesis of induction, we have (i) iff

(ii) for all $\mathrm{u} \in \mathrm{W}$ such that $\langle \mathrm{w}, \mathrm{u}\rangle \in \mathrm{R}$, $\mathrm{M} \models_{\mathrm{f^u,g^u}} \pi(B)$,
where $\langle \mathrm{h}, \mathrm{u}, \mathrm{w}', \mathrm{f^u}, \mathrm{g^u}\rangle \in \Pi$.

Let $\langle \mathrm{h}, \mathrm{w}, \mathrm{w}', \mathrm{f}, \mathrm{g}\rangle \in \Pi$. Then $\mathrm{g} = \mathrm{g^u}$ and (since the only free variables occurring in $\pi(B)$ are x and w), $\mathrm{M} \models_{\mathrm{f^u,g^u}} \pi(B)$ iff $\mathrm{M} \models_{\mathrm{f}[\langle \mathrm{u},\mathrm{h}(x)\rangle/x, \langle \mathrm{u}, \mathrm{d}_0\rangle/w],\mathrm{g}} \pi(B)$.
Therefore, (ii) iff

(iii) for all $\mathrm{u} \in \mathrm{W}$ such that $\langle \mathrm{w}, \mathrm{u}\rangle \in \mathrm{R}$, $\mathrm{M} \models_{\mathrm{f}[\langle \mathrm{u},\mathrm{h}(x)\rangle/x, \langle \mathrm{u}, \mathrm{d}_0\rangle/w]} \pi(B)$.

And (iii) iff

(iv) for all $\delta, \epsilon \in \mathrm{W} \times \mathrm{D}$, if $\mathrm{M} \models_{f[\delta/x^c, \epsilon/w^c]} \mathsf{C}\, x^c x \wedge \mathsf{C}\, w^c w \wedge \mathsf{Coh}\, x^c w^c$ then
$\mathrm{M} \models_{f[\delta/x^c, \epsilon/w^c]} \pi(B)^{x^c}/x^{w^c}/w$.

Finally, (iv) iff $\mathrm{M} \models_{\mathrm{f,g}} \forall x^c \forall w^c[[\mathsf{C}\, x^c x \wedge \mathsf{C}\, w^c w \wedge \mathsf{Coh}\, x^c w^c] \to \pi(B)^{x^c}/x^{w^c}/w]$,
i.e., iff $\mathrm{M} \models_{\mathrm{f,g}} \pi(A)$.

If A is $[@]B$, then $\mathrm{M}' \models_{\mathrm{h,w,w'}} A$ iff

(i) $\mathrm{M}' \models_{\mathrm{h,w',w'}} B$.

By the hypothesis of induction, (i) iff

(ii) $\mathrm{M} \models_{\mathrm{f,g}} \pi(B)$, where $\langle \mathrm{h}, \mathrm{w}', \mathrm{w}', \mathrm{f}, \mathrm{g}\rangle \in \Pi$.

But then $\mathrm{f} = \mathrm{g}$. So, finally, (ii) iff $\mathrm{M} \models_{\mathrm{g,g}} \pi(B)$, and this iff $\mathrm{M} \models_{\mathrm{f,g}} \pi([@]B)$.

If A is $\forall x\, B$, then $\pi(A)$ is $\forall x[\mathsf{Coh}\, xw \to \pi(B)$. Now, $\mathrm{M}' \models_{\mathrm{h,w,w'}} A$ iff for all $\mathrm{d} \in D$, $\mathrm{M}' \models_{\mathrm{h}[\mathrm{d}/x],\mathrm{w,w'}} B$. By the hypothesis of induction, this iff

(i) for all $\mathrm{d} \in \mathrm{D}$, $\mathrm{M} \models_{\mathrm{f^d,g^d}} \pi(B)$, where $\langle \mathrm{h}[\mathrm{d}/x], \mathrm{w}, \mathrm{w}', x, \mathrm{f^d}, \mathrm{g^d}\rangle \in \Pi$.

Let $\langle \mathrm{h}, \mathrm{w}, \mathrm{w}', \mathrm{f}, \mathrm{g}\rangle \in \Pi$. Then $\mathrm{f}[\langle \mathrm{w}, \mathrm{d}\rangle/x]$ and $\mathrm{f^d}$ agree on all variables occurring free in $\pi(B)$, and $\mathrm{g}[\langle \mathrm{w}', \mathrm{d}\rangle/x]$ and $\mathrm{g^d}$ agree on all variables occurring free in $\pi(B)$. In particular, $\mathrm{g^d}(x) = \langle \mathrm{w}', \mathrm{d}\rangle$. So (i) iff

(ii) for all $\mathrm{d} \in \mathrm{D}$, $\mathrm{M} \models_{\mathrm{f}[\langle \mathrm{w},\mathrm{d}\rangle/x],\mathrm{g}[\langle \mathrm{w}',\mathrm{d}\rangle/x]} \pi(B)$.

Now, (ii) iff

(iii) for all $\delta \in \mathrm{W} \times \mathrm{D}$ such that $\mathrm{r}(\delta) = \mathrm{w}$, $\mathrm{M} \models_{f[\delta/x],\mathrm{g}[\mathrm{e}(\delta)/x]} \pi(B)$.

And (iii) amounts to this: for all δ such that $\mathrm{r}(\delta) = \langle \mathrm{w}, \mathrm{d}_0 \rangle$, $\mathrm{M} \models_{\mathrm{f}[\delta/x],\mathrm{g}[\mathfrak{e}(\delta)/x]} \pi(B)$.

So (iii) iff

(iv) for all $\delta \in \mathrm{W} \times \mathrm{D}$, $\mathrm{M} \models_{f[\delta/x],\mathrm{g}[\mathfrak{e}(\delta)/x]} \mathsf{Coh}\, xw \to \pi(B)$.

By Clause (3) of Definition 27, and using Clause (3.1) of Definition 18, (iv) iff $\mathrm{M} \models_{\mathrm{f},\mathrm{g}} \pi(A)$.

This concludes the induction, and the proof of Theorem 8.

It follows from this result that if $\pi(A)$ is valid in $\mathcal{L}^{WLC}_{Coh\forall@}$, then A is valid in the first-order normal modal logic $\mathbf{T}$ (the normal modal logic with axiom $\Box A \to A$) supplemented with an actuality operator.

Theorem 9. If $\pi(A)$ is valid in $\mathcal{L}^{WLC}_{Coh\forall@}$, then A is valid on reflexive frames for $\mathcal{L}_{\Box@\forall}$, for all formulas A of $\mathcal{L}_{\Box@\forall}$.

Proof. Suppose that a model M of $\mathcal{L}_{\Box@\forall}$ on a reflexive frame fails to satisfy A. Then the corresponding model M′ constructed according to the proof of Theorem 8 fails to satisfy $\pi(A)$ in $\mathcal{L}^{WLC}_{Coh\forall@}$, and if $\langle \mathrm{h}, \mathrm{w}, \mathrm{w}' \rangle$ is modally initial and $\langle \mathrm{h}, \mathrm{w}, \mathrm{w}', \mathrm{f}, \mathrm{g} \rangle \in \Pi$, then $\langle \mathrm{f}, \mathrm{g} \rangle$ is counterpart initial.

In the other direction, we construct a model of $\mathcal{L}_{\Box@\forall}$ from a simple model of $\mathcal{L}^{WLC}_{Coh\forall@}$.

Definition 32. Two-dimensional first-order modal model corresponding a simple two-dimensional counterpart model with base world w′.

Let M be a simple model of $\mathcal{L}^{WLC}_{Coh\forall@}$ on a two-dimensional counterpart frame $\mathcal{F} = \langle \mathrm{D}, \mathcal{C}, \mathrm{r}, \mathrm{C}, @, \mathrm{d}' \rangle$, with counterpart function $\textsc{Ctrprt}$. The two-dimensional first-order modal model ′ corresponding to M and its Kripke frame $\langle \mathrm{W}', \mathrm{R}', \mathrm{D}' \rangle$ is defined as follows.

(1) W' is $\mathrm{r}(\mathrm{D})$;

(2) Where $\mathrm{w}, \mathrm{u} \in \mathrm{W}'$, $\langle \mathrm{w}, \mathrm{u} \rangle \in \mathrm{R}'$ iff $\langle \mathrm{u}, \mathrm{w} \rangle \in \mathrm{C}$;

(3) D' is the set $\{\mathrm{d} \,/\, \mathrm{r}(\mathrm{d}) = @\}$;

Finally, where P is an n-place predicate of $\mathcal{L}_{\Box@\forall}$ and $\mathrm{d}_1, \ldots, \mathrm{d}_n \in \mathrm{D}'$, let $\langle \mathrm{d}_1, \ldots, \mathrm{d}_n, \mathrm{w} \rangle \in \mathrm{M}'_P$ iff $\langle \textsc{Ctrprt}(\mathrm{d}_1, \mathrm{w}), \ldots, \textsc{Ctrprt}(\mathrm{d}_n, \mathrm{w}), \mathrm{w} \rangle \in \mathrm{M}_P$.

Definition 33. Satisfaction parameters h, w, w′, corresponding to the variable assignment pair $\langle \mathrm{f}, \mathrm{g} \rangle$.

Let M and M′ be as in Definition 32, and let $\langle \mathrm{f}, \mathrm{g} \rangle$ be a variable assignment pair on M. Then $\langle \mathrm{f}, \mathrm{g}, \mathrm{h}, \mathrm{w}, \mathrm{w}' \rangle \in \Pi'$ iff $\mathrm{w}' = \mathrm{d}' = @$, f is local on w, g is local on w′, and for all plain variables x, $\mathrm{g}(x) = \mathcal{C}(\mathrm{f}(x))$ and $\mathrm{h}(x) = \textsc{Ctrprt}(\mathrm{f}(x), @)$.

Remark 7. If $\langle \mathrm{f}, \mathrm{g}, \mathrm{h}, \mathrm{w}, \mathrm{w}' \rangle \in \Pi'$ then $\langle \mathrm{f}, \mathrm{g} \rangle$ is local on w and w′, $\mathrm{f}(x) = \textsc{Ctrprt}(\mathrm{h}(x), \mathrm{w})$, and $\mathrm{g}(x) = \textsc{Ctrprt}(\mathrm{h}(x), \mathrm{w}')$.

Theorem 10. Let M be a simple two-dimensional model of $\mathcal{L}^{WLC}_{Coh\forall@}$ on a two-dimensional counterpart frame $\mathcal{F} = \langle \mathrm{D}, \mathcal{C}, \mathrm{r}, \mathrm{C}, @, \mathrm{d}' \rangle$ with counterpart function $\textsc{Ctrprt}$, and let M′ be the two-dimensional first-order modal model corresponding to M', as in Definition 32. We then have the following correspondence between the two models:

For all formulas A of $\mathcal{L}_{\Box @ \forall}$, if $\langle \mathrm{f}, \mathrm{g}, \mathrm{h}, \mathrm{w}, \mathrm{w}' \rangle \in \Pi'$, then $\mathrm{M}' \models_{\mathrm{h},\mathrm{w},\mathrm{w}'} A$ iff $\mathrm{M} \models_{\mathrm{f},\mathrm{g}} A$.

Proof. We induce on the complexity of formulas A of $\mathcal{L}_{\Box @ \forall}$, confining ourselves (except for identities) to A containing only one free variable, x. Again, the general case doesn't differ importantly from this special case.

If A is a propositional atom Px, then $\pi(Px) = Pxw$. $\mathrm{M}' \models_{\mathrm{h},\mathrm{w},\mathrm{w}'} A$ iff $\langle \mathrm{h}(x), \mathrm{w} \rangle \in \mathrm{M}'_P$. And this iff $\langle \textsc{Ctrprt}(\mathrm{h}(x), \mathrm{w})), \mathrm{w} \rangle \in \mathrm{M}_P$. This iff $\langle \mathrm{f}(x), \mathrm{w} \rangle \in \mathrm{M}_P$, and, finally, this iff $\mathrm{M} \models_{\mathrm{f},\mathrm{g}} Pxw$.

If A is $x = y$, then $\pi(Px)$ is $x = y$. $\mathrm{M}' \models_{\mathrm{h},\mathrm{w},\mathrm{w}'} A$ iff $\mathrm{h}(x) = \mathrm{h}(y)$. This iff $\textsc{Ctrprt}(\mathrm{h}(x), \mathrm{w}) = \textsc{Ctrprt}(\mathrm{h}(y), \mathrm{w})$. This iff $\mathrm{f}(x) = \mathrm{f}(y)$, and this iff $\mathrm{M} \models_{\mathrm{f},\mathrm{g}} x{=}y$.

If A is $\Box B$, then $\mathrm{M}' \models_{\mathrm{h},\mathrm{w},\mathrm{w}'} A$ iff $\mathrm{M}' \models_{\mathrm{f},\mathrm{u},\mathrm{w}'} B$ for all u such that $\langle \mathrm{w}, \mathrm{u} \rangle \in \mathrm{R}'$. By the hypothesis of induction, this iff

(i) for all u such that $\langle \mathrm{w}, \mathrm{u} \rangle \in \mathrm{R}'$, $\mathrm{M} \models_{\mathrm{f}^{\mathrm{u}},\mathrm{g}^{\mathrm{u}}} \pi(B)$, where $\langle \mathrm{f}^{\mathrm{u}}, \mathrm{g}^{\mathrm{u}}, \mathrm{h}, \mathrm{u}, \mathrm{w}' \rangle \in \Pi'$.

Let $\langle \mathrm{f}, \mathrm{g}, \mathrm{h}, \mathrm{u}, \mathrm{w}' \rangle \in \Pi'$. Now, $\mathrm{f}(x) = \textsc{Ctrprt}(\mathrm{h}(x), \mathrm{w})$, $\mathrm{f}(w) = \mathrm{w}$, $\mathrm{f}^{\mathrm{u}}(x) = \textsc{Ctrprt}(\mathrm{h}(x), \mathrm{u})$, and $\mathrm{f}^{\mathrm{u}}(w) = \mathrm{u}$. Also, $\mathrm{g}^{u}(y) = \textsc{Ctrprt}(\mathrm{h}(y), \mathrm{w}')$ for all plain variables y. So f^{u} and $\mathrm{f}[{}^{\textsc{Ctrprt}(\mathrm{h}(x),\mathrm{u})}/_{x}, {}^{\mathrm{u}}/_{w}]$ agree on all variables occurring free in $\pi(B)$, and g^{u} and g agree on on all variables occurring free in $\pi(B)$.

Therefore, (i) iff

(ii) for all u such that $\langle \mathrm{w}, \mathrm{u} \rangle \in \mathrm{R}'$, $\mathrm{M} \models_{\mathrm{f}[\textsc{Ctrprt}(\mathrm{f}(x),\mathrm{u})/x, \mathrm{u}/w],\mathrm{g}} \pi(B)$.

Now, (ii) iff

(iii) for all $\mathrm{d}, \mathrm{u} \in \mathrm{D}$, if $\mathrm{M} \models_{\mathrm{f}[\mathrm{d}/x^c, \mathrm{u}/w^c],\mathrm{g}} \mathsf{C}\, x^c x \wedge \mathsf{C}\, w^c w \wedge \mathsf{Coh}\, x^c w^c$ then $\mathrm{M} \models_{\mathrm{f}[\mathrm{d}/x^c, \mathrm{u}/w^c]} \pi(B)^{x^c}/x^{w^c}/w$.

And this iff $\mathrm{M} \models_{\mathrm{f,g}} \forall x^c \forall w^c[[\mathsf{C}\, x^c x \wedge \mathsf{C}\, w^c w \wedge \mathsf{Coh}\, x^c w^c] \rightarrow \pi(B)^{x^c}/x^{w^c}/w]$,
i.e., iff $\mathrm{M} \models_{\mathrm{h,g}} \pi(\Box B)$.

If A is $\forall x B$, $\pi(A)$ is $\forall x[\mathsf{Coh}\, xw \rightarrow \pi(B)]$.
Now, $\mathrm{M}' \models_{\mathrm{h,w,w'}} A$ iff for all $\mathrm{d} \in \mathrm{D}'$, $\mathrm{M}' \models_{\mathrm{h[d/}x\mathrm{],w,w'}} B$. By the hypothesis of induction, this iff

(i) for all $\mathrm{d} \in \mathrm{D}'$, $\mathrm{M} \models_{\mathrm{f^d,g^d}} \pi(B)$, where $\langle \mathrm{f^d}, \mathrm{g^d}, \mathrm{h[^d/}x], \mathrm{w}, \mathrm{w'}\rangle \in \Pi'$.

Let $\langle \mathrm{f}, \mathrm{g}, \mathrm{h}, \mathrm{w}, \mathrm{w'}\rangle \in \Pi'$. Then $\mathrm{f^d}[^{\textsc{Ctrprt}(\mathrm{d,w})}/x]$ and $\mathrm{f^d}$ agree on w and x, the only free variables occurring in $\pi(B)$. And $\mathrm{g^d}[^{\textsc{Ctrprt}(\mathrm{d,w})}/x]$ and $\mathrm{g^d}$ also agree on the variables w and x.
Therefore, (i) iff

(ii) for all $\mathrm{d} \in \mathrm{D}'$, $\mathrm{M} \models_{\mathrm{f}[\textsc{Ctrprt}(\mathrm{d,w})/x],\mathrm{g}[\textsc{Ctrprt}(\mathrm{d},w)/x]} \pi(B)$.

Finally, (ii) iff $\mathrm{M} \models_{\mathrm{f,g}} \forall x[\mathsf{Coh}\, xw \rightarrow \pi(B)]$, i.e. iff $\mathrm{M} \models_{\mathrm{f,g}} \pi(\forall x B)$.

If A is $[@]B$, then $\mathrm{M}' \models_{\mathrm{h,w,w'}} (A)$ iff $\mathrm{M}' \models_{\mathrm{h,w'w'}} B$. By the hypothesis of induction, this iff $\mathrm{M} \models_{\mathrm{f,g}} \pi(B)$, where $\langle \mathrm{f}, \mathrm{g}, \mathrm{h}, \mathrm{w'}, \mathrm{w'}\rangle \in \Pi'$. But then $\mathrm{f} = \mathrm{g}$, so this iff $\mathrm{M} \models_{\mathrm{g,g}} \pi(B)$. Finally, this iff $\mathrm{M} \models_{\mathrm{f,g}} \pi([@]B)$.

This completes the induction, and the proof of the theorem.

Theorem 11. If a formula A of $\mathcal{L}_{\Box @ \forall}$ is valid on reflexive frames, then $\pi(A)$ is valid in simple models of $\mathcal{L}^{WLC}_{Coh\forall @}$.

Proof. Suppose that a simple model M of $\mathcal{L}^{WLC}_{Coh\forall @}$ fails to satisfy $\pi(A)$. Then the corresponding model M′ of $\mathcal{L}_{\Box @}$ constructed according to the proof of Theorem 10 fails to satisfy A. And if $\langle \mathrm{f}, \mathrm{g}\rangle$ is counterpart initial and $\langle \mathrm{f}, \mathrm{g}, \mathrm{h}, \mathrm{w'}, \mathrm{w'}\rangle \in \Pi'$ then $\langle \mathrm{h}, \mathrm{w}, \mathrm{w'}\rangle$ is modally initial (i.e., $\mathrm{w} = \mathrm{w'}$).

Finally, putting together Theorems 9 and 11, we have the equivalence of validity in simple models of two-dimensional counterpart theory and validity over reflexive frames in first-order two-dimensional modal logic.

Theorem 12. Validity in $\mathcal{L}_{\Box @}$ on reflexive frames coincides with validity in $\mathcal{L}^{WLC}_{Coh\forall @}$, with respect to the translation π.

Taken together, these results provide general support for the claim that if counterpart theory with actuality is properly constructed, the actuality operator itself does not produce any anomalous or unintuitive patterns of validity when effects due to counterpart theory itself are eliminated.

6.3. A list of allegedly problematic examples

Beyond the assurance provided by Theorem 12, we can, of course, look at the examples that have been presented in published attempts to arguments that counterpart theory cannot provide an adequate account of actuality. The following four formulas are taken from [Fara and Williamson, 2005].

1. $Px \wedge \neg [@] Px$.

   This formula appears literally in $\mathcal{L}^{\mathrm{FOL}}_{\mathrm{2D@}}$ as $Px \wedge \neg [@] Px$, which, as expected, is not satisfiable.

2. $\Diamond \exists x [[@] Fx \leftrightarrow [@] \neg Fx]$.

   The translation of this example into counterpart theory is

$$\exists w^c [\mathsf{C}\, w^c w \wedge \exists x [\mathsf{Coh}\, xw \wedge [@] F xw^c \leftrightarrow [@] \neg F xw^c]].$$

   This formula is not satisfiable in $\mathcal{L}^{\mathrm{FOL}}_{\mathrm{2D@}}$—which seems to be the desired result.

3. $x = y \wedge [@] \neg x = y$.

   This formula is rendered literally in $\mathcal{L}^{\mathrm{FOL}}_{\mathrm{2D@}}$ without any changes. It is not satisfiable, which again is the desired result.

4. $\Diamond \exists x [[@] Fx \wedge [@] \neg Fx]$.

   Again, we need to add an extra variable to the translation to get a sensible result. The translation of this into $\mathcal{L}^{\mathrm{FOL}}_{\mathrm{2D@}}$ is then

$$\exists w^c [\mathsf{C}\, w^c w \wedge \exists x [[@] F xw^c \wedge [@] \neg F xw^c]].$$

   This formula is not satisfiable.

We know of no other counterexamples mentioned in the literature that differ significantly from these.

7. Conclusion

We have avoided metaphysical and foundational issues in this paper, concentrating on purely logical shortcomings that have been alleged to attach to counterpart theory. We have tried to show that the challenge of adding a plausible actuality operator to counterpart-based modality has nothing to do with flaws in Lewis' approach to modality, but originates in the fact that ordinary first-order quantifiers forget their previous values—values that need to be retrievable in order to provide an adequate theory of actuality. And we have shown how adding a very limited memory mechanism to the underlying first-order logic allows a plausible theory of actuality to be developed within counterpart theory.

We do not intend these results to settle debate about the value of counterpart theory. The merits and plausibility of the theory remain at issue. But we hope to have shown convincingly that productive debate on these issues should concentrate on metaphysical issues having to do with individuation and the interpretation of the quantifiers, rather than on alleged logical defects of counterpart theory.

Bibliography

[Blackburn *et al.*, 2001] Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic*. Cambridge University Press, Cambridge, England, 2001.

[Braüner and Ghilardi, 2006] Torsten Braüner and Silvio Ghilardi. First-order modal logic. In Patrick Blackburn, Johan F. A. K. van Benthem, and Frank Wolter, editors, *Handbook of Modal Logic, Volume 3 (Studies in Logic and Practical Reasoning)*, pages 549–620. Elsevier Science Inc., New York, NY, USA, 2006.

[Corsi, 2002] Giovanna Corsi. Counterpart semantics: A foundational study on quantified modal logics. Technical Report ILLC 2002 PP-2002-20, ILLC/Department of Philosophy, University of Amsterdam, Amtsterdam, 2002.

[Cresswell, 2004] Max J. Cresswell. Adequacy conditions for counterpart theory. *Australasian Journal of Philosophy*, 82(1):28–41, 2004.

[Fara and Williamson, 2005] Michael Fara and Timothy Williamson. Counterparts and actuality. *Mind*, 114(453):1–30, 2005.

[Forbes, 1982] Graham Forbes. Canonical counterpart theory. *Analysis*, 42(1):33–37, 1982.

[Garson, 1984] James W. Garson. Quantification in modal logic. In Dov Gabbay and Franz Guenther, editors, *Handbook of Philosophical Logic, Volume II: Extensions of Classical Logic*, pages 249–307. D. Reidel Publishing Co., Dordrecht, 1984.

[Hazen, 1976] Allen P. Hazen. Expressive completeness in modal language. *Journal of Philosophical Logic*, 5(1):25–46, 1976.

[Hazen, 1979] Allen P. Hazen. Counterpart-theoretic semantics for modal logic. *Journal of Philosophy*, 76:319–338, 1979.

[Hodes, 1984] Harold T. Hodes. Axioms for actuality. *Journal of Philosophical Logic*, 13(1):27–34, 1984.

[Kaplan, 1978] David Kaplan. On the logic of demonstratives. *Journal of Philosophical Logic*, 8:81–98, 1978.

[Lewis, 1968] David K. Lewis. Counterpart theory and quantified modal logic. *Journal of Philosophy*, 69:26–39, 1968.

[Ramachandran, 1989] Murali Ramachandran. An alternative translation scheme for counterpart theory. *Analysis*, 49(3):131–141, 1989.